ORIGINAL RESEARCH

Comprehensive review on twin support vector machines

M. Tanveer[1] · **T. Rajani[1]** · **R. Rastogi[2]** · **Y. H. Shao[3]** · **M. A. Ganaie[1]**

Abstract
Twin support vector machine (TWSVM) and twin support vector regression (TSVR) are newly emerging efficient machine learning techniques which offer promising solutions for classification and regression challenges respectively. TWSVM is based upon the idea to identify two nonparallel hyperplanes which classify the data points to their respective classes. It requires to solve two small sized quadratic programming problems (QPPs) in lieu of solving single large size QPP in support vector machine (SVM) while TSVR is formulated on the lines of TWSVM and requires to solve two SVM kind problems. Although there has been good research progress on these techniques; there is limited literature on the comparison of different variants of TSVR. Thus, this review presents a rigorous analysis of recent research in TWSVM and TSVR simultaneously mentioning their limitations and advantages. To begin with, we first introduce the basic theory of support vector machine, TWSVM and then focus on the various improvements and applications of TWSVM, and then we introduce TSVR and its various enhancements. Finally, we suggest future research and development prospects.

Keywords Machine learning · Twin support vector machines (SVM) · Twin SVM · Survey of twin SVM · Review of twin SVM · Classification · Regression · Clustering

✉ M. Tanveer
mtanveer@iiti.ac.in

T. Rajani
taanirajani@gmail.com

R. Rastogi
reshma.khemchandani@sau.ac.in

Y. H. Shao
shaoyuanhai21@163.com

M. A. Ganaie
phd1901141006@iiti.ac.in

[1] Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore 453552, India

[2] Department of Computer Science, South Asian University, New Delhi, India

[3] School of Management, Hainan University, Haikou 570228, China

1 Introduction

SVM (Cortes and Vapnik 1995) is a prominent classification technique widely used since its inception. It was first introduced by Cortes and Vapnik (1995) in 1995 for binary classification problems. SVM seeks to find decision hyperplanes that determine the decision boundary which can classify the data points into two classes. These decision planes are called support hyperplanes and the distance between them is maximized by solving a quadratic programming problem (QPP). SVM is computationally powerful even in non-linearly separable cases by using kernel trick (Cristianini et al. 2000). SVM has remarkable advantages as it utilizes the idea of structural risk minimization (SRM) principle which provides better generalization as well as reduces error in the training phase. As a result of its superior performance even in non-linear classification problems, it has been implemented in a diverse spectrum of research fields, ranging from text classification, face recognition, financial application, brain-computer interface, bio-medicine, human action recognition, horse race odds prediction and multiple instance learning (Tong and Koller 2001; Agarwal and Tomar 2014; Schölkopf et al. 2004; Tay and Cao 2001; Gupta et al. 2019a; Noble 2004; Osuna et al. 1997; Hua and Sun 2001; Byvatov and Schneider 2003; Morra et al. 2010; Tanveer et al. 2020b; Vapnik 2013; Edelman 2007; Poursaeidi and Kundakcioglu 2014). SVM has also been used in feature selection methods (Le Thi and Nguyen 2017). Robust SVM penalized all the samples via new loss function for better generalization in presence of noise (Wang et al. 2010b). Although SVM has outperformed most other systems, it still has many limitations in dealing with complex data due to its high computational cost of solving QPPs and its performance highly depends upon the choice of kernel functions and its parameters. Many improvements have been made in the last decade to enhance the accuracy of SVM (Li et al. 2019, 2020). One such critical enhancement was generalized eigenvalue proximal SVM (GEPSVM) (Mangasarian and Wild 2001; Khemchandani and Chandra 2017; Shao et al. 2013b) that led the foundation of TWSVM (Khemchandani and Chandra 2007; Jayadeva and Chandra 2016). The idea behind GEPSVM is to determine two nonparallel hyperplanes via solving two generalized eigenvalue problems. TWSVM (Khemchandani and Chandra 2007; Jayadeva and Chandra 2016) enhanced the generalization ability of GEPSVM. TWSVM also determines two nonparallel hyperplanes and needs to solve a pair of small QPPs in lieu of solving one complex QPP in SVM. Computational cost of TWSVM is approximately one-fourth of the SVM. TWSVM is faster and has better generalization performance than SVM and GEPSVM. Authors also extended TWSVM for nonlinear classification problems by introducing kernel. But, in nonlinear cases when a kernel is used, the formulation requires to solve inverse of matrices. Also, its performance highly depends upon the selection of kernel functions and its parameters for nonlinear classification. Although TWSVM is still in its rudimentary stage, many improvements and variants have been proposed by researchers due to its favorable performance especially in-case of handling large datasets which is not possible with the conventional SVM. Huang et al. (2018) reviewed the research progress of TWSVM until 2017 for classification problems.

Support Vector Machine has also been used effectively for regression and is called support vector regression (SVR) (Basak et al. 2007). SVR is different than SVM in some aspects as it sets a margin of tolerance ϵ and find the optimal hyperplane such that the error is minimized. Thus, it finds a function such that error can be maximum of ϵ distance, thus any error within ϵ deviation is acceptable. Learning speed of SVR is low as it needs to solve a large sized QPP. Researchers have proposed many variants of SVR to improve its performance in terms of reducing computational complexity and improving accuracy (Drucker et al. 1997; Smola and

Schölkopf 2004). Some other variants of SVR includes ϵ−support vector regression (ϵ−SVR) (Shao et al. 2013e), Fuzzy SVM (Chuang 2007), ν−SVR (Chang and Lin 2002), robust and sparse SVR (Tanveer et al. 2016b) and some other algorithms (Yang et al. 2009; Elattar et al. 2010). ϵ−SVR has high computational cost and in order to remediate this, an efficient twin support vector regression (TSVR) (Peng 2010b) determined two nonparallel hyperplanes that clusters the data points in two classes by optimizing the regression function. TSVR uses ϵ−insensitive up and down-bound functions to optimize the final regression function. TSVR formulation is computationally less complex as it needs to solve a pair of smaller QPPs in lieu of solving one large QPP in SVR, thus it's speed is much more than SVR. However, in TSVR, the spirit of TWSVM was missing, improved TSVR (TWSVR) (Khemchandani et al. 2016) which works on the similar lines of TWSVM was proposed. To further boost the performance of TSVR, many algorithms have been introduced by researchers which are discussed in Sect. 3.

In the past years, twin support vector classification and regression algorithms have been developed rapidly and are implemented to solve some real-life challenges. However, there is limited literature on twin support vector regression as it is a relatively new theory and needs further study and improvement. Thus, the objective of this paper is to present an compendium of recent developments in TWSVM and TSVR, identify limitations and advantages of these techniques and promote future developments.

The framework of this paper is as follows, Sect. 2 presents brief about SVM and TWSVM, Sect. 3 and 4 include the recent advancements and applications of TWSVM respectively, Sect. 5 presents a brief about TSVR, Sects. 6 and 7 include the recent advancements and applications of TSVR respectively and at last Sect. 8 provides synopsis, future research and development prospects.

2 Related work

Suppose a binary classification problem with dataset $X \in \mathbb{R}^{l \times n}$ in which l_1 data points are in class $+1$ (termed as positive class) and l_2 data points are in class -1 (termed as negative class) and these are represented by matrix A and B respectively. For classification problems, the data label $Y \in \{-1, 1\}$ and for regression $Y \in \mathbb{R}$. Accordingly, these matrices will be of size $(l_1 \times n)$ and $(l_2 \times n)$ where n is feature space dimension and $l = l_1 + l_2$.

2.1 Support vector machine (Cortes and Vapnik 1995)

The optimization problem of support vector machine is given as follows:

$$\min_{u,\xi} \ \frac{1}{2} u^T u + c \sum_{k=1}^{l} \xi_k$$
$$s.t. \ \ y_k(u^T \phi(x_k) + b) \geq 1 - \xi_k, \ \ k = 1, \ldots, l$$
$$\xi_k \geq 0, k = 1, \ldots, l, \tag{1}$$

where, $c > 0$ is the regularisation parameter, and $\phi(x)$ represents the non-linear mapping of the input x and ξ is the vector of slack variables.

Form the above given optimization problem, one can see that all the data samples appear in the constraints of the optimization problem. Hence, the complexity of the model is high

as it involves solving a single large quadratic programming problem. Thus, the complexity of SVM is $O(l^3)$.

2.2 Twin support vector machines (Khemchandani and Chandra 2007)

Standard twin SVM is a binary classification model. To categorize data samples which cannot be separated by linear functions, TWSVM uses kernel functions to convert the higher dimensional data space to the required form. The two kernel generated surfaces are given as below:

$$K(x^T, D^T)u_+ + b_+ = 0 \text{ and } K(x^T, D^T)u_- + b_- = 0, \qquad (2)$$

here $D = [A; B]$; and K is a kernel function. The formulation for nonlinear TWSVM classification is defined as below:

$$\min_{u_+, b_+, \xi_1} \frac{1}{2}\|K(A, D^T)u_+ + e_1 b_+\|^2 + c_1 e_2^T \xi_1$$
$$s.t. \quad -(K(B, D^T)u_+ + e_2 b_+) + \xi_1 \geq e_2, \ \xi_1 \geq 0 \qquad (3)$$

and

$$\min_{u_-, b_-, \xi_2} \frac{1}{2}\|K(B, D^T)u_- + e_2 b_-\|^2 + c_2 e_1^T \xi_2$$
$$s.t. \quad K(A, D^T)u_- + e_1 b_- + \xi_2 \geq e_1, \ \xi_2 \geq 0, \qquad (4)$$

here $c_1, c_2 \geq 0$, e_1, e_2 are vectors of ones of suitable dimensions and ξ_1, ξ_2 are called slack variables. After using the Lagrange multipliers $\alpha \geq 0$, $\beta \geq 0$ and the Karush–Kuhn–Tucker (K.K.T.) (Kuhn and Tucker 1951) conditions, the duals of the QPPs in (3) and (4) are defined as below:

$$\max_{\alpha} e_2^T \alpha - \frac{1}{2}\alpha^T Q(P^T P)^{-1} Q^T \alpha$$
$$s.t. \ 0 \leq \alpha \leq c_1 \qquad (5)$$

and

$$\max_{\beta} e_1^T \beta - \frac{1}{2}\beta^T P(Q^T Q)^{-1} P^T \beta$$
$$s.t. \ 0 \leq \beta \leq c_2, \qquad (6)$$

where $P = [K(A, D^T) \ e_1]$ and $Q = [K(B, D^T) \ e_2]$. After solving (5) and (6), the proximal hyperplanes are given as follows:

$$\begin{bmatrix} u_+ \\ b_+ \end{bmatrix} = -(P^T P + \delta I)^{-1} Q^T \alpha, \qquad (7)$$

$$\begin{bmatrix} u_- \\ b_- \end{bmatrix} = (Q^T Q + \delta I)^{-1} P^T \beta, \qquad (8)$$

where δI is a regularization term and $\delta > 0$.

A new sample $x \in \mathbb{R}^n$ is allocated to either class on the basis of the proximity of kernel generated surface to x, i.e.,

$$class(x) = \text{sign}\left(\frac{K(x^T, D^T)u_+ + b_+}{\|u_+\|} + \frac{K(x^T, D^T)u_- + b_-}{\|u_-\|} \right), \tag{9}$$

where $sign(\cdot)$ is signum function.

From optimization problems (3) and (4), one can see that the constraints of only one class appear in the optimization problem while generating the hyperplane for the other class. Thus, the size of the constraints in QPPs of TWSVM is approximately half compared to the SVM formulation (assuming the data is balanced between the classes). TWSVM is approximately four times faster than the usual SVM. The complexity of TWSVM is $2 * O((l/2)^3)$. In terms of generalization, TWSVM and SVM have comparable generalization performance (Khemchandani and Chandra 2007).

3 Research progress on twin support vector machines

In this section, we discuss the progress of TWSVM based models in classification problems. The variants of TWSVM (given in Fig. 1) are

3.1 Least squares twin support vector machines

To reduce TWSVM training time, Kumar and Gopal (2009) formulated least squares TWSVM (LS-TWSVM) algorithm. The major advantage of LS-TWSVM over TWSVM is that it only deals with linear equations in place of QPPs in TWSVM which reduces the computational complexity of the model. LS-TWSVM has comparable accuracy but low computational time than TWSVM. Although the computational time of LS-TWSVM is less than TWSVM, its generalization performance is poor as same penalties are allotted to the samples either being positive or negative. LS-TWSVM apply the empirical risk minimization (ERM) principle which affects accuracy and also causes over-fitting problem. Also, it doesn't take into account the effects of samples having different locations. To avoid this limitation, weighted LS-TWSVM (Xu et al. 2014a) gives different weights to the samples depending upon their locations. Experimental results have shown that weighted LS-TWSVM yields better testing accuracy than TWSVM and LS-TWSVM but its computational time is more than these algorithms.

To incorporate expert's knowledge in LS-TWSVM classifier, knowledge-based LS-TWSVM (KBLS-TWSVM) (Kumar et al. 2010) included polyhedral knowledge sets in the formulation of LS-TWSVM. Experimental results have shown that KBLS-TWSVM is simple and more apt classifier compared to LS-TWSVM. In order to improve accuracy of LS-TWSVM, Wang et al. (2010a) incorporated the manifold geometric structure of data of each class. It requires to solve a set of linear equations.

Although, LS-TWSVM performed well with large datasets compared to TWSVM, Gao et al. (2011) designed l_1-norm LS-TWSVM (NLS-TWSVM) to automatically select the relevant features in order to strengthen the algorithm in dealing with high-dimensional datasets. NLS-TWSVM is based on LS-TWSVM and includes a Tikhonov regularization term. To outdo LS-TWSVM in accuracy and avoid over-fitting problem, Improved LS-TWSVM (ILS-TWSVM) (Xu et al. 2012) improved the classification accuracy by implementing structural risk minimization principle. ILS-TWSVM is faster and yields comparable generalization

and computational time to LS-TWSVM. Since L_2 norm magnifies the outlier effect in least squares TWSVM models, hence, capped $L_{2,p}$ norm based least squares TWSVM (Yuan and Yang 2021) was formulated to reduce the effect of outliers and noise.

To take advantage of the correlation between some data points and reduce the effect of noise, Ye et al. (2012a) used density of the sample to allot weights for each sample (DWLSC). Experimental results demonstrate better classification accuracy of the DWLSC than TWSVM and LS-TWSVM.

There are several real-life problems which are essentially multicategory problems. The extensions of TWSVM to multicategory is also an active research domain where several research papers have evolved addressing the same. However, class imbalance problem is common in multicategory classification as all binary SVMs are trained with all patterns. To deal with these issues Nasiri et al. (2014) formulated energy-based LS-TWSVM (ELS-TWSVM) in which the constraints of LS-TWSVM are converted to an energy model to decrease the impact of noise and outliers. TWSVM, LS-TWSVM, and ELS-TWSVM satisfy the empirical risk minimization (ERM) principle and also in these techniques the matrices are positive semi-definite. Thus, to remediate this problem, Tanveer et al. (2016a) embodied the SRM principle in ELS-TWSVM and proposed robust ELS-TWSVM (RELSTSVM) which makes the matrices positive definite. Moreover, RELSTSVM uses an energy parameter to handle the noise and thus make the algorithm more robust. Results have shown the promising generalization performance of RELSTSVM with less computational time when compared with the baseline algorithms (Table 1). In the recent comprehensive evaluation (Tanveer et al. 2019) of 187 classifiers on 90 datasets, RELSTSVM (Tanveer et al. 2016a) emerged as the best classifier. Khemchandani and Sharma (2016) proposed robust least squares TWSVM (RLS-TWSVM) which is fast and yields better generalization performance and is also robust to handle the noise. Further, the application in the field of human activity recognition is explored in the paper. The authors in Khemchandani and Sharma (2016) also proposed Incremental RLS-TWSVM to increase the training speed of RLS-TWSVM and Incremental RLS-TWSVM also deals with noise and outliers effectively. Tomar and Agarwal (2014) proposed feature selection based LS-TWSVM to diagnose heart diseases using F-score to choose the most relevant features which enhances the accuracy of the model. To handle imbalanced datasets, a novel weighted LS-TWSVM (Tomar et al. 2014c) for imbalanced data, which has better accuracy and geometric mean than SVM and TWSVM, was proposed. To further enhance the accuracy on imbalanced datasets, Hybrid Feature Selection Based WLS-TWSVM (HFS based WLS-TWSVM) (Tomar and Agarwal 2015a) approach was proposed for diagnosing diseases like Breast Cancer, Diabetes and Hepatitis which can tackle data imbalance issues. Xu et al. (2015) included data distribution information into the structural LS-TWSVM (SLS-TWSVM) classifier. It is based on structural TWSVM and performs clustering before classification and embody the structural information into the model. SLS-TWSVM has a better generalization performance than ν-SVM (Chen et al. 2005), ν-TWSVM (Peng 2010c), STWSVM (Qi et al. 2013a) and LS-TWSVM (Kumar and Gopal 2009). It also improves the noise insensitivity of LS-TWSVM.

Mei and Xu (2019) proposed a novel multi-task LS-TWSVM algorithm which is build on directed multi-task TWSVM (DMTWSVM) and LS-TWSVM. It focuses on multitask learning instead of commonly applied single task learning in TWSVM and LS-TWSVM. This algorithm is computationally effective as it only requires to solve linear equations in lieu of QPPs in DMTWSVM.

Least square twin support vector hypersphere (LSTSVH) (Tomar and Agarwal 2015a) is an enhancement of twin support vector hypersphere (TSVH) (Peng and Xu 2013b). TSVH is different from TWSVM as it obtains two hyperspheres through solving two small SVM

type problems. Experimental results demonstrate that LSTSVH has almost same accuracy as SVM, LS-SVM, TWSVM, LS-TWSVM, and TSVH but has high computational time than LS-SVM and LS-TWSVM and less than SVM and TWSVM. Tanveer et al. (2020a) proposed efficient large scale least squares twin support vector machines (LS-LSTWSVM) which uses different Lagrangian functions to eliminate the need for calculating the inverse matrices and thus enhance the performance of TWSVM on large scale datasets.

3.2 Projection twin support vector machine

Projection TWSVM (P-TWSVM) (Chen et al. 2011) is a multiple-surface classification (MSC) technique based on the multi-weight vector projection SVM (MVSVM) and TWSVM. Chen et al. (2011) introduced P-TWSVM with minimizing within class variance principle. Unlike finding hyperplanes in TWSVM, P-TWSVM finds projection axes to distinctly separate samples. To enhance performance of P-TWSVM, authors proposed a recursive algorithm in which for each class multiple projection axes are considered. P-TWSVM is a better classifier as hyperplanes are more fitted for single plane based samples while for datasets having complex sample distribution like XOR problems, projection directions can generate better accuracy. Experimental results have shown that P-TWSVM has better accuracy than GEPSVM, TWSVM, LS-TWSVM, and MVSVM. However, it is computationally more complex than TWSVM. Thus, Shao et al. (2012a) proposed improved P-TWSVM and used the idea in LS-SVM (Suykens et al. 1999) and LS-TWSVM (Kumar et al. 2010) and added a regularization term into P-TWSVM simultaneously maintaining the optimization problem to be positive definite. It is simple, fast and has similar classification accuracy but less computational time than P-TWSVM.

Although P-TWSVM is an efficient classifier, it only implements empirical risk minimization principle which can incur possible singularity problems which need to be dealt with principal component analysis (PCA) and linear discriminant analysis (LDA). Shao et al. (2013c) formulated a new P-TWSVM variant by introducing a maximum margin regularization term called RP-TWSVM in which empirical risk is replaced by regularized risk. Authors further proposed a successive over-relaxation (SOR) technique and a genetic algorithm (GA) for parameter selection to optimally solve the primal problems. RP-TWSVM not only has better accuracy but also has a better generalization than P-TWSVM.

To implement least square P-TWSVM to non-linear problems, Ding and Hua (2014) introduced kernel to LSP-TWSVM (Shao et al. 2012a) which also deals with linear equations similar to LSP-TWSVM and opposed to P-TWSVM. Authors introduced nonlinear recursive algorithm in order to improve its performance. Experimental results have shown good classification accuracy than LSP-TWSVM and P-TWSVM.

Guo et al. (2014) proposed a feature selection approach for LSP-TWSVM which finds two projection directions and has comparable prediction accuracy to that of TWSVM, LS-TWSVM and LSP-TWSVM and a similar generalization ability to TWSVM and LS-TWSVM.

LSP-TWSVM enhances the performance of P-TWSVM but fails to include underlying data information to enhance classification accuracy. To overcome this limitation, weighted LSP-TWSVM (LIWLSP-TWSVM) (Hua and Ding 2015) exploited local information into the algorithm. LIWLSP-TWSVM is more effective than LSP-TWSVM as it uses weighted mean rather than standard mean in LSP-TWSVM. It has better generalization ability but its computational cost is high when solving multi class problems (Table 2).

Table 1 Classification accuracy on non-linear kernel (Tanveer et al. 2016a)

| Dataset (train size, test size)	TWSVM (Khemchandani and Chandra 2007)	TBSVM (Shao et al. 2011)	LS-TWSVM (Kumar and Gopal 2009)	ELS-TWSVM (Nasiri et al. 2014)	RELSTSVM (Tanveer et al. 2016a)
Ripley (250×2, 1000×2)	88.00	83.50	86.40	83.50	88.30
Heart-c (177×13, 120×13)	60.81	70.00	61.67	61.67	67.50
Heart-stat (200×13, 70×13)	82.85	84.28	85.71	78.57	85.71
Ionosphere (246×33, 105×33)	92.38	87.46	87.61	92.38	96.19
Bupa-liver (241×6, 104×6)	63.46	70.19	55.77	66.35	69.23
Votes (306×16, 129×16)	96.90	96.90	96.12	95.35	96.90
WPBC (137×34, 57×34)	80.70	80.19	78.94	75.44	80.70
Pima-Indian (537×8, 231×8)	77.48	76.09	64.07	78.35	80.52
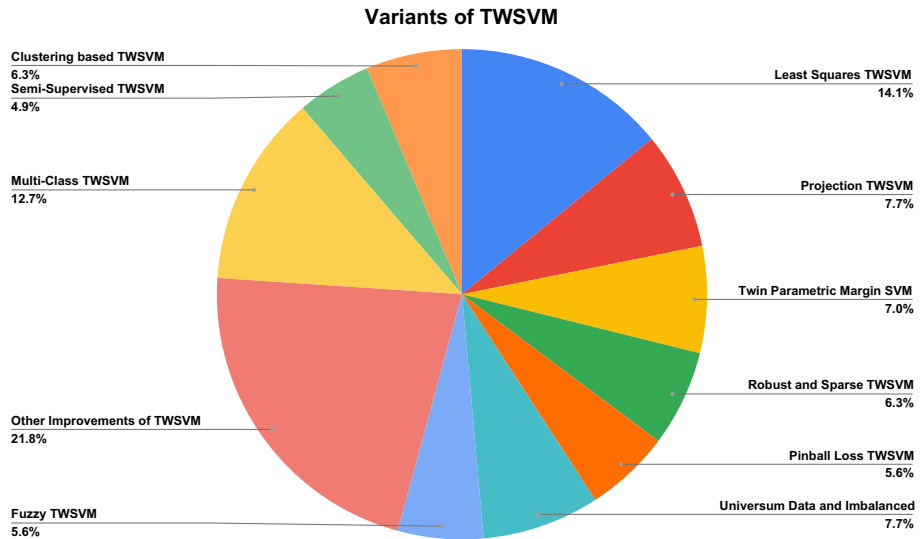
Fig. 1 Variants of TWSVM

Hua et al. (2017) formulated a novel projection TWSVM (NP-TWSVM). NP-TWSVM has many advantages over P-TWSVM. It is faster than P-TWSVM as calculation of inverse matrices is not avoided in NP-TWSVM. It determines the projection axes so that margin across the sample and class is greater than or equal to zero. Experimental results have shown that it obtains better generalization and accuracy when compared with other baseline algorithms. An efficient nonparallel sparse projection SVM (NPrSVM) (Chen et al. 2020c) finds optimal projection for each class such that the projection values of within class instance are clustered as much as possible within an insensitive tube while those of other class instance are kept away. Extensive experiments show that NPrSVM is superior than PTWSVM in terms of generalization performance, training time, sparsity and robustness to outliers. $\nu-$ projection twin SVM (Chen et al. 2020b) seeks projection axis for each class in a manner that ν controls the fraction of support vectors and error margin, and also avoids the matrix inversions. Robust rescaled hinge loss based projection TWSVM model (Ren and Yang 2021) use different parameters to control the effect of outliers and the model results in better performance.

3.3 Twin parametric margin support vector machine

Motivated by the idea of TWSVM and Par-ν-SVM (Hao 2010), Peng (2011a) formulated twin parametric-margin SVM (TPMSVM). Like par-ν-SVM, TPMSVM seeks to determine two parametric margin hyperplanes which defines the positive and negative margin respectively. It's an indirect classifier even suitable when noise is heteroscedastic as it automatically adjusts a flexible margin. Experimental results have shown that TPMSVM has comparable generalization ability to SVM, Par-ν-SVM, and TWSVM. Also, TPMSVM has much lower training time than Par-ν-SVM as it solves two small sized QPPs like TWSVM. TPMSVM classifier is not sparse due to the formulations of its parametric- margin hyperplanes. TPMSVM has good generalization but it is computationally complex as it solves two QPPs. Wang et al. (2013) added a quadratic function to TPMSVM in primal space to get better training speed.

Table 2 Classification accuracy on non-linear kernel (Hua and Ding 2015)

Dataset	TWSVM (Khemchandani and Chandra 2007)	WLTSVM (Ye et al. 2012b)	P-TWSVM (Chen et al. 2011)	LSP-TWSVM (Shao et al. 2012a)	LIWLSP-TWSVM (Hua and Ding 2015)
Cleve (296×13)	85.12	84.90	85.93	83.80	85.30
Bupa-liver (345×6)	74.13	73.58	72.18	72.47	73.39
Heart (270×13)	84.44	84.82	84.44	84.44	84.82
Monks2 (432×6)	68.77	66.21	68.71	67.84	68.95
Australian (690×14)	86.97	86.70	86.99	87.27	88.13
Wpbc (198×33)	79.36	80.35	78.03	78.56	81.87
Parkinsons (195×22)	87.22	85.00	86.36	86.62	87.48
Vertebral (310×6)	85.81	86.13	84.52	84.84	87.10
Spect (267×23)	84.42	85.19	84.42	84.80	85.11
Spectf (267×44)	83.43	82.96	82.49	83.26	84.25

The proposed algorithm is called smooth twin parametric-margin SVM (STPMSVM). Also, the authors proposed the use of genetic algorithm (GA) in STPMSVM to overcome the drawback of regularizing at least four parameters in TPMSVM. To obtain sparsity and feature noise insensitivity, truncated pinball loss TPMSVM (Wang et al. 2021a) was proposed. The optimal separating hyperplanes are obtained via concave-convex procedure (CCCP).

Shao et al. (2013d) proposed least squares TPMSVM (LSTPMSVM) to decrease the training cost of TPMSVM as LSTPMSVM solves two primal problems rather than dual problems. This change makes it less complex and increases training speed. It has comparable or better classification accuracy but with remarkably less computational time than TPMSVM. For model selection, the authors proposed a particle swarm optimizer (PSO) which effectively optimizes the four parameters defined in LSTPMSVM. In terms of generalization, LSTPMSVM performs better than TPMSVM and LS-TWSVM. In order to consider prior structural data information, Peng et al. (2013) proposed structural twin parametric-margin SVM (STPMSVM). Based on cluster granularity, the class data structures are included in the problem. STPMSVM not only obtained good generalization ability but also showed fast learning speed, and better performance than TPMSVM. TPMSVM is an efficient classifier but it losses the sparsity due to the weight vectors of the hyperplanes. To solve this issue, centroid-based TPSVM (CTPSVM) (Peng et al. 2015b) was introduced which uses the projection of the centroid points and leads to a sparse optimal hyperplane by optimizing the centroid projection values.

To incorporate structural information present in data, Rastogi et al. (2018d) proposed robust parametric TWSVM for pattern classification (RP-TWSVM) which seek to find two parametric margin hyperplanes that has the capability to adjust margin to capture heteroscedastic noise data information. Khemchandani et al. (2018b), formulated angle-based TPSVM (ATPSVM) is proposed, which can efficiently handle heteroscedastic noise. Other improvements which maximizes angle between the twin hyperplanes are proposed by Rastogi et al. Rastogi et al. (2018c). Taking motivation from TPMSVM, ATPSVM determines a pair of hyperplanes so that the angle between their normals is maximized. Recently, Richhariya and Tanveer proposed a novel angle based universum LS-TWSVM (AULSTWSVM) for pattern classification. In contrast to ATPSVM (Rastogi et al. 2018c), the AULSTWSVM minimizes the angle between universum hyperplane and classifying hyperplane. To incorporate the prior information about the distribution of the data using the universum, AULSTWSVM used linear loss (Shao et al. 2015) in the formulation. Numerical experiments show the promising generalization performance with very less computational time. Further, the application in the diagnosis of Alzheimer's disease is explored.

Moreover, the proposed AULSTWSVM includes linear loss (Shao et al. 2015) in the optimization problem, while incorporating prior information about data distribution using the universum. Moreover, the solution of AULSTWSVM involves system of linear equations leading to less computation time (Kumar and Gopal 2009)

3.4 Robust and sparse twin support vector machine

TWSVM achieves better accuracy and is faster when compared to conventional SVM but it loses sparsity as it considers l_2 norm of distances in the objective function. Thus, in order to make TWSVM sparse, Peng (2011b) proposed TWSVM in primal space which provides a sparse hyperplane with better generalization ability. To improve the robustness, a regularization term is also added. It has comparable generalization and a rapid learning speed to TWSVM and LS-TWSVM but the computational cost is high. Peng and Xu (2013c)

proposed robust minimum class variance TWSVM (RMCV-TWSVM) classifier to enhance generalization and robustness of TWSVM. This algorithm has an extra regularization term which makes its learning speed comparable to TWSVM but obtains better generalization ability than TWSVM.

Qi et al. (2013b) proposed robust TWSVM using second-order cone programming formulation. It is effective with noise-corrupted data. This algorithm overcomes the limitation of TWSVM and computes the inverse of matrices which are not suitable for large datasets, it takes only the inner products about samples by which kernel trick can be applied directly and does not need to solve the extra inverse of matrices. It is superior in computational time and accuracy than other TWSVMs. Tian et al. (2014a) proposed sparse nonparallel support vector machine (SN-SVM). While TWSVM losses the sparseness, SN-SVM has the inherent sparseness as it uses the two loss functions instead of one in existing TWSVM. Only the empirical risk is considered in TWSVM, SN-SVM introduces the regularization term by maximizing the margin. TWSVM minimizes the loss function based on l_1 or l_2-norm. Thus, Zhang et al. (2014) proposed l_p-norm-LS-TWSVM which is an adaptive as p can be automatically chosen by data. Robust non-parallel SVM via second order cone programming (López et al. 2019) is robust to outliers and noise, and constructs the two separating hyperplanes via maximisation of probabilistic framework.

In order to intensify the robustness and sparsity in the original formulation of TWSVM, Tanveer (2015b), incorporated regularization technique and formulated a linear programming l_1-norm TWSVM (NLP-TWSVM) which needs to solve linear equations rather than solving QPPs in TWSVM which makes it fast, robust, sparse and a simple algorithm. Experimental results demonstrate that NLP-TWSVM's generalization ability is better and computational time is less than GEPSVM, SVM, and TWSVM. Tanveer (2015c, 2013) proposed smoothing approaches for 1-norm linear programming TWSVM (SLPTWSVM). The solution of SLPTWSVM reduces to solving two systems of linear equations which makes the algorithm extremely simple and computationally efficient. Tanveer et al. (2019c) proposed sparse pinball TWSVM (SP-TWSVM) by introducing ϵ insensitive zone pinball loss function in the orginal TWSVM formulation. SP-TWSVM is noise insensitive, retain sparsity and more stable for re-sampling. Robust capped l_1 norm TWSVM (Wang et al. 2019a) reduced the effect of outliers which resulted in better performance. To reduce the overfitting issues, capped l_1 norm twin bounded support vector machines (Ma et al. 2020) was proposed. Efficient and robust l_1 norm TWSVM (Yan et al. 2019) used l_1 norm to maximise the ratio of interclass scatter to intraclass scatter. The authors used iterative procedure to get the optimal separating hyperplanes. Adaptive capped $L_{\theta,\epsilon}$-loss based TWSVM (Ma et al. 2021) is a generalized TWSVM model wherein θ and ϵ parameters are optimized to meet the objectives of robust to noise and outliers.

3.5 Pinball loss twin support vector machine

Xu et al. (2016c) proposed a novel TPMSVM with the pinball loss (Pin-TPMSVM). The authors introduced pinball loss function that is based on maximizing quantile distances between the two classes instead of hinge loss function which maximizes the distance between the closest samples of the two classes. This leads to noise insensitivity and as well as re-sampling stability. Pin-TPMSVM has excellent capability to handle noise which makes the model a more robust classifier but little attention is given on sparsity due to which the testing time is high and there is instability in re-sampling. Xu et al. (2016d) also formulated a maximum margin and minimum volume hyperspheres with pinball loss (Pin-M^3HM).

Authors proposed this algorithm to enhance the generalization of twin hypersphere SVM (TH-SVM) for noise present in datasets. The algorithm classifies samples of two classes by a pair of hyper-spheres, each containing either majority or the minority class samples simultaneously maintaining maximum margin between the hyperplanes. Pin-M^3HM is fast and has better accuracy than TWSVM. Sharma and Rastogi (2018) proposed two models called ϵ-Pin-TWSVM and Flex-Pin-TWSVM. ϵ-Pin-TWSVM introduces ϵ parameter to reduce the impact of noise and to attain a sparse solution while Flex-Pin-TWSVM uses a self-optimized framework which makes this algorithm flexible to estimate the size of insensitive zone. It adapts to the structure of data. Pin-TPMSVM has impressive ability to handle noise but to reduce parameter tuning time, Yang et al. (2018) introduced a new solution approach for the TPMSVM in which one instance is considered at a time and it is solved analytically without solving optimization problem. It is fast, simple and flexible. Since pinball loss function is not differentiable at zero, hence, smooth pinball loss TWSVM (Li and Lv 2021) used smooth approximation function and solved the objective functions via Newton-Armijo method.

For imbalanced data classification Xu et al. (2018b) proposed a maximum margin twin spheres which uses pinball loss. This algorithm finds homocentric spheres so that the smaller one captures positive class samples and the larger one repel negative samples. It requires to solve a QPP and a Linear programming (LP) problem. This algorithm has good generalization performance and noise insensitivity, however, suffers due to large complexity. To reduce the complexity, bound estimation-based safe acceleration for maximum margin of twin spheres machine with pinball loss (Yuan and Xu 2021) was proposed. Sharma et al. (2019) proposed a Stochastic Quasi-Netwon method based TPSVM using Pinball Loss Function (SQN-PTWSVM) which can scale the training process to handle millions of data points while simultaneously deals with noise and re-sampling data issues. Twin neural networks (Pant et al. 2019) uses feature maps which allows better discrimination among classes. The twin neural network is also extended to multiclass problems wherein the number of neural networks trained is proportional to number of classes.

The aforementioned pinball loss TWSVM algorithms are based on TPMSVM and not on the original TWSVM algorithms. Tanveer et al. (2019b) introduced pinball loss to the original TWSVM, termed as general TWSVM with pinball loss (Pin-GTSVM). Pin-GTSVM (Tanveer et al. 2019b; Ganaie and Tanveer 2021) is less sensitive to the outliers and more stable algorithm for re-sampling as compared to the original TWSVM. To retain the sparsity of original TWSVM and Pin-GTSVM, Tanveer et al. (2019c) proposed a novel sparse pinball TWSVM (SP-TWSVM) which uses ϵ-insensitive zone pinball loss function. SP-TWSVM has better classification accuracy than TWSVM and Sparse Pin SVM and it is insensitive to outliers, retains sparseness and suitable for re-sampling. Recently, Tanveer et al. (2019a) proposed improved sparse pinball TWSVM (ISPTWSVM) by adding a regularization term to the objective function of SP-TWSVM. ISPTWSVM implements SRM principle and also the matrices appear in the dual formulation are positive definite which makes the proposed algorithm computationally less complex and it also achieves better accuracy than other baseline algorithms. Most of the twin SVM based models involve matrix inversion operations which limits their applicability to large scale data. Hence, large scale pinball TWSVM (Tanveer et al. 2021g) uses pinball loss function to reduce the issues of feature noise and reformulated the Lagrangian in a manner that matrix inversions are no longer involved in the optimization problems. This scaled the model to large scale data.

3.6 Twin support vector machine for universum data and imbalanced datasets

Qi et al. (2012a) formulated TWSVM for universum data classification (U-TWSVM). Universum data is defined as the data samples which does not belong to any given class. This algorithm utilizes the universum data to enhance TWSVM classification accuracy. U-TWSVM employs new data points to the either class based on its proximity to the hyperplanes. Experimental results have shown that U-TWSVM has better accuracy than TWSVM. In order to construct a robust classifier by including the prior information embedded in the universum samples Qi et al. (2014) formulated nonparallel SVM (U-NSVM) which maximizes the two margins related to the two nearest adjacent classes.

As many real life problems consists of datasets which are imbalanced in nature i.e. classes do not contain same number of data samples, due to which many machine learning algorithms cannot be implemented. Thus, to enhance the performance of TWSVM while dealing with imbalanced datasets, Shao et al. (2014b) proposed weighted Lagrangian TWSVM for imbalanced data classification. Authors introduced a quadratic loss function to the formulation of TWSVM which enabled faster training of data points. Also, this is more robust to outliers and has better computational speed and accuracy than TWSVM. Tomar et al. (2014c) proposed weighted LS-TWSVM for imbalanced datasets by adjusting the classifier and assigning distinct weight to training samples. The results of the experiment shows that its accuracy is more than SVM, TWSVM, and LS-TWSVM. Furthermore, in 2015, Xu et al. (2016a) formulated least squares TWSVM (ULS-TWSVM) which exploits the universum data. It introduces a regularization term thus implements SRM principle and is computationally less complex as it requires to solve linear equations in place of QPPs in TWSVM. Cao and Shen (2016) proposed combining re-sampling with TWSVM for imbalanced data classification. In this algorithm, authors presented a hybrid re-sampling technique which utilizes the one side selection (OSS) algorithm and synthetic minority oversampling technique (SMOTE) algorithm to balance the training data. This is combined with TWSVM for classification purpose. However, SMOTE algorithm based on K-nearest neighbors can often result in over-fitting. Robust rescaled hinge loss TWSVM (Huang et al. 2019) and density weighted TWSVM (Hazarika and Gupta 2021) for imbalanced datasets resulted in improved performance.

Encouraged by the performance of ν-TWSVM (Peng 2010c) and U-TWSVM (Qi et al. 2012a), Xu et al. (2016b) proposed ν-TWSVM for universum data classification ($U\nu$-TWSVM). It is more flexible in using the prior knowledge from universum data to improve the generalization ability. It uses two hinge loss functions so that data can remain in a nonparallel insensitive loss tube. Experimental results have shown that $U\nu$-TWSVM has better accuracy and also costs lower running time than other baseline algorithms.

Xu (2017) proposed maximum margin twin spheres SVM algorithm (MMTSSVM) to overcome many limitations of TWSVM while dealing with imbalanced data. This algorithm seeks to determine two homocentric spheres so that smaller one captures as many positive class samples and the larger one repel negative samples simultaneously increases the margin between the spheres.

Richhariya et al. (2018) added a regularization term into the optimization problem of universum TWSVM to make matrices non-singular and improve the generalization performance. The proposed algorithm is called improved universum TWSVM. It has better generalization and training time than USVM and UTWSVM. Richhariya and Tanveer (2019) proposed a fuzzy universum SVM (FUSVM) based on information entropy. Universum support vector machines are more efficient than other SVM methods as it includes prior information of samples. But, it is not effective in case of noise-corrupted datasets. FUSVM introduces

weights such that it gives fewer weights to the outlier universum points. Further, the authors proposed a fuzzy universum TWSVM (FUTWSVM). Both the algorithms have better generalization performance than SVM, USVM, TWSVM, and UTWSVM. Recently, Richhariya and Tanveer (2020b) proposed a reduced universum twin support vector machine for class imbalance learning (RUTWSVM-CIL) with the idea of prior information about the data distribution. RUTWSVM-CIL used reduced kernel matrix, and thus applicable for the large sized imbalanced datasets. Numerical experiments on several imbalanced benchmark datasets showed the applicability of RUTWSVM-CIL. Richhariya and Tanveer (2020a) also proposed a novel parametric model for universum based twin support vector machine and extended its application for the classification of Alzheimer's disease data. Fuzzy universum least squares TWSVM (Richhariya and Tanveer 2021a; Borah et al. 2018) and fuzzy least squares TWSVM (Ganaie et al. 2021b; Gupta and Richhariya 2018) solved a linear system of equations instead of solving the QPPs, hence, the model is fast and requires no external toolbox for solving the optimization problems. Robust TBSVM (Borah and Gupta 2021) is proposed to make the model more robust to noise in imbalance datasets.

3.7 Fuzzy twin support vector machines

Chen and Wu (2018) proposed a novel fuzzy TWSVM (NFTWSVM) which takes care of the problem of dealing with one class playing major role in classification, by assigning fuzzy membership to different samples. The proposed NFTWSVM includes fuzzy neural networks and provides more generalized results. Richhariya and Tanveer (2019) proposed a fuzzy universum SVM (FUSVM) based on information entropy. This algorithm is also discussed in Sect. 3.6. Richhariya and Tanveer (2018a) also proposed a robust fuzzy LS-TWSVM (RFLS-TWSVM-CIL) to boost the performance of TWSVM on imbalanced datasets. The optimization problem in this algorithm is strictly convex as it uses 2-norm of the slack variables. Authors further proposed a fuzzy membership function to deal with imbalanced problems as it provides weights to samples and includes data imbalance ratio. RFLS-TWSVM-CIL obtains better generalization and is computationally more efficient as it requires to solve two linear equations. However, RFLSTWSVM-CIL algorithm implements empirical risk minimization principle which requires inverse of matrices to be positive semi-definite. In order to improvise, Ganaie et al. (2020b) introduced a regularization term to the primal formulation of RFLSTWSVM-CIL. The proposed algorithm is regularized robust fuzzy least squares (RRFLSTWSVM) which doesn't require the extra assumption of inverse of matrices to be positive semi-definite and performs better than RFLSTWSVM-CIL.

Khemchandani et al. (2018a) formulated fuzzy LS version of TSVC in which each data point has a fuzzy membership value and is allotted to different clusters. This algorithm is also discussed in Section 3.11. Chen et al. (2020a) proposed entropy-based fuzzy least squares TWSVM which considers the fuzzy membership for each data point basis entropy. Rezvani et al. (2019) formulated intuitionistic fuzzy TWSVM (IFTWSVM) which is modified version of fuzzy TWSVM as it considers the position of input data in the feature space and calculate adequate fuzzy membership and also reduces the effect of noise. Zhang et al. (2019) proposed fuzzy TWSVM which assigns fuzzy membership based on intra-class hyperplane and sample affinity. Experimental results showed that this algorithm has better accuracy than TWSVM and classic fuzzy TWSVM. Gupta et al. (2019b) proposed entropy-based fuzzy twin support vector machine (EFTWSVM-CIL) which is an efficient algorithm for imbalanced datasets as it assigns fuzzy membership values based on entropy of the samples. Chen et al. (2020a) proposed entropy-based fuzzy least squares twin support vector machine (EFLSTWSVM)

which is an improvised version of EFTWSVM-CIL by formulating the QPPs in least squares sense and retains the superior characteristics of LSTWSVM. Experimental results shown that this algorithm performed better than other baseline fuzzy TWSVM algorithms (Table 3).

3.8 Some other improvements of twin support vector machines

Kumar and Gopal (2008) enhanced TWSVM using smoothing techniques. Authors proposed to transform the primal problems of TWSVM into smooth unconstrained minimization problems and used Newton−Armijo algorithm for optimization. Smooth TWSVM (STWSVM) has comparable generalization to TWSVM but is significantly a faster algorithm. Khemchandani and Chandra (2009) proposed iterative alternating algorithm to make kernels learn efficiently. Zhang (2009) proposed boosting based TWSVM for clustered microcalcifications detection. Results have shown that this method improved detection accuracy. Bagging algorithm was combined with boosting to solve the unstable problem of TWSVM. This method is called BB-TWSVM. Twin support vector machine in linear programs and robust TWSVM (Qi et al. 2013b) have also been proposed for the classification problems (Li and Tian 2014).

Shao et al. (2010) proposed a bi-level programming method to multiple instance classification, called MI-TWSVM. Multiple instance learning (MIL) is a supervised technique wherein the training data points consist of labeled bags and these bags include multiple unlabeled instances. For binary classification, a bag is a negative sample if all the instances in it are negative and it is labeled a positive sample if at least one instance is positive. Similar to TWSVM, the proposed algorithm seeks two hyperplanes thereby the positive hyperplane is closest to the positive instances and as distant as possible from the negative instances and vice-versa. Ghorai et al. (2010) proposed a unity norm TWSVM classifier (UNTWSVM) which includes unity norm equality constraints and a quadratic loss function in the objective function of TWSVM. This algorithm can be solved by sequential quadratic optimization method. UNTWSVM has more computational cost than TWSVM especially for large datasets because of the nonlinear formulation. Generalized TWSVM (Moosaei et al. 2021) proposed two models, in the first model the authors used the l_1 and l_∞ norm in the optimization problems and in second model the convex, piecewise quadratic objective function is solved via generalized Newton method.

In order to improve TWSVM's generalization ability and to decrease support vectors (SVs), Peng (2010c) introduced a variable p and parameter ν in TWSVM and proposed ν-TWSVM. An improved ν-TWSVM (Xu and Guo 2014a) has also been proposed for the classification problems. The parameter ν controls the trade-off between the SV and marginal error while adaptive p overcomes the limitations of constraints being over-restricted in TWSVM and thus reduces the number of support vectors. But, it gives the same penalties to each misclassified sample and results in over-fitting. Thus, Xu et al. (2012) introduced the rough set theory into ν-TWSVM to remediate this limitation. This algorithm is more effective in avoiding over-fitting as it gives different penalties to different points depending upon location. It also has better generalization than ν-TWSVM.

Although rough ν-TWSVM (Xu et al. 2012) performs better than ν-TWSVM, it provides different weights to negative samples and same weights to all the positive samples. Thus, Xu et al. (2014b) formulated K nearest neighbor (KNN) weighted rough ν-TWSVM which gives different penalties to the samples of both the classes. It has better accuracy and lower computational cost than ν-TWSVM and Rough ν-TWSVM. In another attempt to enhance the results of ν-TWSVM, Khemchandani et al. (2016) formulated two models for binary classification: $I\nu$TWSVM and $I\nu$TWSVM Fast. Both the algorithms are faster than ν-TWSVM as

Table 3 Classification accuracy on non-linear kernel (Chen et al. 2020a)

Dataset (Train size, Test size)	TWSVM (Khemchandani and Chandra 2007)	LSTWSVM (Kumar and Gopal 2009)	EFSVM (Fan et al. 2017)	FTWSVM (Li and Ma 2013)	EFTWSVM-CIL (Gupta et al. 2019b)	EFLSTWSVM (Chen et al. 2020a)
Australian (690×14)	87.83	87.54	87.10	87.39	87.68	87.97
Bupa-Liver (345×6)	75.07	75.94	73.91	74.78	74.49	75.65
House-Votes (435×16)	96.32	96.55	95.86	96.32	96.09	96.55
Heart-c (303×13)	85.14	84.50	84.84	83.83	86.13	85.82
Heart-Statlog (270×13)	85.56	85.19	85.19	85.86	86.30	85.93
Ionosphere (351×34)	94.32	94.30	95.73	94.60	94.60	96.58
Musk (476×166)	95.79	94.96	94.54	96.43	95.58	96.63
PimaIndian (768×8)	78.52	78.30	77.74	78.13	78.64	79.16
Sonar (208×60)	88.95	91.39	90.88	89.94	91.86	91.81
Spect (267×44)	82.42	83.89	82.78	83.15	83.52	84.65
Wpbc (198×34)	83.86	83.36	82.83	83.33	84.36	84.88

$I\nu$TWSVM solves one small QPP and an unconstrained minimization problem (UMP) while $I\nu$TWSVM Fast solves one unimodal function and one UMP. To overcome the disadvantage of TWSVM being sensitive to outliers, Xie and Sun (2015) implemented class centroid and proposed multitask centroid TWSVM for multitask problems.

For handling large scale datasets, Shao et al. (2015) introduce weighted linear loss in TWSVM and proposed weighted linear loss TWSVM (WL-TWSVM) to classify large scale datasets. The proposed algorithm only deals with linear equations which increases the training speed and also has better generalization than TWSVM. To further fasten up the training process, Sharma and Rastogi (2018) and Wang et al. (2018), recently proposed stochastic conjugate gradient method based twin support vector machine (SCH-TWSVM). The resulting model showed effective performance on binary activity classification problem.

Shao et al. (2011) introduced the regularization term in TWSVM to embody SRM principle and formulated a new algorithm called twin bounded support vector machines (TBSVM). To further speed up training, SOR technique is used. Numerical results on various datasets show that TBSVM is faster and has better generalization ability than TWSVM. Ye et al. (2011) formulated localized TWSVM via convex minimization (LC-TWSVM) which effectively constructed two nearest-neighbor graphs in the original input space to reduce the space complexity of TWSVM. Shao and Deng (2013) in 2012 considered the unity norm constraints and added a regularization term to minimize structural risk in TWSVM and proposed margin-based TWSVM with unity norm hyperplanes (UNHMTWSVM). The proposed algorithm is fast, has better generalization and accurate compared to TWSVM. Lagrangian TBSVM with L_2 norm (Gupta and Gupta 2019) replaced L_1 norm of the slack variables with L_2 norm to improve the performance.

To include statistical data information in TWSVM, Peng and Xu (2012) formulated twin Mahalanobis distance-based SVM (TMSVM) that uses each classes covariance to determine hyperplanes. It has better training speed and generalization than TWSVM. Shao et al. (2012b) introduced a probability based TWSVM model (P-TWSVM) which is more accurate than TWSVM. In 2012, Shao and Deng (2012) formulated a coordinate descent based TWSVM to increase the efficiency of TWSVM by introducing a regularization term. Further, to solve the dual problems, the authors proposed a coordinate descent method which reduces the training time even in case of large datasets as it deals with single data point at once.

To include the underlying correlation information between data points in TWSVM, Ye et al. (2012b) formulated a novel nonparallel plane classifier, called Weighted TWSVM with Local Information (WL-TWSVM). A major limitation of this algorithm is that it doesn't work for large-scale problems as it finds the K-nearest neighbors for all the samples. Based on TWSVM, Peng and Xu (2013b) formulated a twin-hypersphere SVM (TH-SVM). Unlike TWSVM, it determines two hyperspheres and avoids the matrix inversions in its dual formulations.

To incorporate the structural data information in TWSVM, Qi et al. (2013a) formulated structural-TWSVM which is superior in training time as well as accuracy to that of TWSVM. However, it still ignores the importance of different samples within each cluster. To overcome this drawback, Pan et al. (2015) uses K-nearest neighbor based structural TWSVM which provides different penalties to the samples of different classes. However, it suffers from overfitting due to empirical risk minimisation. To reduce the overfitting issues, efficient KNN weighted TWSVM (Xie and Xu 2019) introduced the regularisation term to avoid the issues of overfitting. This also ensured the minimization of structural risk which results in better generalization.

Shao et al. (2014a) formulated a different nonparallel hyperplane SVM in which hyperplanes are determined by clustering the samples based on similarity between the two classes.

This algorithm has better or comparable accuracy with low computational time. To optimally select the parameters in TWSVM, Ding et al. (2016) proposed TWSVM build on Fruit Fly Optimization Algorithm (FOA-TWSVM). It can optimally select the parameters in less time and with better accuracy. In 2017, Pan et al. (2017) proposed safe screening rules to make TWSVM efficient for large-scale classification problems. The safe screening rules reduce the scale by eliminating training samples and giving same solution as the original problem. Experimental results show that safe screening rules can greatly reduce the computational time while giving the same solutions as original ones. Yang and Xu (2018) proposed safe sample screening rule (SSSR) for Laplacian twin parametric-margin SVM (LTPSVM) to address the problems while handling large-scale problems. Pang and Xu (2019) proposed safe screening rule for Weighted TWSVM with local information (WLTWSVM) to implement the algorithm for large-scale datasets. Experimental results demonstrate the effectiveness of SSSR for WLTWSVM as it performes better than SVM and TWSVM with significantly less computational time. Recently, Zhao et al. (2019) proposed an efficient non-parallel hyperplane Universum support vector machine (U-NHSVM) for classification problems. U-NHSVM is flexible to exploit the prior information in universum.

Tanveer (2015c) proposed unconstrained minimization problem (UMP) formulation of Linear programming TWSVM to enhance robustness and sparsity in TWSVM. Tanveer (2015a) also proposed an implicit Lagrangian TWSVM which is solved by using finite Newton method. Tanveer and Shubham (2017a) proposed smooth TWSVM via UMP which increases the generalization ability and training speed of TWSVM.

Multi-label learning deals with data having multiple labels and has gained a lot of attention recently. Chen et al. (2016) proposed multi-label TWSVM (MLTWSVM) that exploits multi-label information from instances. Azad-Manjiri et al. (2020) proposed structural twin support vector machine for multi-label learning (ML-STWSVM) which embeds the prior structural information of data into the optimization function of MLTWSVM based on the same clustering technology of S-TWSVM. This algorithm achieved better performance compared to other baseline multi-label learning algorithms.

Rastogi et al. (2018b) proposed a new loss function termed as (ϵ_1, ϵ_2)-insensitive zone pinball loss function which generalizes other existing loss functions e.g. pinball loss, hinge loss. The resulting model takes care of noise insensitivity, instability to re-sampling and scatterdness present in the datasets. Tanveer et al. (2019) presented rigorous comparison of 187 classifiers which includes 8 variants of TWSVM and exhaustive evaluation of these classifiers was performed on 90 UCI benchmark datasets. Results have shown that RELS TSVM achieved highest performance than all other classifiers for binary classification task. Thus, RELSTSVM is the best TWSVM based classifier. For more details, one can refer to Tanveer et al. (2019). Ganaie and Tanveer (2020) proposed a novel classification approach using pre-trained functional link to enhance the feature space. Authors performed the classification task by LSTWSVM on the enhanced features and validated the performance on various datasets. Some other recent research on TWSVM include (Ganaie et al. 2020a) where authors proposed a novel way for generating oblique decision trees. The classification of training samples is done based on the Bhattachrayya distance with randomly selected feature subset and then hyperplanes are generated using TBSVM. The major advantage of the proposed model is that there is no need for any extra regularization as matrices are positive definite. The ensemble (Ganaie et al. 2021a) based models of twin SVM based models was proposed in Tanveer et al. (2021a).

3.9 Twin support vector machine for multi-class classification

Earlier, TWSVM was only implemented to solve binary class problems, however, the majority of problems in the real-world applications are generally based on multicategory classification. Thus, Xu et al. (2013) formulated Twin-KSVC. It implements "1-versus-1-rest" form to provide ternary output $(-1, 0, 1)$. This algorithm requires to solve two smaller-sized QPPs. It has better accuracy than 1-versus-rest TWSVM but losses sparsity. Yang et al. (2013) proposed multiple birth SVM (MBSVM). It is computationally better than TWSVM even when number of classes are large.

Xie et al. (2013) extended TWSVM application for multi-class problems and proposed one-versus-all TWSVM (OVA-TWSVM) which solve k-category problem using one-versus-all (OVA) approach to develop k TWSVM. To strengthen the performance of multi-TWSVM, Shao et al. (2013a) formulated a separating decision tree TWSVM (DTTWSVM). The basic idea of DTTWSVM is to embody the best separating principle rule to create a binary tree and then built binary TWSVM model on each node. Experimental results have shown that DTTWSVM has low computational complexity and better generalization.

Xu and Guo (2014b) formulated twin hyper-sphere multi-class SVM (THKSVM) which employs the "rest-versus-1" structure instead of "1-versus-rest" structure in TWSVM. In order to find hyperspheres, it constructs k (no. of classes) classifiers and finds k centers and k radiuses for each hypersphere. Each hypersphere covers maximum points of K_1 classes and is as distant as possible from the rest classes. It has fast computational speed as compared to Twin-KSVC and also inverse operation of matrices are not required while solving dual QPPs of THKSVM. Thus, it performs better on large datasets and has better accuracy than "1-versus-rest" TWSVM but lower than Twin-KSVC. To enhance the performance of Twin-KSVC, Nasiri et al. (2015) formulated LS version of Twin-KSVC that works similar to Twin-KSVC but solves linear system of equations rather than pair of QPPs in Twin-KSVC. The proposed algorithm is fast, simple and has better accuracy and lower training time than Twin-KSVC. To enhance the performance of Twin-KSVC and to include local information of samples, Xu (2016) proposed K-nearest neighbor-based weighted multi-class TWSVM which uses information from within class and applies a weight in the objective function. This algorithm has low computational cost and better accuracy than Twin-KSVC.

Based on the multi-class extension of the binary LS-TWSVM, weighted multi-class LST-WSVM algorithm (Tomar and Agarwal 2015b) and regularized least squares twin SVM (Ali et al. 2022) have been proposed for multi-class imbalanced data. To control sensitivity of classifier, weight setting is employed in loss function for determining each hyperplane. Experimental results have shown the superiority and feasibility of the proposed algorithm for multi-class imbalanced problems. The authors (Tomar and Agarwal 2015c) extended LS-TWSVM for multi-class classification and compared various concepts of a multi-class classifier like "One-versus-All", "One-versus-One", "All-versus-One" and "Directed Acyclic Graph (DAG)". DAG MLS-TWSVM performance is superior and has high computational efficiency.

Yang et al. (2016) formulated least squares recursive projection TWSVM. For each class, it determines k projection axes and needs to solve linear equations. It has similar performance as MP-TWSVM. Based on P-TWSVM, Li et al. (2016) proposed multiple recursive projection TWSVM (Multi-P-TWSVM) which solves k QPPs in order to determine k-projection axes (for k classes). Authors introduced regularization term and recursive procedure which increases the generalization but this algorithm is complex when more orthogonal projection axes are generated.

Ding et al. (2017) review various multi-class algorithms as per their structures: "one-versus-rest", "one-versus-one", "binary tree", "one-versus-one-versus-rest", "all-versus-one" and "direct acyclic graph" based multi-class TWSVM. All these multi-class TWSVMs have some advantages and disadvantages. In general, one-versus-one TWSVMs have higher performance. Ai et al. (2018) proposed a multi-class classification weighted least squares TSVH using local density information in order to improve the performance of LSTSVH. Authors introduced local density information into LSTSVH to provide weight for each data point in order to avoid noise sensitivity. Pang et al. (2018) proposed K-nearest neighbor-based weighted multi-class TWSVM (KMTWSVM) that incorporated "1-versus-1-versus-rest" strategy for multi-class classification and also takes into account the distribution of all instances. However, it is computationally extensive especially for large-scale problems. Thus, authors in Pang et al. (2018) also proposed safe instance reduction rule (SIR-KMTWSVM) to reduce its computational time. de Lima et al. (2018) proposed improvements on least squares twin multi-class classification SVM which is "one-versus-one-versus-rest" strategy and generated ternary output. The proposed algorithm only needs to deal with linear equations. Numerical results demonstrate that it achieves better classification accuracy than Twin-KSVC (Xu et al. 2013) and LSTKSVC (Nasiri et al. 2015). Qiang et al. (2020) proposed improvement on LSTKSVC by proposing robust weighted linear loss twin multi-class support vector machine (WLT-KSVC) which takes care of the two drawbacks of LSTKSVC; sensitive to outliers and misclassifying some rest class samples due to the use of quadratic loss. Experiments on the UCI and NDC datasets showed promising results of this algorithm but its training accuracy significantly decreases as the number of classes increases. Li et al. (2019) proposed a nonparallel support vector machine (NSVM) for multiclass classification problem. Numerical experiments on several benchmark datasets clearly show the advantage of NSVM. In 2020, Tanveer et al. (2021d) proposed a fast and improved version of KWMTWSVM (Xu and Wang 2014) called least square K-nearest neighbor weighted multi-class TWSVM (LS-KWMTWSVM). Numerical experiments on various KEEL imbalance datasets showed high accuracy and low computational time for the proposed LS-KWMTWSVM as compared to other baseline algorithms. A multiclass nonparallel parametric margin SVM (Du et al. 2021) has also been proposed for multiclass classification. Kernel free least squares TWSVM (Gao et al. 2021) via special fourth order polynomial surface resulted in improved performance in multiclass problems.

3.10 Twin support vector machine for semi-supervised learning

Semi-supervised learning (SSL) techniques have achieved extensive attention from many researchers in the last few years due to its promising applications in machine learning and data mining. In many real-world challenges, labeled data is not easily available and thus it deteriorates the performance of supervised learning algorithms due to insufficient supervised information. SSL overcomes this limitation and uses both unlabeled and labeled data.

Qi et al. (2012b) formulated a novel Laplacian – TWSVM for the semi-supervised classification problem. This algorithm assumes that data points lie in low dimensional space and uses the geometric information of the unlabeled data points. Similar to TWSVM it solves a pair of QPPs with inverse matrix operations. Results have shown that Lap-TWSVM has better flexibility, prediction accuracy and generalization performance than conventional TWSVM. This algorithm has excellent performance for semi-supervised classification problems but due to QPPs and inverse matrix operations, its computational cost is high. Thus, to increase the training speed of Lap-TWSVM, Chen et al. (2014a) formulated LS version of Lap-

TWSVM (Lap-LS-TWSVM). Unlike Lap-TWSVM, it needs to deal with linear equations and is done using conjugate gradient algorithm. It has less training time than Lap-TWSVM. Another algorithm was formulated by Chen et al. (2014b) to improve the performance of Lap-TWSVM by transforming the QPPs to UMPs in primal space and smoothing method is used which is effectively solved by Newton-Armijo algorithm. It achieved comparable accuracy as Lap-TWSVM with less computational time. Khemchandani and Pal (2016b) extended this to multi-class classification and formulated Lap-LS-TWSVM which evaluates training samples to "1-versus-1-versus-rest" and provides ternary output $(-1, 0, +1)$. It has better accuracy and less training time than Lap-TWSVM.

Based on the similar idea of Lap-TWSVM, Yang and Xu (2016) proposed an extension of traditional TPMSVM, which makes use of graph Laplacian and creates a connection graph of the training data points whose solution can be obtained by solving two SVM-type QPPs. Experiments reveal that LTPMSVM has higher classification accuracy than SVM, TPMSVM, Lap-TWSVM, and TWSVM.

Khemchandani and Pal (2017), blended the Laplacian—TWSVM and Decision Tree–TWSVM classifier and formulated a tree based classifier for semi-supervised multi-class classification. Extensive experiments on color images shows the feasibility of the model. Another interesting approach in this direction in 2019, has been suggested by Rastogi and Sharma (2019) called as Fast Laplacian TWSVM for Active Learning $(FLap-TWSVM_{AL})$ where the authors proposed to identify the most informative and representative training points whose labels are queried for domain experts for annotations. Once the corresponding labels are acquired, this limited labeled and unlabeled data are used to train a fast model that involves solving a QPP and an unconstrained minimization problem to seek the classification hyperplanes.

3.11 Twin support vector machine for clustering

Wang et al. (2015b) formulated twin support vector clustering algorithm (TSVC) build upon TWSVM. It divides the data samples into k clusters such that the data samples are around k cluster center points. It exploits the information from clusters (both between and within clusters) and the center planes are determined by solving a series of QPP. Authors in Wang et al. (2015b) also proposed a nearest neighbor graph (NNG)-based initialization to make the model more stable and efficient. Improved TSVC (Moezzi et al. 2019) decomposed the multiclass clustering problem into multiple two cluster problems.

Khemchandani et al. (2018a) formulated fuzzy least squares version of TSVC in which each data point has a fuzzy membership value and is allotted to different clusters. The proposed algorithm solves primal problems instead of dual problems in TSVC. Experimental results show that the proposed algorithm obtains comparable clustering accuracy to that of TSVC but has less training time.

Khemchandani and Pal (2016a) replaced the hinge loss function of TSVC with weighted linear loss function and introduced a regularization term in TSVC which needs to solve linear equations and also implements the structural risk component. The proposed algorithm called weighted linear loss TSVC achieves higher accuracy than TSVC. The loss function proposed in Khemchandani and Pal (2016a) is not continuous and therefore, authors in Rastogi and Pal (2019) modified the weighted linear loss function to form a continuous loss function and implemented TSVC and WLL-TSVC in semi-supervised framework and also introduced a fuzzy extension of semi-supervised WLL-TSVC to make a robust algorithm which is less sensitive to noise. Experimental results have demonstrated that the proposed algorithm achieved

better clustering accuracy and less computational time than TSVC and WLL-TSVC. Tree-based localized fuzzy twin support vector clustering (Tree-TSVC) was proposed by Rastogi and Saigal (2017). Tree-TSVC is a novel clustering algorithm that builds the cluster that represents a node on a binary tree, where each node comprises of proposed TWSVM based classifier. Due to the tree structure and the formulation that leads to solving a series of systems of linear equations, Tree-TSVC model is efficient in terms of training time and accuracy.

TSVC uses squared L_2-norm distance which leads to noise sensitivity. Also, each cluster plane learning iteration requires to solve a QPP which makes it computationally complex. To address these issues in TSVC, Ye et al. (2018) used L_1 norm distance and proposed L_1-norm distance minimization-based robust TSVC which only deals with linear equations instead of series of QPPs. Ye et al. (2018) further proposed RTSVC and Fast RTSVC to speed up the computation of TSVC in nonlinear case. Numerical experiments shown that this model has higher accuracy as compared to other k-clustering methods and has less computational time than TSVC. Zhen et al. (2019c) proposed ramp-based TSVC by introducing the ramp cost function in TSVC. The solution of the proposed algorithm is obtained by solving non-convex programming problem using an alternating iteration algorithm. Bai et al. (2019) introduced regularization in clustering and proposed large margin TSVC. Bai et al. (2019) proposed a fast least squares TSVC with uniform coding output. The algorithm achieves better performance than TSVC and other plane-clustering methods. Wang et al. (2019d) proposed a general model for plane-based clustering which includes various extensions of k-plane clustering (kPC). It optimizes the problem by minimizing the total loss which is derived from both within-cluster and between-cluster. It can capture data distribution accurately. Richhariya et al. (2020a) proposed projection based least square TSVC (LSPTSVC) and employed the concave-convex procedure (CCCP) to solve the optimization problem. LSPTSVC only needs to solve a set of linear equations and thus leads to significantly less computational time. TSVC uses hinge loss function, hence suffers from the issues of noise and has low sampling stability. To overcome these issues, pinball loss twin support vector clustering (Tanveer et al. 2021b) used pinball loss function to penalize the samples.However, it suffers from the issues of overfitting as it minimises the empirical risk. Hence, pinball loss twin bounded support vector clustering (Tanveer et al. 2021e) introduced the regularisation term to minimise the structural risk and avoids the issues of overfitting. Introduction of pinball loss function leads to loss of sparsity in the model. Hence, sparse pinball loss TSVC (Tanveer et al. 2021c, f) used ϵ-insensitive pinball loss function to make the model sparse. It implements the empirical risk minimisation principle and thus suffers due to overfitting. To minimise the overfitting issues, sparse pinball twin bounded support vector clustering (Tanveer et al. 2021e) used the regularisation term to minimise the structural risk and hence, avoid the issues of overfitting.

4 Applications of twin support vector classification

TWSVM has been implemented to solve many real-life classification challenges and has shown promising results in some applications (Fig. 2).

4.1 Biomedical applications

TWSVM has been applied to detect breast cancer in early stages by detecting a mass in digital mammograms (Si and Jing 2009). A hybrid feature selection based approach has also been implemented for detecting breast cancer, Hepatitis, and Diabetes (Tomar and Agarwal 2015a).
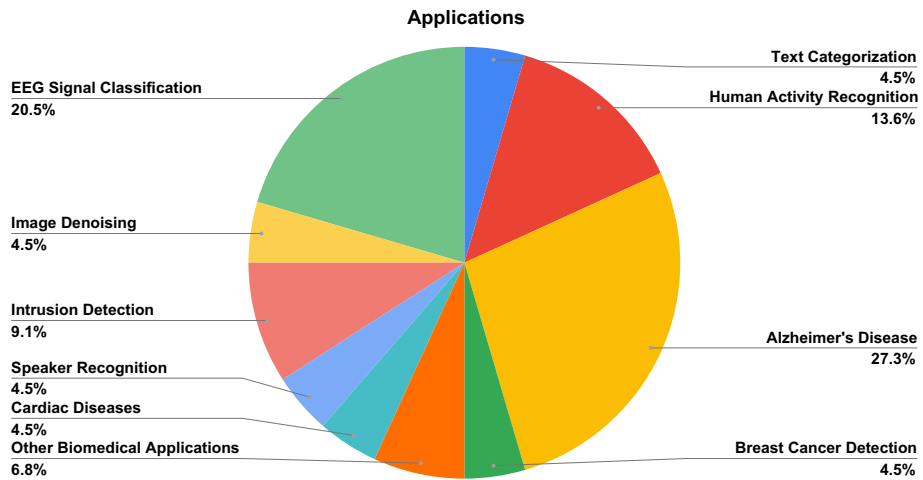
Applications

Fig. 2 Applications of twin support vector classification

Wang et al. (2016b) proposed TWSVM in combination with dual-tree complex wavelet transform (DTCWT) method for Pathological Brain Detection. TWSVM is also implemented for detecting cardiac diseases, one such application was proposed by Houssein et al. Houssein et al. (2018), heartbeats were detected using Swarm-TWSVM and this algorithm achieved better accuracy than TWSVM. Refahi et al. (2018) used LSTWSVM and DAG LS-TWSVM classifiers for predicting arrhythmia heart disease. Chandra and Bedi (2018) proposed linear norm fuzzy based TWSVM for color based classification of human skin which achieved better accuracy than other conventional methods. Xu et al. (2015) proposed semi-supervised TWSVM for detection of Acute-on-chronic liver failure (ACLF). Also, many researchers have applied TWSVM to detect Alzheimer's disease in its early stages (Tanveer et al. 2020; Zhang and Wang 2015; Wang et al. 2016c; Alam et al. 2017; Tomar and Agarwal 2014; Wang et al. 2016b; Tomar et al. 2014b; Wang et al. 2016a; Tomar et al. 2014a; Wang et al. 2015a).

4.2 Alzheimer's disease prediction

Richhariya and Tanveer (2021b) proposed an angle based universum least squares TWSVM (AULSTWSVM) which performed with 95% accuracy for detecting Alzheimer's disease (AD). Richhariya et al. (2020b) also proposed universum support vector machine based recursive feature elimination (USVM-RFE) for detecting AD. Khan et al. (2021) proposed an approach to improve the classification accuracy in mild cognitive impairment (MCI), normal control (NC), and AD subjects using structural magnetic resonance imaging (sMRI). Authors used FreeSurfer to process MRI data and derive cortical features which are used in TWSVM, LSTWSVM and RELSTSVM to detect AD. Sharma et al. (2022) proposed fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease.

4.3 Speaker recognition

Cong et al. (2008) formulated multi-class TWSVM for speaker recognition with feature extraction based on gaussian mixture models (GMMs). It gives better results than traditional

SVM. Yang and Wu (2009) proposed multi-TWSVM which find hyperplane for every class and takes constraints from other classes separately on the QPP. This algorithm performed better than many other algorithms for speaker recognition.

4.4 Text categorization

Kumar and Gopal (2009) formulated a least squares TWSVM and experiments have shown the validity of the model for text applications. Francis and Sreenath (2019) proposed manifold regularized TWSVM for text recognition and the proposed method achieved highest accuracy among SVM, LSTWSVM and other methods. Non-parallel SVM (Tian et al. 2014b) and efficient pinball TWSVM (Rastogi and Pal 2021) has also demonstrated better performance in text categorization.

4.5 Intrusion detection

It is a system which monitors or protects the network against any malicious activity. It has been a critical component for network security. Researchers (He and Zheng 2014; Ding et al. 2008; Mousavi et al. 2015; Nie et al. 2013) applied TWSVM for intrusion detection and results have shown that it achieves better accuracy than other intrusion detection algorithms.

4.6 Human activity recognition

Khemchandani and Sharma (2016) proposed least square TWSVM for human activity recognition which gives promising results even with the outliers. Nasiri et al. (2014) formulated energy-based LS-TWSVM algorithm. Khemchandani and Sharma (2017) also proposed robust parametric twin support vector machine which can effectively deal with the noise. Mozafari et al. (2011) used the Harris detector algorithm and applied LS-TWSVM for action recognition and achieved the highest accuracy than other state-of-the-art methods. Kumar and Rajagopal (2018) proposed Multi-class TWSVM for detecting human face happiness combined with Constrained Local Model. Kumar and Rajagopal (2019) also proposed semi-supervised multi TWSVM to predict human facial emotions with 13 minimal features that can detect six basic human emotions. Algorithm achieved highest accuracy and least computation time with minimal feature vectors.

4.7 Image denoising

Yang et al. (2014) proposed edge/texture-preserving image denoising based on TWSVM which is very effective to preserve the informative features such as edges and textures and better than other image denoising methods available. Shahdoosti and Hazavei (2018) proposed a ripplet formulation of the total variation method for denoising images. This algorithm attains promising results in improving the image quality in terms of both subjective and objective inspections.

4.8 Electroencephalogram (EEG)

Classification of electroencephalogram (EEG) for different mental activities has been an active research topic. Richhariya and Tanveer (2018b) proposed universum TWSVM which

is insensitive to outliers as it selects universum from the EEG datasets itself to generate universum data points which remove the effect of outliers. Soman (2015) used the classifiable metric to choose discriminative frequency bands and used the TWSVM to learn imbalanced datasets. Tanveer et al. (2018) proposed entropy based features in Flexible analytic wavelet transform (FAWT) framework and RELSTSVM Tanveer et al. (2016a) for classification to detect epileptic seizure EEG. Tanveer et al. (2018) used FAWT framework for classification of seizure and seizure-free EEG signals with Hjorth parameters as features and implemented TWSVM, LSTWSVM and RELSTSVM (Tanveer et al. 2016a) for classifying signals. Li et al. (2018) proposed LSTWSVM with a frequency band selection common spatial pattern algorithm for detecting motor imagery electroencephalography. This algorithm achieved faster training time compared to other SVM baseline algorithms. Some more researchers have applied TWSVM for EEG classification, biometric identification and other leak detection challenges (Kumar and Gupta 2021; Gupta et al. 2021; She et al. 2015; Li et al. 2018; Kostílek and Št'astnỳ 2012; Lang et al. 2017; Dalal et al. 2019)

4.9 Other applications

Cao et al. (2018) proposed improved TWSVM with multi-objective cuckoo search to predict software bugs. Authors employed TWSVM to predict defected modules and used cuckoo search to optimize TWSVM and this proposed method achieved better accuracy than other software defect prediction methods. Chu et al. (2018) proposed Multi-information TWSVM for detecting steel surface defects. The TWSVM models have also been used in image recognition or face recognition (Qi et al. 2013b; Peng and Xu 2012; Chen and Wu 2018; Peng and Xu 2013a) and facial expression recognition (Richhariya and Gupta 2019) and privacy preservation (Anand et al. 2019).

The pinball loss function is defined as follows:

$$L_\tau \left(X_y, y, g_y, \left(X_y \right) \right) = \begin{cases} e^T (0 - y g_y(X_y)), & (0 - y g_y(X_y)) \geq 0, \\ -\tau e^T (y g_y(X_y) - 0), & (0 - y g_y(X_y)) < 0. \end{cases} \tag{10}$$

where $g_y(x) = g(x; w_y, b_y) = w_y^T x + b_y = 0$, $\underline{D}(g_y(A, 0))$ denotes the intraclass distance which represents objective function while $\overline{D}(g_y(A), g_y(B))$ is interclass distance which corresponds to constraints, $D(g_y(x))$ is the perpendicular distance of point x from the hyperplane $g_y(x) = 0$.

Optimization framework of various TWSVM algorithms are discussed in Table 4 (Li et al. 2019).

Table 5 shows the differences in major TWSVM methods based on the SRM principle, sparsity, matrix inversion and noise insensitivity.

5 Basic theory of twin support vector regression

Peng (2010b) proposed an efficient twin support vector regression (TSVR) algorithm in line with TWSVM, called twin support vector regresson (TSVR). Like TWSVM, it also requires to solve two QPPs. It finds an end regressor that is the mean of ε-insensitive up and down bound functions.TSVR has less computational time than a standard SVR and has better generalization ability. The down- and up-bound functions for linear case are given below:

For any $x \in R^n$, the two hyperplanes are defined as follows:

Table 4 Optimization framework for TWSVM algorithms

Method	Regularization term	$\underline{D}(g_y(X_y), 0)$	$\overline{D}(g_y(X_j), g_y(X_y))$	$D(g_y(x))$	Distance
GEPSVM (Mangasarian and Wild 2006)	$\left\|\begin{bmatrix} w_y \\ b_y \end{bmatrix}\right\|^2$	$\left\|g_y(X_y)\right\|^2$	$\left\|g_y(X_j)\right\|^2$	$\frac{\|g_y(x)\|}{\|w_y\|}$	$\|\cdot\|$
TWSVM (Khemchandani and Chandra 2007)	0	$\left\|g_y(X_y)\right\|^2$	$e^T\left\|g_y(X_j) - e\right\|_+$	$\|g_y(x)\|$	$\|\cdot\|/(.)_+$
TBSVM (Shao et al. 2011)	$\left\|\begin{bmatrix} w_y \\ b_y \end{bmatrix}\right\|^2$	$\left\|g_y(X_y)\right\|^2$	$e^T\left\|g_y(X_j) - e\right\|_+$	$\frac{\|g_y(x)\|}{\|w_y\|}$	$\|\cdot\|/(.)_+$
LSTWSVM (Kumar and Gopal 2009)	0	$\left\|g_y(X_y)\right\|^2$	$\left\|g_y(X_j) - e\right\|^2$	$\|g_y(x)\|$	$\|\cdot\|$
Pin-TWSVM (Xu et al. 2016c)	$\left\|w_y\right\|^2$	$L_\tau(X_y, y, g_y, (X_y))$	$g_y(X_j)_1$	$\frac{\|g_y(x)\|}{\|w_y\|}$	$\|\cdot\|/L_T(.)/\|\cdot\|_1$
Pin-GTSVM (Tanveer et al. 2019b)	$\left\|w_y\right\|^2$	$\left\|g_y(X_y)\right\|^2$	$L_\tau(X_y, y, g_y, (X_y))$	$\frac{\|g_y(x)\|}{\|w_y\|}$	$\|\cdot\|/L_T(.)/\|\cdot\|_1$
NPSVM (Tian et al. 2014a)	$\left\|w_y\right\|^2$	$\left\|g_y(X_y) - \epsilon\right\|_1$	$e^T(g_y(X_j) - e)_+$	$\|g_y(x)\|$	$\|\cdot\|/\|\cdot\|_1/(.)_+$
Par-TWSVM (Peng 2011a)	$\left\|w_y\right\|^2$	$e^T g_y(X_j)$	$\left\|g_y(X_y)\right\|_+$	$\|g_y(x)\|$	$\|\cdot\|$

Table 5 Properties of different TWSVM algorithms

Models\characteristics	SRM	Sparsity	Matrix inversion	Insensitive to noise
TWSVM (Khemchandani and Chandra 2007)		✓	✓	✓
TBSVM (Shao et al. 2011)	✓	✓	✓	✓
LSTWSVM (Kumar and Gopal 2009)			✓	
RELS-TSVM (Tanveer et al. 2016a)	✓		✓	✓
Projection TWSVM (Hua et al. 2017)	✓		✓	✓
TPMSVM (Peng 2011a)	✓	✓		✓
Pin-GTSVM (Tanveer et al. 2019b)			✓	✓
SP-TWSVM (Tanveer et al. 2019c)		✓	✓	✓
Pin-TPMSVM (Xu et al. 2016c)	✓			✓
ISPTWSVM (Tanveer et al. 2019a)	✓	✓	✓	✓

$$f_1(x) = u_1^T x + b_1 \quad \text{and} \quad f_2(x) = u_2^T x + b_2, \tag{11}$$

The two QPPs in linear case are defined as below:

$$\min_{(u_1, b_1, \zeta_1) \in R^{n+1+m}} \frac{1}{2} \, ||y - e\varepsilon_1 - (Au_1 + b_1 e)||^2 + C_1 e^T \zeta_1$$

$$s.t. \qquad y - (Au_1 + b_1 e) \geq e\varepsilon_1 - \zeta_1, \quad \zeta_1 \geq 0 \tag{12}$$

and

$$\min_{(u_2, b_2, \zeta_2) \in R^{n+1+m}} \frac{1}{2} \, ||y + e\varepsilon_2 - (Au_2 + b_2 e)||^2 + C_2 e^T \zeta_2$$

$$s.t. \qquad (Au_2 + b_2 e) - y \geq e\varepsilon_2 - \zeta_2, \quad \zeta_2 \geq 0, \tag{13}$$

here, $C_1, C_2 > 0$; $\varepsilon_1, \varepsilon_2 > 0$ and ζ_1, ζ_2 are slack variables. The final regressor is the mean of up and down regressors in (11), which is given as follows

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) \quad \text{for all} \quad x \in R^n. \tag{14}$$

For nonlinear case, kernel surfaces are used rather than hyperplanes which are given below:

$$f_1(x) = k(x^T, A^T)u_1 + b_1 \quad \text{and} \quad f_2(x) = k(x^T, A^T)u_2 + b_2. \tag{15}$$

The two QPPs in non-linear case are defined as below:

$$\min_{(u_1, b_1, \zeta_1) \in R^{m+1+m}} \frac{1}{2} \, \left|\left|y - e\varepsilon_1 - (k(A, A^T)u_1 + b_1 e)\right|\right|^2 + C_1 e^T \zeta_1$$

$$s.t. \qquad y - (k(A, A^T)u_1 + b_1 e) \geq e\varepsilon_1 - \zeta_1, \zeta_1 \geq 0 \tag{16}$$

and

$$\min_{(u_2, b_2, \zeta_2) \in R^{m+1+m}} \frac{1}{2} \, \left|\left|y + e\varepsilon_2 - (k(A, A^t)u_2 + b_2 e)\right|\right|^2 + C_2 e^T \zeta_2$$

$$s.t. \qquad (k(A, A^T)u_2 + b_2 e) - y \geq e\varepsilon_2 - \zeta_2, \zeta_2 \geq 0. \tag{17}$$

For more details, one can refer to Peng (2010b).

Although taking motivation from TWSVM formulation, Peng (2010b) attempted to propose TSVR where the regressor is obtained via solving a pair of quadratic programming problems (QPPs). However, later authors in Khemchandani et al. (2016) argued that TSVR formulation is not in the true spirit of TWSVM. Further, taking the motivation from Bi and Bennett (2003), they proposed an alternative approach to find a formulation for TSVR which is in the true spirit of TWSVM. They have shown that their proposed TSVR formulation can be derived from TWSVM for an appropriately constructed classification problem.

6 Research progress on twin support vector regression

In this section, we discuss the progress of twin SVM based models for the regression problems.

6.1 Weighted twin support vector regression

Xu and Wang (2012) introduced weighted TSVR which gives different weights to samples and have different influence over bound functions. Computational results have demonstrated that this algorithm avoids over-fitting and also yields good generalization ability. The authors Xu and Wang (2014) also proposed K nearest neighbor based weighted TSVR (KNNWTSVR) which gives different penalties to the samples based on their local information on number of K-nearest neighbors. The weights are assigns based on number of K-nearest neighbors. KNNWTSVR has better accuracy but similar computational complexity as its optimization is similar to TSVR. However, the above algorithm only implements empirical risk minimization and suffer from the inverse of positive semi-definite matrices. To overcome these limitations, Tanveer et al. (2016) introduced an efficient regularized KNNWTSVR (RKNNWTSVR) algorithm to make it strongly convex. RKNNWTSVR leads to better generalization and robust solution (Table 6). Computational cost is also reduced as it only deals with linear equations. To overcome noise sensitivity in TSVR, Ye et al. (2016) introduced weighted matrix in Lagrangian ϵ twin support vector regression. It uses quadratic loss functions and provides different weights to samples through weighted matrix. It obtains better generalization and also has less training time than TSVR and ϵ-TSVR models.

6.2 Projection twin support vector regression

TSVR (Peng 2010b) and twin parametric insensitive SVR (Peng 2012) have obtained better generalization performance than classical SVR but both these algorithms only implement empirical risk minimization and do not include any prior information about the data samples which can lead to noise sensitivity. Thus, Peng et al. (2014) proposed an efficient twin projection SVR (TPSVR) algorithm which exploits the prior structural information of data into the learning process. It seeks two projection axes such that projected points have as small as possible empirical variance values on the down-and up-bound functions. This algorithm has better generalization and requires small number of support vectors. The aforementioned models, use uniform weighting approach and assumes that all the samples are equally important. However, this assumption may be wrong due to outliers and noise. Hence, wavelet weighted projection TWSVM for regression (Wang et al. 2019b, 2021b) used wavelet based weights to reduce the effect of outliers.

6.3 Robust and sparse twin support vector regression

Although TSVR has proven to be an effective classifier with good generalization ability, it is less robust due to the square of the 2-norm in the QPP of TSVR. Zhong et al. (2012) improved TSVR by using 1-norm rather than 2-norm distance in TSVR's QPP. It has less training time and better generalization. Chen et al. (2012) formulated smooth TSVR (STSVR) using smooth function in order to make the QPP of TSVR positive definite to obtain a unique global solution. Authors converted the QPPs to unconstrained minimization problems (UMPs) and applied Newton method to solve it effectively. All these algorithms still have high computational time due to quadratic or linear programming problems. To avoid this shortcoming, a least squares version of TSVR (TLSSVR) was formulated by Zhao et al. (2013) which leads to faster computational speed as it only deals with set of linear equations. Authors also proposed sparse TLSSVR.

Chen et al. (2014c) introduced the regularization into the formulation of TSVR and implemented l_1-norm loss function to make it robust and sparse simultaneously. Huang et al. (2016) formulated a sparse version of least square TSVR by introducing a regularization term to make it strongly convex and also converted the primal problems to linear programming problems. This leads to a sparse solution with significantly less computational time. Tanveer (2017) formulated 1-norm TSVR to improve robustness and sparsity in original TSVR. 1-norm TSVR has better accuracy, generalization, and less computational time than TSVR. In 2020, Singla et al. (2020) proposed a novel TSVR (Res-TSVR) which is robust and not sensitive to noise in data. The optimization problem is non-convex with smooth l_2 regularizer and thus, to solve it efficiently, the authors converted it to a convex optimization problem. Res-TSVR performed best as compared to other robust TSVR algorithms in terms of robustness to noise and better generalization. Gu et al. (2020) also proposed a TSVR variant suitable to handle noise called fast clustering-based weighted TSVR (FCWTSVR) which classify the samples into different categories based on their similarities and provides different weights to samples located at different positions. The proposed algorithm performed better than TSVR, ε-TSVR, KNNWTSVR and WL-ε-TSVR. ϵ-non parallel support vector regression (Carrasco et al. 2019) uses two ϵ-tubes for better alignment of hyperplanes and to get the more robust upbound and down bound regressor. Robust huber loss based twin SVR (Balasundaram and Prasad 2020) penalizes the large deviation samples linearly and small error samples are squared. This results in robustness to noise and outliers.

6.4 Other improvements on twin support vector regression

TSVR has proven to provide better generalization results but it needs to solve two QPPs which increases the learning time for TSVR. Thus, Peng (2010a) formulated a primal TSVR (P-TSVR), which only deals with linear equations. This improves the learning speed of TSVR and shows comparable generalization. Further, to increase the sparsity of TSVR, the author introduced the back-fitting strategy for optimizing the unconstrained QPP. TSVR requires two set of constraints one with each QPP which increases the computational time for large datasets. To overcome this disadvantage, Singh et al. (2011) introduced rectangular kernels in the formulation of TSVR and proposed reduced TSVR which resulted in the significant saving of computational time and thus promoting its application for large datasets. To further reduce computational time, a LS version of TSVR (TLSSVR) was formulated by Zhao et al. (2013) which only deals with a set of linear equations. Authors also proposed sparse TLSSVR.

Huang and Ding (2013) further attempted to reduce the computational time by proposing LS-TWSVM in primal space (TLSSVR) rather than dual space (PLSTSVR). This also requires to find solution of two linear equations and has comparable accuracy to TSVR. To make TSVR suitable to handle heteroscedastic noise structure, Peng (2012) proposed twin parametric insensitive SVR (TPISVR) which determines a set of nonparallel parametric-insensitive up and down-bound functions. It also works effectively when noise depends upon the input value. It requires to solve two SVM-type problems, which increases its learning speed. Computational results showed that it also has good generalization ability. Shao et al. (2013e) implemented the SRM principle in TSVR primal space and proposed a new regressor ϵ—TSVR which seeks to find a pair of ϵ-insensitive proximal functions. Further, to reduce complexity, the successive over-relaxation (SOR) technique is employed. Experimental results show that ϵ−TSVR has better generalization and fast training speed than TSVR.

Balasundaram and Tanveer (2013a) proposed linearly convergent Lagrangian TSVR (LTSVR) algorithm. Experimental results have exhibited its suitability and applicability due to the better generalization and less computational time than TSVR. Inspired by this algorithm, Balasundaram and Gupta (2014) proposed Lagrangian dual of the 2-norm TSVR. Results have demonstrated an increase in learning speed with better accuracy when compared to TSVR. Tanveer et al. (2016) introduced the regularization term to the objective function of TSVR and formulated regularized Lagrangian TSVR (RLTSVR).This algorithm implements the SRM principle and requires to solve linear system of equations in place of QPP in TSVR. Optimization problems are positive definite and avoid the singularity in the solution. It has better accuracy and speed than conventional TSVR (Table 7).

Balasundaram and Tanveer (2013b) proposed smooth Newton method for LTSVR which needs to solve linear equations in each iteration using Newton-Armijo algorithm. It has comparable generalization ability but it is at least two times faster than TSVR. Khemchandani et al. (2013) proposed TSVR for simultaneous learning. This algorithm is more accurate, computationally less complex and more robust as it uses l_1 norm error. Lagrangian twin parametric insensitive twin SVR (Gupta et al. 2020; Gupta and Richhariya 2021) employed l_2 norm of the square variables, also it is faster as it uses linearly convergent iterative scheme for obtaining the end regressor. Asymmetric possibility and necessity regression by twin-support vector networks (Hao 2020) and reularization based twin SVR with huber loss (Gupta and Gupta 2021) improved the generalization performance of the end regressor.

Peng et al. (2015a) implemented the use of interval data to handle interval input-output data (ITSVR). Rastogi et al. (2017a, 2018a) provided an extension of ν-SVR i.e ν-TWSVR and large margin distribution machine based regression that it is in the true spirit of TWSVM. Balasundaram and Meena (2016) proposed unconstrained TSVR formulation in the primal space (UP-TSVR) which is speed and obtains better generalization than TSVR. This model is solved by a gradient based iterative approach.

Parastalooi et al. (2016) proposed a modified version of TSVR by including the structural information from data and its distribution. Clustering is done based on the mean and covariance matrix of the data which increases accuracy. Furthermore, to increase the training speed, authors applied SOR technique and also optimized parameter selection by implementing PSO algorithm. Rastogi et al. (2017b) proposed an improved version of ν-TSVR which can automatically adjust the values of upper and lower bound parameters to attain better accuracy. Experimental results have shown the superiority of the proposed algorithm over ϵ-TSVR.

TSVR gives same weights to all the samples but in fact, different positions will influence differently on the regressor which are ignored in TSVR. Thus, Xu et al. (2018a) proposed

Table 6 Performance of various non-linear twin support vector regression (TSVR) based algorithms (Tanveer et al. 2016)

Datasets	Models	RMSE	MAE	SSE/SST	SSR/SST
IBM	SVR (Vapnik and Vapnik 1998)	0.1283	0.0920	0.2872	0.4055
	WSVR (Han and Clemmensen 2014)	0.0459	0.0379	0.0416	0.8365
	TSVR (Peng 2010b)	0.0765	0.0551	0.1125	1.3937
	KNNWTSVR (Xu and Wang 2014)	0.0330	0.0243	0.0217	1.0549
	RKNNWTSVR (Tanveer et al. 2016)	0.0330	0.0243	0.0217	1.0549
Intel	SVR (Vapnik and Vapnik 1998)	0.0500	0.0406	0.0552	0.8174
	WSVR (Han and Clemmensen 2014)	0.0384	0.0295	0.0330	1.0524
	TSVR (Peng 2010b)	0.0405	0.0314	0.0368	0.8208
	KNNWTSVR (Xu and Wang 2014)	0.0971	0.0791	0.1922	0.5566
	RKNNWTSVR (Tanveer et al. 2016)	0.0382	0.0285	0.0328	0.8390
SNP-500	SVR (Vapnik and Vapnik 1998)	0.0311	0.0253	0.0192	0.8825
	WSVR (Han and Clemmensen 2014)	0.0296	0.0222	0.0174	1.064
	TSVR (Peng 2010b)	0.0288	0.0212	0.0166	0.9702
	KNNWTSVR (Xu and Wang 2014)	0.0296	0.0219	0.0174	0.9904
	RKNNWTSVR (Tanveer et al. 2016)	0.0273	0.0193	0.0148	0.9907

Table 7 Performance of various non-linear twin support vector regression (TSVR) based algorithms (Tanveer and Shubham 2017b)

Datasets	Models	RMSE	MAE	SSE/SST	SSR/SST
Gas furnace	SVR (Vapnik and Vapnik 1998)	0.0700	0.0459	0.088	0.8719
	TSVR (Peng 2010b)	0.0578	0.0389	0.0599	0.8744
	LTSVR (Balasundaram and Tanveer 2013a)	0.0634	0.0427	0.0718	0.8689
	RLTSVR (Tanveer and Shubham 2017b)	0.0636	0.0439	0.0718	0.8623
IBM	SVR (Vapnik and Vapnik 1998)	0.1283	0.0920	0.2872	0.4055
	TSVR (Peng 2010b)	0.0765	0.0551	0.1125	1.3937
	LTSVR (Balasundaram and Tanveer 2013a)	0.0328	0.0241	0.0214	1.0567
	RLTSVR (Tanveer and Shubham 2017b)	0.0330	0.0245	0.0217	1.0604
Intel	SVR (Vapnik and Vapnik 1998)	0.0500	0.0406	0.0552	0.8174
	TSVR (Peng 2010b)	0.0405	0.0314	0.0368	0.8208
	LTSVR (Balasundaram and Tanveer 2013a)	0.0343	0.0257	0.0266	0.9180
	RLTSVR (Tanveer and Shubham 2017b)	0.0336	0.0251	0.0255	0.9421

asymmetric ν-TSVR based on pinball loss function. Pinball loss function gives different penalties to the points lying above and below the bounds. It is insensitive to noise and also has better generalization ability. Tanveer and Shubham (2017b) added regularization term in TSVR in the primal form which yields better accuracy and more robust solution (Tables 6, 7).

Table 8 Properties of different TSVR algorithms

Models\characteristics	SRM	Sparsity	Matrix inversion	Insensitive to noise
TSVR (Peng 2010b)			✓	✓
TWSVR (Khemchandani et al. 2016)	✓		✓	✓
ϵ−TSVR (Shao et al. 2013e)	✓		✓	✓
WTSVR (Xu and Wang 2012)			✓	✓
LTSVR (Balasundaram and Tanveer 2013a)			✓	✓
RLTSVR (Tanveer and Shubham 2017b)	✓		✓	✓
PTSVR (Peng 2010a)	✓	✓	✓	✓
KNNWTSVR (Xu and Wang 2014)			✓	✓
RKNNWTSVR (Tanveer et al. 2016)	✓		✓	✓

7 Applications of twin support vector regression

Ye et al. (2013b) implemented $L_1 - \epsilon$- TSVR for forecasting inflation. This algorithm proved to be excellent for feature ranking and determined important features for inflation in China. Experimental results showed that its accuracy is better than the ordinary least square (OLS) models. Ye et al. (2013a) also implemented ϵ-wavelet TSVR for inflation forecast. Authors employed the wavelet kernel that can be used for any curve in quadratic continuous integral space. This algorithm derives lower root mean squared error (RMSE) and thus, is an efficient method for inflation forecast. Ding et al. (2013) predicted stock prices using polynomial smooth twin support vector regression. Numerical experiments reveal that this algorithm can obtain better regression performance compared with SVR and TSVR. Le et al. (2014) proposed a novel genetic algorithm (GA) based TSVR to improve the precision of indoor positioning. It performs better than K- nearest neighbor and neural network but comparable to SVR with significantly less computational time. Houssein (2017) proposed particle swarm optimization (PSO) based TSVR for forecasting wind speed. The computational results proved that it achieves better forecasting accuracy and outperformed genetic algorithm with TSVR and TSVR using radial basis kernel function. Wang and Xu (2018) proposed safe screening rule (SSR) based on variational inequality (VI) to make TSVR efficient for large-scale problems as SSR reduces computational time significantly. Authors also implemented dual coordinate descent method (DCDM) to further increase the computational speed of TSVR. Improved twin SVR (Ganaie et al. 2021c, 2022) was formulated for brain age prediction. Twin SVR models have been benchmarked for the prediction of brain age in Alzheimer's disease (Beheshti et al. 2021).

Table 8 shows the differences in major TSVR methods based on the SRM principle, Sparsity, Matrix Inversion and Noise Insensitivity.

8 Future research and development prospects

TWSVM classification algorithms have high training speed and accuracy than conventional SVM but it's still in the primitive stage of development and lacks practical application background. TWSVM has low generalization ability and also lacks in sparsity. Therefore, TWSVM needs further research and improvements to effectively apply to real-life challenges. Future research prospects for TWSVM can be as follows :

- An interesting aspect can be coupling other machine learning algorithms with the TWSVM. For example, a deep convolutional neural network can extract features which can be classified using TWSVM with better accuracy.
- There is limited research on TWSVM applications for large-scale classification. Thus, how to develop TWSVM algorithms for big data classification problems effectively is worthwhile.
- For non-linear classification problems, TWSVM performance highly depends upon kernel function selection and there is not enough research on this to guide researchers to choose kernels as per different applications to get desired accuracy and performance of the TWSVM algorithm. Thus, kernel selection and optimal parameters selection need further study and improvement.
- Currently, only few TWSVM algorithms have been implemented for multi-class classification but it leads to class imbalance problem and often losses sparsity. Thus, further study is required for TWSVM implementation for multi-class classification.
- The main concept of GEPSVM/TWSVM is based on linear discriminant analysis (LDA). A well cross study on TWSVM and LDA is worthy of future work.
- TWSVM applications to health care is currently limited. Thus, how to implement TWSVM effectively for early diagnosis of human diseases like Alzheimer, Epilepsy, Breast cancer etc is worthy of study.
- Clustering, which aims at dividing the data samples into different clusters, is major fundamental problem in classification. Clustering based TWSVM approach is less studied currently and needs further study and development.
- There is limited research on TWSVM applications for remote-sensing. Thus, how to build efficient classifiers for remote-sensing can be explored.

TSVR is much faster than conventional SVR and also has better generalization ability. But, it suffers from a lack of model complexity control and results in over-fitting. It also losses sparsity similar to TWSVM and is sensitive to outliers. Further research on TSVR can be on the following:

- A significant limitation of TSVR is high computational time as it loses the sparsity. More work is required to find an efficient sparse TSVR algorithm.
- Data cleaning, transforming and pre-processing is an important issue for every machine learning technique and can tremendously improve results and even help in identifying novel interactions within data. As TSVR is a relatively new technique, various data handling, cleaning, pre-processing techniques like missing value imputation can be explored for improving the performance of TSVR.
- TSVR is evaluated only on few types of continuous variable problems. Application of TSVR can be explored on a wide range of problems in different domains.
- The current TSVR requires off-setting of multiple hyperparameters and hence optimal parameter selection is an issue. Thus, further research in the direction of identifying and choosing parameters should be done.

References

Agarwal, S., & Tomar, D. (2014). Siddhant: Prediction of software defects using twin support vector machine. In *2014 international conference on information systems and computer networks (ISCON)* (pp. 128–132). IEEE.

Ai, Q., Wang, A., Zhang, A., Wang, Y., & Sun, H. (2018). A multi-class classification weighted least squares twin support vector hypersphere using local density information. *IEEE Access, 6*, 17284–17291.

Alam, S., Kwon, G.-R., Kim, J.-I., & Park, C.-S. (2017). Twin SVM-based classification of Alzheimer disease using complex dual-tree wavelet principal coefficients and LDA. *Journal of Healthcare Engineering*.

Ali, J., Aldhaifallah, M., Nisar, K. S., Aljabr, A., & Tanveer, M. (2022). Regularized least squares twin svm for multiclass classification. *Big Data Research, 27*, 100295.

Anand, P., Pandey, J.P., Rastogi, R., & Chandra, S. (2019). A privacy-preserving twin support vector machine classifier for vertical partitioned data. In *Computational intelligence: Theories, applications and future directions* (Vol. I, pp. 539–552). Springer.

Azad-Manjiri, M., Amiri, A., & Sedghpour, A. S. (2020). ML-SLSTSVM: A new structural least square twin support vector machine for multi-label learning. *Pattern Analysis and Applications, 23*(1), 295–308.

Bai, L., Shao, Y.-H., Wang, Z., & Li, C.-N. (2019). Clustering by twin support vector machine and least square twin support vector classifier with uniform output coding. *Knowledge-Based Systems, 163*, 227–240.

Balasundaram, S., & Gupta, D. (2014). Training Lagrangian twin support vector regression via unconstrained convex minimization. *Knowledge-Based Systems, 59*, 85–96.

Balasundaram, S., & Meena, Y. (2016). Training primal twin support vector regression via unconstrained convex minimization. *Applied Intelligence, 44*(4), 931–955.

Balasundaram, S., & Prasad, S. C. (2020). Robust twin support vector regression based on Huber loss function. *Neural Computing and Applications, 32*(15), 11285–11309.

Balasundaram, S., & Tanveer, M. (2013a). On Lagrangian twin support vector regression. *Neural Computing and Applications, 22*(1), 257–267.

Balasundaram, S., & Tanveer, M. (2013b). Smooth newton method for implicit Lagrangian twin support vector regression. *International Journal of Knowledge-based and Intelligent Engineering Systems, 17*(4), 267–278.

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews, 11*(10), 203–224.

Beheshti, I., Ganaie, M., Paliwal, V., Rastogi, A., Razzak, I., & Tanveer, M. (2021). Predicting brain age using machine learning algorithms: A comprehensive evaluation. *IEEE Journal of Biomedical and Health Informatics*.

Bi, J., & Bennett, K. P. (2003). A geometric approach to support vector regression. *Neurocomputing, 55*, 79–108.

Borah, P., & Gupta, D. (2021). Robust twin bounded support vector machines for outliers and imbalanced data. *Applied Intelligence*, 1–30.

Borah, P., Gupta, D., & Prasad, M. (2018). Improved 2-norm based fuzzy least squares twin support vector machine. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 412–419). IEEE.

Byvatov, E., & Schneider, G. (2003). Support vector machine applications in bioinformatics. *Applied Bioinformatics, 2*(2), 67–77.

Cao, L., & Shen, H. (2016). Combining re-sampling with twin support vector machine for imbalanced data classification. In *2016 17th international conference on parallel and distributed computing, applications and technologies (PDCAT)* (pp. 325–329). IEEE.

Cao, Y., Ding, Z., Xue, F., & Rong, X. (2018). An improved twin support vector machine based on multi-objective cuckoo search for software defect prediction. *International Journal of Bio-Inspired Computation, 11*(4), 282–291.

Carrasco, M., López, J., & Maldonado, S. (2019). Epsilon-nonparallel support vector regression. *Applied Intelligence, 49*(12), 4223–4236.

Chandra, M. A., & Bedi, S. (2018). A twin support vector machine based approach to classifying human skin. In *2018 4th international conference on computing communication and automation (ICCCA)* (pp. 1–5). IEEE.

Chang, C.-C., & Lin, C.-J. (2002). Training ν-support vector regression: Theory and algorithms. *Neural Computation, 14*(8), 1959–1977.

Chen, P.-H., Lin, C.-J., & Schölkopf, B. (2005). A tutorial on ν-support vector machines. *Applied Stochastic Models in Business and Industry, 21*(2), 111–136.

Chen, S., Cao, J., Chen, F., & Liu, B. (2020a). Entropy-based fuzzy least squares twin support vector machine for pattern classification. *Neural Processing Letters, 51*(1), 41–66.

Chen, S.-G., & Wu, X.-J. (2018). A new fuzzy twin support vector machine for pattern classification. *International Journal of Machine Learning and Cybernetics, 9*(9), 1553–1564.

Chen, W.-J., Shao, Y.-H., Deng, N.-Y., & Feng, Z.-L. (2014a). Laplacian least squares twin support vector machine for semi-supervised classification. *Neurocomputing, 145*, 465–476.

Chen, W.-J., Shao, Y.-H., & Hong, N. (2014b). Laplacian smooth twin support vector machine for semi-supervised classification. *International Journal of Machine Learning and Cybernetics, 5*(3), 459–468.

Chen, W.-J., Shao, Y.-H., Li, C.-N., & Deng, N.-Y. (2016). MLTSVM: A novel twin support vector machine to multi-label learning. *Pattern Recognition, 52*, 61–74.

Chen, W.-J., Shao, Y.-H., Li, C.-N., Liu, M.-Z., Wang, Z., & Deng, N.-Y. (2020b). ν-projection twin support vector machine for pattern classification. *Neurocomputing, 376*, 10–24.

Chen, W.-J., Shao, Y.-H., Li, C.-N., Wang, Y.-Q., Liu, M.-Z., & Wang, Z. (2020c). NPrSVM: Nonparallel sparse projection support vector machine with efficient algorithm. *Applied Soft Computing, 90,* 106142.

Chen, X., Yang, J., & Chen, L. (2014c). An improved robust and sparse twin support vector regression via linear programming. *Soft Computing, 18*(12), 2335–2348.

Chen, X., Yang, J., Liang, J., & Ye, Q. (2012). Smooth twin support vector regression. *Neural Computing and Applications, 21*(3), 505–513.

Chen, X., Yang, J., Ye, Q., & Liang, J. (2011). Recursive projection twin support vector machine via within-class variance minimization. *Pattern Recognition, 44*(10–11), 2643–2655.

Chuang, C.-C. (2007). Fuzzy weighted support vector regression with a fuzzy partition. *IEEE Transactions on Systems, Man, and Cybernetics Part B (Cybernetics), 37*(3), 630–640.

Chu, M., Liu, X., Gong, R., & Liu, L. (2018). Multi-class classification method using twin support vector machines with multi-information for steel surface defects. *Chemometrics and Intelligent Laboratory Systems, 176*, 108–118.

Cong, H., Yang, C., & Pu, X. (2008). Efficient speaker recognition based on multi-class twin support vector machines and GMMs. In *2008 IEEE conference on robotics, automation and mechatronics* (pp. 348–352). IEEE.

Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning, 20*(3), 273–297.

Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods.

Dalal, M., Tanveer, M., & Pachori, R. B. (2019). Automated identification system for focal EEG signals using fractal dimension of FAWT-Based sub-bands signals, 583–596.

de Lima, M. D., Costa, N. L., & Barbosa, R. (2018). Improvements on least squares twin multi-class classification support vector machine. *Neurocomputing, 313*, 196–205.

Ding, S., & Hua, X. (2014). Recursive least squares projection twin support vector machines for nonlinear classification. *Neurocomputing, 130*, 3–9.

Ding, S., Huang, H., & Nie, R. (2013). Forecasting method of stock price based on polynomial smooth twin support vector regression. In *International conference on intelligent computing* (pp. 96–105). Springer.

Ding, S., Zhang, X., & Yu, J. (2016). Twin support vector machines based on fruit fly optimization algorithm. *International Journal of Machine Learning and Cybernetics, 7*(2), 193–203.

Ding, S., Zhao, X., Zhang, J., Zhang, X., & Xue, Y. (2017). A review on multi-class TWSVM. *Artificial Intelligence Review*, 1–27.

Ding, X., Zhang, G., Ke, Y., Ma, B., & Li, Z. (2008). High efficient intrusion detection methodology with twin support vector machines. In *2008 international symposium on information science and engineering* (Vol. 1, pp. 560–564). IEEE.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems* (pp. 155–161).

Du, S.-W., Zhang, M.-C., Chen, P., Sun, H.-F., Chen, W.-J., & Shao, Y.-H. (2021). A multiclass nonparallel parametric-margin support vector machine. *Information, 12*(12), 515.

Edelman, D. (2007). Adapting support vector machine methods for horserace odds prediction. *Annals of Operations Research, 151*(1), 325–336.

Elattar, E. E., Goulermas, J., & Wu, Q. H. (2010). Electric load forecasting based on locally weighted support vector regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(4), 438–447.

Fan, Q., Wang, Z., Li, D., Gao, D., & Zha, H. (2017). Entropy-based fuzzy support vector machine for imbalanced datasets. *Knowledge-Based Systems, 115*, 87–99.

Francis, L. M., & Sreenath, N. (2019). Robust scene text recognition: Using manifold regularized twin-support vector machine. *Journal of King Saud University-Computer and Information Sciences*.

Ganaie, M., Beheshti, I., & Tanveer, M. (2022). Brain age prediction with improved least squares twin SVR. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2022.3147524

Ganaie, M., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2021a). Ensemble deep learning: A review. arXiv preprint arXiv:2104.02395.

Ganaie, M., Tanveer, M., & Initiative, A. D. N. (2021b). Fuzzy least squares projection twin support vector machines for class imbalance learning. *Applied Soft Computing*, 107933

Ganaie, M., & Tanveer, M. (2021). Robust general twin support vector machine with pinball loss function. *Machine Learning for Intelligent Multimedia Analytics*, 103–125.

Ganaie, M., & Tanveer, M. (2020). LSTSVM classifier with enhanced features from pre-trained functional link network. *Applied Soft Computing, 93,* 106305.

Ganaie, M., Tanveer, M., & Beheshti, I. (2021c). Brain age prediction using improved twin SVR. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-06518-1

Ganaie, M. A., Tanveer, M., & Suganthan, P. N. (2020a). Oblique decision tree ensemble via twin bounded SVM. *Expert Systems with Applications, 143,* 113072.

Ganaie, M. A., Tanveer, M., & Suganthan, P. N. (2020b). Regularized robust fuzzy least squares twin support vector machine for class imbalance learning. In *2020 international joint conference on neural networks, IJCNN* (pp. 1–8). IEEE.

Gao, Z., Fang, S.-C., Gao, X., Luo, J., & Medhin, N. (2021). A novel kernel-free least squares twin support vector machine for fast and accurate multi-class classification. *Knowledge-Based Systems, 226*, 107123.

Gao, S., Ye, Q., & Ye, N. (2011). 1-norm least squares twin support vector machines. *Neurocomputing, 74*(17), 3590–3597.

Ghorai, S., Hossian, S.J., Mukherjee, A., & Dutta, P. K. (2010). Unity norm twin support vector machine classifier. In *2010 annual IEEE India conference (INDICON)* (pp. 1–4). IEEE.

Gu, B., Fang, J., Pan, F., & Bai, Z. (2020). Fast clustering-based weighted twin support vector regression. *Soft Computing*, 1–17.

Guo, J., Yi, P., Wang, R., Ye, Q., & Zhao, C. (2014). Feature selection for least squares projection twin support vector machine. *Neurocomputing, 144*, 174–183.

Gupta, D., Acharjee, K., & Richhariya, B. (2020). Lagrangian twin parametric insensitive support vector regression (LTPISVR). *Neural Computing and Applications, 32*(10), 5989–6007.

Gupta, D., Borah, P., Sharma, U. M., & Prasad, M. (2021). Data-driven mechanism based on fuzzy Lagrangian twin parametric-margin support vector machine for biomedical data analysis. *Neural Computing and Applications*, 1–11.

Gupta, D., Pratama, M., Ma, Z., Li, J., & Prasad, M. (2019a). Financial time series forecasting using twin support vector regression. *PLoS ONE, 14*(3), 0211402.

Gupta, D., & Richhariya, B. (2018). Entropy based fuzzy least squares twin support vector machine for class imbalance learning. *Applied Intelligence, 48*(11), 4212–4231.

Gupta, D., & Richhariya, B. (2021). Efficient implicit Lagrangian twin parametric insensitive support vector regression via unconstrained minimization problems. *Annals of Mathematics and Artificial Intelligence, 89*(3), 301–332.

Gupta, U., & Gupta, D. (2019). Lagrangian twin-bounded support vector machine based on l2-norm. In *Recent developments in machine learning and data analytics* (pp. 431–444). Springer.

Gupta, U., & Gupta, D. (2021). On regularization based twin support vector regression with Huber loss. *Neural Processing Letters, 53*(1), 459–515.

Gupta, D., Richhariya, B., & Borah, P. (2019b). A fuzzy twin support vector machine based on information entropy for class imbalance learning. *Neural Computing and Applications, 31*(11), 7153–7164.

Han, X., & Clemmensen, L. (2014). On weighted support vector regression. *Quality and Reliability Engineering International, 30*(6), 891–903.

Hao, P.-Y. (2020). Asymmetric possibility and necessity regression by twin support vector networks. *IEEE Transactions on Fuzzy Systems*.

Hao, P.-Y. (2010). New support vector algorithms with parametric insensitive/margin model. *Neural Networks, 23*(1), 60–73.

Hazarika, B. B., & Gupta, D. (2021). Density weighted twin support vector machines for binary class imbalance learning. *Neural Processing Letters*, 1–40.

He, J., & Zheng, S.-H. (2014). Intrusion detection model with twin support vector machines. *Journal of Shanghai Jiaotong University (Science), 19*(4), 448–454.

Houssein, E. H. (2017). Particle swarm optimization-enhanced twin support vector regression for wind speed forecasting. *Journal of Intelligent Systems*.

Houssein, E. H., Ewees, A. A., & ElAziz, M. A. (2018). Improving twin support vector machine based on hybrid swarm optimizer for heartbeat classification. *Pattern Recognition and Image Analysis, 28*(2), 243–253.

Huang, H., Wei, X., & Zhou, Y. (2016). A sparse method for least squares twin support vector regression. *Neurocomputing, 211*, 150–158.

Huang, H., Wei, X., & Zhou, Y. (2018). Twin support vector machines: A survey. *Neurocomputing, 300*, 34–43.

Hua, S., & Sun, Z. (2001). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics, 17*(8), 721–728.

Hua, X., & Ding, S. (2015). Weighted least squares projection twin support vector machines with local information. *Neurocomputing, 160*, 228–237.

Hua, X., Xu, S., & Gao, J. (2017). A novel projection twin support vector machine for pattern recognition. In *2017 international smart cities conference (ISC2)*. IEEE.

Huang, H.-J., & Ding, S. (2013). Primal least squares twin support vector regression. *Journal of Zhejiang University SCIENCE C, 14*(9), 722–732.

Huang, L.-W., Shao, Y.-H., Zhang, J., Zhao, Y.-T., & Teng, J.-Y. (2019). Robust rescaled hinge loss twin support vector machine for imbalanced noisy classification. *IEEE Access, 7*, 65390–65404.

Jayadeva, R. K., & Chandra, S. (2016). Twin support vector machines: Models, extension and applications, *659*.

Khan, R. U., Tanveer, M., & Pachori, R. B. (2021). A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging. *Expert Systems, 38*(1), 12566.

Khemchandani, R., & Chandra, S. (2007). Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(5), 905–910.

Khemchandani, R., & Chandra, S. (2009). Optimal kernel selection in twin support vector machines. *Optimization Letters, 3*(1), 77–88.

Khemchandani, R., Chandra, S., et al. (2017). Generalized eigenvalue proximal support vector machines. In *Twin support vector machines* (pp. 25–42). Springer.

Khemchandani, R., Goyal, K., & Chandra, S. (2016). TWSVR: Regression via twin support vector machine. *Neural Networks, 74*, 14–21.

Khemchandani, R., Karpatne, A., & Chandra, S. (2013). Twin support vector regression for the simultaneous learning of a function and its derivatives. *International Journal of Machine Learning and Cybernetics, 4*(1), 51–63.

Khemchandani, R., & Pal, A. (2016a). Weighted linear loss twin support vector clustering. In *Proceedings of the 3rd IKDD conference on data science* (p. 18). ACM.

Khemchandani, R., & Pal, A. (2016b). Multi-category Laplacian least squares twin support vector machine. *Applied Intelligence, 45*(2), 458–474.

Khemchandani, R., & Pal, A. (2017). Tree based multi-category Laplacian TWSVM for content based image retrieval. *International Journal of Machine Learning and Cybernetics, 8*(4), 1197–1210.

Khemchandani, R., Pal, A., & Chandra, S. (2018a). Fuzzy least squares twin support vector clustering. *Neural Computing and Applications, 29*(2), 553–563.

Khemchandani, R., Saigal, P., & Chandra, S. (2016). Improvements on ν-twin support vector machine. *Neural Networks, 79*, 97–107.

Khemchandani, R., Saigal, P., & Chandra, S. (2018b). Angle-based twin support vector machines. *Annals of OR, 269*, 387–417.

Khemchandani, R., & Sharma, S. (2016). Robust least squares twin support vector machine for human activity recognition. *Applied Soft Computing, 47*, 33–46.

Khemchandani, R., & Sharma, S. (2017). Robust parametric twin support vector machine and its application in human activity recognition. In proceedings of international conference on computer vision and image processing (pp. 193–203). Springer.

Kostílek, M., & Šťastnỳ, J. (2012). EEG biometric identification: repeatability and influence of movement-related EEG. In *2012 International Conference on Applied Electronics*, (pp. 147–150). IEEE.

Kuhn, H.W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley symposium*, (pp. 481–492). Berkeley: University of California Press.

Kumar, B., & Gupta, D. (2021). Universum based Lagrangian twin bounded support vector machine to classify EEG signals. *Computer Methods and Programs in Biomedicine, 208*, 106244.

Kumar, M. A., & Gopal, M. (2008). Application of smoothing technique on twin support vector machines. *Pattern Recognition Letters, 29*(13), 1842–1848.

Kumar, M. A., & Gopal, M. (2009). Least squares twin support vector machines for pattern classification. *Expert Systems with Applications, 36*(4), 7535–7543.

Kumar, M. A., Khemchandani, R., Gopal, M., & Chandra, S. (2010). Knowledge based least squares twin support vector machines. *Information Sciences, 180*(23), 4606–4618.

Kumar, M. P., & Rajagopal, M. K. (2018). Detecting happiness in human face using unsupervised twin-support vector machines. *International Journal of Intelligent Systems and Applications, 10*(8), 85.

Kumar, M. P., & Rajagopal, M. K. (2019). Detecting facial emotions using normalized minimal feature vectors and semi-supervised twin support vector machines classifier. *Applied Intelligence*, 1–25.

Lang, X., Li, P., Hu, Z., Ren, H., & Li, Y. (2017). Leak detection and location of pipelines based on LMD and least squares twin support vector machine. *IEEE Access, 5*, 8659–8668.

Le Thi, H. A., & Nguyen, M. C. (2017). DCA based algorithms for feature selection in multi-class support vector machine. *Annals of Operations Research, 249*(1–2), 273–300.

Le, W., Wang, Z., Wang, J., Zhao, G., & Miao, H. (2014). A novel wifi indoor positioning method based on genetic algorithm and twin support vector regression. In *The 26th Chinese control and decision conference (2014 CCDC)* (pp. 4859–4862). IEEE.

Li, K., & Lv, Z. (2021). Smooth twin bounded support vector machine with pinball loss. *Applied Intelligence*, 1–17

Li, K., & Ma, H. (2013). A fuzzy twin support vector machine algorithm. *International Journal of Application or Innovation in Engineering and Management (IJAIEM), 2*(3), 459–465.

Li, C.-N., Huang, Y.-F., Wu, H.-J., Shao, Y.-H., & Yang, Z.-M. (2016). Multiple recursive projection twin support vector machine for multi-class classification. *International Journal of Machine Learning and Cybernetics, 7*(5), 729–740.

Li, C.-N., Ren, P.-W., Shao, Y.-H., Ye, Y.-F., & Guo, Y.-R. (2020). Generalized elastic net l_p-norm nonparallel support vector machine. *Engineering Applications of Artificial Intelligence, 88*, 103397.

Li, C.-N., Shao, Y.-H., Wang, H., Zhao, Y.-T., Huang, L.-W., Xiu, N., & Deng, N.-Y.: Single versus union: Non-parallel support vector machine frameworks. arXiv preprint arXiv:1910.09734 (2019)

Li, D., & Tian, Y. (2014). Twin support vector machine in linear programs. *Procedia Computer Science, 29*, 1770–1778.

Li, D., Zhang, H., Khan, M. S., & Mi, F. (2018). A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition. *Biomedical Signal Processing and Control, 41*, 222–232.

López, J., Maldonado, S., & Carrasco, M. (2019). Robust nonparallel support vector machines via second-order cone programming. *Neurocomputing, 364*, 227–238.

Ma, J., Yang, L., & Sun, Q. (2020). Capped l_1-norm distance metric-based fast robust twin bounded support vector machine. *Neurocomputing, 412*, 295–311.

Ma, J., Yang, L., & Sun, Q. (2021). Adaptive robust learning framework for twin support vector machine classification. *Knowledge-Based Systems, 211*, 106536.

Mangasarian, O. L., & Wild, E. W. (2001). Proximal support vector machine classifiers. In: Proceedings KDD-2001: Knowledge discovery and data mining. Citeseer.

Mangasarian, O. L., & Wild, E. W. (2006). Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(1), 69–74.

Mei, B., & Xu, Y. (2019). Multi-task least squares twin support vector machine for classification. Neurocomputing.

Moezzi, S., Jalali, M., & Forghani, Y. (2019). TWSVC+: Improved twin support vector machine-based clustering. *Ingénierie des Systèmes d'Information, 24*(5), 463–471.

Moosaei, H., Ketabchi, S., Razzaghi, M., & Tanveer, M. (2021). Generalized twin support vector machines. *Neural Processing Letters, 53*(2), 1545–1564.

Morra, J. H., Tu, Z., Apostolova, L. G., Green, A. E., Toga, A. W., & Thompson, P. M. (2010). Comparison of adaboost and support vector machines for detecting Alzheimer disease through automated hippocampal segmentation. *IEEE Transactions on Medical Imaging, 29*(1), 30–43.

Mousavi, A., Ghidary, S. S., & Karimi, Z. (2015). Semi-supervised intrusion detection via online Laplacian twin support vector machine. In *2015 signal processing and intelligent systems conference (SPIS)* (pp. 138–142). IEEE.

Mozafari, K., Nasiri, J. A., Charkari, N. M., & Jalili, S. (2011). Action recognition by local space-time features and least square Twin SVM (LS-TSVM). In *2011 First international conference on informatics and computational intelligence* (pp. 287–292). IEEE.

Nasiri, J. A., Charkari, N. M., & Jalili, S. (2015). Least squares twin multi-class classification support vector machine. *Pattern Recognition, 48*(3), 984–992.

Nasiri, J. A., Charkari, N. M., & Mozafari, K. (2014). Energy-based model of least squares twin support vector machines for human action recognition. *Signal Processing, 104*, 248–257.

Nie, P., Zang, L., & Liu, L. (2013). Application of multi-class classification algorithm based on twin support vector machine in intrusion detection. *Jisuanji Yingyong/ Journal of Computer Applications, 33*(2), 426–429.

Noble, W. S. (2004). Support vector machine applications in computational biology. *Kernel Methods in Computational Biology, 71*, 92.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: An application to face detection. In *CVPR* (p. 130). IEEE.

Pan, X., Luo, Y., & Xu, Y. (2015). *K*-nearest neighbor based structural twin support vector machine. *Knowledge-Based Systems, 88*, 34–44.

Pang, X., Xu, C., & Xu, Y. (2018). Scaling KNN multi-class twin support vector machine via safe instance reduction. *Knowledge-Based Systems, 148*, 17–30.

Pang, X., & Xu, Y. (2019). A safe screening rule for accelerating weighted twin support vector machine. *Soft Computing, 23*(17), 7725–7739.

Pan, X., Yang, Z., Xu, Y., & Wang, L. (2017). Safe screening rules for accelerating twin support vector machine classification. *IEEE Transactions on Neural Networks and Learning Systems, 29*(5), 1876–1887.

Pant, H., Sharma, M., & Soman, S. (2019). Twin neural networks for the classification of large unbalanced datasets. *Neurocomputing, 343*, 34–49.

Parastalooi, N., Amiri, A., & Aliheidari, P. (2016). Modified twin support vector regression. *Neurocomputing, 211*, 84–97.

Peng, X. (2010a). Primal twin support vector regression and its sparse approximation. *Neurocomputing, 73*(16–18), 2846–2858.

Peng, X. (2010b). TSVR: An efficient twin support vector machine for regression. *Neural Networks, 23*(3), 365–372.

Peng, X. (2010c). A ν-twin support vector machine (ν-TSVM) classifier and its geometric algorithms. *Information Sciences, 180*(20), 3863–3875.

Peng, X. (2011a). TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition. *Pattern Recognition, 44*(10–11), 2678–2692.

Peng, X. (2011b). Building sparse twin support vector machine classifiers in primal space. *Information Sciences, 181*(18), 3967–3980.

Peng, X. (2012). Efficient twin parametric insensitive support vector regression model. *Neurocomputing, 79*, 26–38.

Peng, X., Chen, D., Kong, L., & Xu, D. (2015a). Interval twin support vector regression algorithm for interval input-output data. *International Journal of Machine Learning and Cybernetics, 6*(5), 719–732.

Peng, X., Kong, L., & Chen, D. (2015b). Improvements on twin parametric-margin support vector machine. *Neurocomputing, 151*, 857–863.

Peng, X., & Xu, D. (2012). Twin mahalanobis distance-based support vector machines for pattern recognition. *Information Sciences, 200*, 22–37.

Peng, X., & Xu, D. (2013a). Bi-density twin support vector machines for pattern recognition. *Neurocomputing, 99*, 134–143.

Peng, X., & Xu, D. (2013b). A twin-hypersphere support vector machine classifier and the fast learning algorithm. *Information Sciences, 221*, 12–27.

Peng, X., & Xu, D. (2013c). Robust minimum class variance twin support vector machine classifier. *Neural Computing and Applications, 22*(5), 999–1011.

Peng, X., Xu, D., & Shen, J. (2014). A twin projection support vector machine for data regression. *Neurocomputing, 138*, 131–141.

Peng, X., Wang, Y., & Xu, D. (2013). Structural twin parametric-margin support vector machine for binary classification. *Knowledge-Based Systems, 49*, 63–72.

Poursaeidi, M. H., & Kundakcioglu, O. E. (2014). Robust support vector machines for multiple instance learning. *Annals of Operations Research, 216*(1), 205–227.

Qiang, W., Zhang, J., Zhen, L., & Jing, L. (2020). Robust weighted linear loss twin multi-class support vector regression for large-scale classification. *Signal Processing, 170*, 107449.

Qi, Z., Tian, Y., & Shi, Y. (2012a). Twin support vector machine with universum data. *Neural Networks, 36*, 112–119.

Qi, Z., Tian, Y., & Shi, Y. (2012b). Laplacian twin support vector machine for semi-supervised classification. *Neural Networks, 35*, 46–53.

Qi, Z., Tian, Y., & Shi, Y. (2013a). Structural twin support vector machine for classification. *Knowledge-Based Systems, 43*, 74–81.

Qi, Z., Tian, Y., & Shi, Y. (2013b). Robust twin support vector machine for pattern classification. *Pattern Recognition, 46*(1), 305–316.

Qi, Z., Tian, Y., & Shi, Y. (2014). A nonparallel support vector machine for a classification problem with universum learning. *Journal of Computational and Applied Mathematics, 263*, 288–298.

Rastogi, R., Anand, P., & Chandra, S. (2017a). ν-norm twin support vector machine-based regression. *Optimization, 66*(11), 1895–1911.

Rastogi, R., Anand, P., & Chandra, S. (2017b). A ν-twin support vector machine based regression with automatic accuracy control. *Applied Intelligence, 46*(3), 670–683.

Rastogi, R., Anand, P., & Chandra, S. (2018a). Large-margin distribution machine-based regression. *Neural Computing and Applications, 3*, 1–16.

Rastogi, R., & Pal, A. (2019). Fuzzy semi-supervised weighted linear loss twin support vector clustering. *Knowledge-Based Systems, 165*, 132–148.

Rastogi, R., & Pal, A. (2021). Efficient learning of pinball twsvm using privileged information and its applications. arXiv preprint arXiv:2107.06744

Rastogi, R., Pal, A., & Chandra, S. (2018b). Generalized pinball loss SVMs. *Neurocomputing, 322*, 151–165.

Rastogi, R., & Saigal, P. (2017). Tree-based localized fuzzy twin support vector clustering with square loss function. *Applied Intelligence, 47*(1), 96–113.

Rastogi, R., Saigal, P., & Chandra, S. (2018c). Angle-based twin parametric-margin support vector machine for pattern classification. *Knowledge-Based Systems, 139*, 64–77.

Rastogi, R., & Sharma, S. (2019). Fast Laplacian twin support vector machine with active learning for pattern classification. *Applied Soft Computing, 74*, 424–439.

Rastogi, R., Sharma, S., & Chandra, S. (2018d). Robust parametric twin support vector machine for pattern classification. *Neural Processing Letters, 47*(1), 293–323.

Refahi, M. S., Nasiri, J. A., & Ahadi, S. (2018). ECG arrhythmia classification using least squares twin support vector machines. In *Iranian conference on electrical engineering (ICEE)* (pp. 1619–1623). IEEE.

Ren, Q., & Yang, L. (2021). A robust projection twin support vector machine with a generalized correntropy-based loss. *Applied Intelligence*, 1–17.

Rezvani, S., Wang, X., & Pourpanah, F. (2019). Intuitionistic fuzzy twin support vector machines. *IEEE Transactions on Fuzzy Systems, 27*(11), 2140–2151.

Richhariya, B., & Gupta, D. (2019). Facial expression recognition using iterative universum twin support vector machine. *Applied Soft Computing, 76*, 53–67.

Richhariya, B., Sharma, A., & Tanveer, M. (2018). Improved universum twin support vector machine. In *2018 IEEE symposium series on computational intelligence (SSCI) (pp. 2045–2052)*. IEEE.

Richhariya, B., & Tanveer, M. (2018a). A robust fuzzy least squares twin support vector machine for class imbalance learning. *Applied Soft Computing, 71*, 418–432.

Richhariya, B., & Tanveer, M. (2018b). EEG signal classification using universum support vector machine. *Expert Systems with Applications, 106*, 169–182.

Richhariya, B., & Tanveer, M. (2019). A fuzzy universum support vector machine based on information entropy. *Machine Intelligence and Signal Analysis*, 569–582.

Richhariya, B., & Tanveer, M. (2020a). Universum least squares twin parametric margin support vector machine. In *2020 international joint conference on neural networks (IJCNN)*, 1–8.

Richhariya, B., & Tanveer, M. (2020b). A reduced universum twin support vector machine for class imbalance learning. *Pattern Recognition, 102*,107150.

Richhariya, B., & Tanveer, M. (2021a). A fuzzy universum least squares twin support vector machine (FUL-STSVM). *Neural Computing and Applications*, 1–12.

Richhariya, B., & Tanveer, M. (2021b). An efficient angle-based universum least squares twin support vector machine for classification. *ACM Transactions on Internet Technology, 21*, 1–24. https://doi.org/10.1145/3387131.

Richhariya, B., Tanveer, M., & Initiative, A. D. N. (2020a). Least squares projection twin support vector clustering (LSPTSVC). *Information Sciences, 533*, 1–23.

Richhariya, B., Tanveer, M., Rashid, A. H., & Initiative, A. D. N. (2020b). Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomedical Signal Processing and Control, 59*, 101903.

Schölkopf, B., Tsuda, K., & Vert, J.-P. (2004). Support vector machine applications in computational biology.

Shahdoosti, H. R., & Hazavei, S. M. (2018). Combined ripplet and total variation image denoising methods using twin support vector machines. *Multimedia Tools and Applications, 77*(6), 7013–7031.

Shao, Y.-H., Chen, W.-J., & Deng, N.-Y. (2014a). Nonparallel hyperplane support vector machine for binary classification problems. *Information Sciences, 263*, 22–35.

Shao, Y.-H., Chen, W.-J., Huang, W.-B., Yang, Z.-M., & Deng, N.-Y. (2013a). The best separating decision tree twin support vector machine for multi-class classification. *Procedia Computer Science, 17,* 1032–1038.

Shao, Y.-H., Chen, W.-J., Wang, Z., Li, C.-N., & Deng, N.-Y. (2015). Weighted linear loss twin support vector machine for large-scale classification. *Knowledge-Based Systems, 73*, 276–288.

Shao, Y.-H., Chen, W.-J., Zhang, J.-J., Wang, Z., & Deng, N.-Y. (2014b). An efficient weighted Lagrangian twin support vector machine for imbalanced data classification. *Pattern Recognition, 47*(9), 3158–3167.

Shao, Y.-H., & Deng, N.-Y. (2012). A coordinate descent margin based-twin support vector machine for classification. *Neural Networks, 25*, 114–121.

Shao, Y.-H., & Deng, N.-Y. (2013). A novel margin-based twin support vector machine with unity norm hyperplanes. *Neural Computing and Applications, 22*(7–8), 1627–1635.

Shao, Y.-H., Deng, N.-Y., Chen, W.-J., & Wang, Z. (2013b). Improved generalized eigenvalue proximal support vector machine. *IEEE Signal Processing Letters, 20*(3), 213–216.

Shao, Y.-H., Deng, N.-Y., & Yang, Z.-M. (2012a). Least squares recursive projection twin support vector machine for classification. *Pattern Recognition, 45*(6), 2299–2307.

Shao, Y.-H., Deng, N.-Y., Yang, Z.-M., Chen, W.-J., & Wang, Z. (2012b). Probabilistic outputs for twin support vector machines. *Knowledge-Based Systems, 33,* 145–151.

Shao, Y.-H., Wang, Z., Chen, W.-J., & Deng, N.-Y. (2013c). A regularization for the projection twin support vector machine. *Knowledge-Based Systems, 37,* 203–210.

Shao, Y.-H., Wang, Z., Chen, W.-J., & Deng, N.-Y. (2013d). Least squares twin parametric-margin support vector machine for classification. *Applied Intelligence, 39*(3), 451–464.

Shao, Y.-H., Yang, Z.-X., Wang, X.-B., & Deng, N.-Y. (2010). Multiple instance twin support vector machines. *Lecture Notes in Operations Research, 12*, 433–442.

Shao, Y.-H., Zhang, C.-H., Wang, X.-B., & Deng, N.-Y. (2011). Improvements on twin support vector machines. *IEEE Transactions on Neural Networks, 22*(6), 962–968.

Shao, Y.-H., Zhang, C.-H., Yang, Z.-M., Jing, L., & Deng, N.-Y. (2013e). An ε-twin support vector machine for regression. *Neural Computing and Applications, 23*(1), 175–185.

Sharma, R., Goel, T., Tanveer, M., & Murugan, R. (2022). FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans. *Applied Soft Computing, 115*, 108099.

Sharma, S., & Rastogi, R. (2018a). Insensitive zone based pinball loss twin support vector machine for pattern classification. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 2238–2245). IEEE.

Sharma, S., & Rastogi, R. (2018b). *Stochastic conjugate gradient descent twin support vector machine for large scale pattern classification* (pp. 590–602). Springer.

Sharma, S., Rastogi, R., & Chandra, S. (2019). Large-scale twin parametric support vector machine using pinball loss function. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

She, Q., Ma, Y., Meng, M., & Luo, Z. (2015). Multiclass posterior probability twin svm for motor imagery EEG classification. *Computational Intelligence and Neuroscience*, *95*.

Si, X., & Jing, L. (2009). Mass detection in digital mammograms using twin support vector machine-based CAD system. In *2009 WASE international conference on information engineering* (Vol. 1, pp. 240–243). IEEE.

Singh, M., Chadha, J., Ahuja, P., & Chandra, S. (2011). Reduced twin support vector regression. *Neurocomputing, 74*(9), 1474–1477.

Singla, M., Ghosh, D., Shukla, K., & Pedrycz, W. (2020). Robust twin support vector regression based on rescaled hinge loss. *Pattern Recognition*, 107395.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.

Soman, S. (2015). High performance EEG signal classification using classifiability and the twin SVM. *Applied Soft Computing, 30*, 305–318.

Suykens, J., Lukas, L., Van Dooren, P., De Moor, B., & Vandewalle, J. (1999). Least squares support vector machine classifiers: A large scale algorithm. In *European conference on circuit theory and design, ECCTD* (Vol. 99, pp. 839–842). Citeseer.

Tanveer, M. (2013). Smoothing technique on linear programming twin support vector machines. *International Journal of Machine Learning and Computing, 3*(2), 240.

Tanveer, M. (2015a). Newton method for implicit Lagrangian twin support vector machines. *International Journal of Machine Learning and Cybernetics, 6*(6), 1029–1040.

Tanveer, M. (2015b). Robust and sparse linear programming twin support vector machines. *Cognitive Computation, 7*(1), 137–149.

Tanveer, M. (2015c). Application of smoothing techniques for linear programming twin support vector machines. *Knowledge and Information Systems, 45*(1), 191–214.

Tanveer, M. (2017). Linear programming twin support vector regression. *Filomat, 31*(7), 2123–2142.

Tanveer, M., Ganaie, M., & Suganthan, P. (2021a). Ensemble of classification models with weighted functional link network. *Applied Soft Computing, 107*,107322.

Tanveer, M., Gautam, C., & Suganthan, P. N. (2019). Comprehensive evaluation of twin svm based classifiers on UCI datasets. *Applied Soft Computing, 83*, 105617.

Tanveer, M., Gupta, T., Shah, M., & Initiative, A. D. N. (2021b). Pinball loss twin support vector clustering. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17*(2), 1–23.

Tanveer, M., Gupta, T., Shah, M., & Richhariya, B. (2021c). Sparse twin support vector clustering using pinball loss. *IEEE Journal of Biomedical and Health Informatics, 25*, 3776–3783. https://doi.org/10.1109/JBHI. 2021.3059910.

Tanveer, M., Khan, M. A., & Ho, S.-S. (2016a). Robust energy-based least squares twin support vector machines. *Applied Intelligence, 45*(1), 174–186.

Tanveer, M., Mangal, M., Ahmad, I., & Shao, Y.-H. (2016b). One norm linear programming support vector regression. *Neurocomputing, 173*, 1508–1518.

Tanveer, M., Pachori, R. B., & Angami, N. V. (2018a). Classification of seizure and seizure-free EEG signals using hjorth parameters. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 2180–2185). IEEE.

Tanveer, M., Pachori, R. B., & Angami, N. V. (2018b). Entropy based features in FAWT framework for automated detection of epileptic seizure EEG signals. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 1946–1952). IEEE.

Tanveer, M., Rajani, T., & Ganaie, M. (2019a). Improved sparse pinball twin SVM. In: *2019 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE.

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C.-T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications, 16*(1s), 1–35.

Tanveer, M., Sharma, A., & Suganthan, P. (2019b). General twin support vector machine with pinball loss function. *Information Sciences, 494,* 311–327.

Tanveer, M., Sharma, A., & Suganthan, P. N. (2021d). Least squares KNN-based weighted multiclass twin SVM. *Neurocomputing, 459*, 454–464.

Tanveer, M., Sharma, S., & Muhammad, K. (2020a). Large scale least squares twin svms. *ACM Transactions on Internet Technology*, *21*, 1–19. https://doi.org/10.1145/3398379.

Tanveer, M., Sharma, S., Rastogi, R., & Anand, P. (2020b). Sparse support vector machine with pinball loss. *Wiley Transactions on Emerging Telecommunications Technologies (ETT).* https://doi.org/10.1002/ett.3820

Tanveer, M., & Shubham, K. (2017a). Smooth twin support vector machines via unconstrained convex minimization. *Filomat, 31*(8), 2195–2210.

Tanveer, M., & Shubham, K. (2017b). A regularization on Lagrangian twin support vector regression. *International Journal of Machine Learning and Cybernetics, 8*(3), 807–821.

Tanveer, M., Shubham, K., Aldhaifallah, M., & Nisar, K. (2016). An efficient implicit regularized Lagrangian twin support vector regression. *Applied Intelligence, 44*(4), 831–848.

Tanveer, M., Tabish, M., & Jangir, J. (2021e). Pinball twin bounded support vector clustering. In *2021 IEEE EMBS international conference on biomedical and health informatics (BHI)* (pp. 1–4). IEEE.

Tanveer, M., Tabish, M., & Jangir, J. (2021f). Sparse pinball twin bounded support vector clustering. *IEEE Transactions on Computational Social Systems.* https://doi.org/10.1109/TCSS.2021.3122828

Tanveer, M., Tiwari, A., Choudhary, R., & Ganaie, M. (2021g). Large-scale pinball twin support vector machines. *Machine Learning*, 1–24.

Tanveer, M., Tiwari, A., Choudhary, R., & Jalan, S. (2019c). Sparse pinball twin support vector machines. *Applied Soft Computing, 78*, 164–175.

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega, 29*(4), 309–317.

Tian, Y., Ju, X., & Qi, Z. (2014a). Efficient sparse nonparallel support vector machines for classification. *Neural Computing and Applications, 24*(5), 1089–1099.

Tian, Y., Qi, Z., Ju, X., Shi, Y., & Liu, X. (2014b). Nonparallel support vector machines for pattern classification. *IEEE Transactions on Cybernetics, 44*(7), 1067–1079.

Tomar, D., & Agarwal, S. (2014). Feature selection based least square twin support vector machine for diagnosis of heart disease. *International Journal of Bio-Science and Bio-Technology, 6*(2), 69–82.

Tomar, D., & Agarwal, S. (2015a). Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes. *Advances in Artificial Neural Systems*, *1*.

Tomar, D., & Agarwal, S. (2015b). An effective weighted multi-class least squares twin support vector machine for imbalanced data classification. *International Journal of Computational Intelligence Systems, 8*(4), 761–778.

Tomar, D., & Agarwal, S. (2015c). A comparison on multi-class classification methods based on least squares twin support vector machine. *Knowledge-Based Systems, 81*, 131–147.

Tomar, D., Ojha, D., & Agarwal, S. (2014a). An emotion detection system based on multi least squares twin support vector machine. *Advances in Artificial Intelligence*, *8*.

Tomar, D., Prasad, B. R., & Agarwal, S. (2014b). An efficient Parkinson disease diagnosis system based on least squares twin support vector machine and particle swarm optimization. In *2014 9th international conference on industrial and information systems (ICIIS)* (pp. 1–6). IEEE.

Tomar, D., Singhal, S., & Agarwal, S. (2014c). Weighted least square twin support vector machine for imbalanced dataset. *International Journal of Database Theory and Application, 7*(2), 25–36.

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research, 2*, 45–66.

Vapnik, V. (2013). The nature of statistical learning theory.

Vapnik, V., & Vapnik, V. (1998). Statistical learning theory, Vol. 1. Wiley.

Wang, C., Ye, Q., Luo, P., Ye, N., & Fu, L. (2019a). Robust capped l_1-norm twin support vector machine. *Neural Networks, 114*, 47–59.

Wang, D., Ye, Q., Ye, N.(2010a). Localized multi-plane TWSVM classifier via manifold regularization. In *2010 Second international conference on intelligent human-machine systems and cybernetics* (vol. 2, pp. 70–73). IEEE.

Wang, H., & Xu, Y. (2018). Scaling up twin support vector regression with safe screening rule. *Information Sciences, 465*, 174–190.

Wang, H., Xu, Y., & Zhou, Z. (2021a). Twin-parametric margin support vector machine with truncated pinball loss. *Neural Computing and Applications, 33*(8), 3781–3798.

Wang, L., Gao, C., Zhao, N., & Chen, X. (2019b). A projection wavelet weighted twin support vector regression and its primal solution. *Applied Intelligence, 49*(8), 3061–3081.

Wang, L., Ma, Y., Chang, X., Gao, C., Qu, Q., & Chen, X. (2021b). Projection wavelet weighted twin support vector regression for OFDM system channel estimation. *Artificial Intelligence Review, 54*(1), 469–489.

Wang, S., Chen, M., Li, Y., Shao, Y., Zhang, Y., Du, S., & Wu, J. (2016a). Morphological analysis of dendrites and spines by hybridization of ridge detection with twin support vector machine. *PeerJ, 4*, 2207.

Wang, S., Jiang, W., & Tsui, K.-L. (2010b). Adjusted support vector machines based on a new loss function. *Annals of Operations Research, 174*(1), 83–101.

Wang, S., Lu, S., Dong, Z., Yang, J., Yang, M., & Zhang, Y. (2016b). Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection. *Applied Sciences, 6*(6), 169.

Wang, S., Zhang, Y., Liu, G., Phillips, P., & Yuan, T.-F. (2016c). Detection of Alzheimer disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. *Journal of Alzheimer Disease, 50*(1), 233–248.

Wang, S., Zhang, Y., Yang, X., Sun, P., Dong, Z., Liu, A., & Yuan, T.-F. (2015a). Pathological brain detection by a novel image feature fractional Fourier entropy. *Entropy, 17*(12), 8278–8296.

Wang, Z., Chen, X., Shao, Y. H., & Li, C. N. (2019c). Ramp-based twin support vector clustering. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-019-04511-3

Wang, Z., Shao, Y.-H., Bai, L., & Deng, N.-Y. (2015b). Twin support vector machine for clustering. *IEEE Transactions on Neural Networks and Learning Systems, 26*(10), 2583–2588.

Wang, Z., Shao, Y.-H., Bai, L., Li, C.-N., & Liu, L.-M. (2019d). A general model for plane-based clustering with loss function. arXiv preprint arXiv:1901.09178.

Wang, Z., Shao, Y.-H., Bai, L., Li, C.-N., Liu, L.-M., & Deng, N.-Y. (2018). Insensitive stochastic gradient twin support vector machines for large scale problems. *Information Sciences, 462*, 114–131.

Wang, Z., Shao, Y.-H., & Wu, T.-R. (2013). A GA-based model selection for smooth twin parametric-margin support vector machine. *Pattern Recognition, 46*(8), 2267–2277.

Xie, F., & Xu, Y. (2019). An efficient regularized k-nearest neighbor structural twin support vector machine. *Applied Intelligence, 49*(12), 4258–4275.

Xie, J., Hone, K., Xie, W., Gao, X., Shi, Y., & Liu, X. (2013). Extending twin support vector machine classifier for multi-category classification problems. *Intelligent Data Analysis, 17*(4), 649–664.

Xie, X., & Sun, S. (2015). Multitask centroid twin support vector machines. *Neurocomputing, 149*, 1085–1091.

Xu, Y. (2016). K-nearest neighbor-based weighted multi-class twin support vector machine. *Neurocomputing, 205*, 430–438.

Xu, Y. (2017). Maximum margin of twin spheres support vector machine for imbalanced data classification. *IEEE Transactions on Cybernetics, 47*(6), 1540–1550.

Xu, Y., Chen, M., & Li, G. (2016a). Least squares twin support vector machine with universum data for classification. *International Journal of Systems Science, 47*(15), 3637–3645.

Xu, Y., Chen, M., Yang, Z., & Li, G. (2016b). ν-twin support vector machine with universum data for classification. *Applied Intelligence, 44*(4), 956–968.

Xu, Y., & Guo, R. (2014a). An improved ν-twin support vector machine. *Applied Intelligence, 41*(1), 42–54.

Xu, Y., & Guo, R. (2014b). A twin hyper-sphere multi-class classification support vector machine. *Journal of Intelligent & Fuzzy Systems, 27*(4), 1783–1790.

Xu, Y., Guo, R., & Wang, L. (2013). A twin multi-class classification support vector machine. *Cognitive Computation, 5*(4), 580–588.

Xu, Y., Li, X., Pan, X., & Yang, Z. (2018a). Asymmetric ν-twin support vector regression. *Neural Computing and Applications, 30*(12), 3799–3814.

Xu, Y., Lv, X., Wang, Z., & Wang, L. (2014a). A weighted least squares twin support vector machine. *Journal of Information Science and Engineering, 30*(6), 1773–1787.

Xu, Y., Pan, X., Zhou, Z., Yang, Z., & Zhang, Y. (2015). Structural least square twin support vector machine for classification. *Applied Intelligence, 42*(3), 527–536.

Xu, Y., & Wang, L. (2012). A weighted twin support vector regression. *Knowledge-Based Systems, 33*, 92–101.

Xu, Y., & Wang, L. (2014). K-nearest neighbor-based weighted twin support vector regression. *Applied Intelligence, 41*(1), 299–309.

Xu, Y., Wang, L., & Zhong, P. (2012). A rough margin-based ν-twin support vector machine. *Neural Computing and Applications, 21*(6), 1307–1317.

Xu, Y., Wang, Q., Pang, X., & Tian, Y. (2018b). Maximum margin of twin spheres machine with pinball loss for imbalanced data classification. *Applied Intelligence, 48*(1), 23–34.

Xu, Y., Xi, W., Lv, X., & Guo, R. (2012). An improved least squares twin support vector machine. *Journal of Information and Computational Science, 9*(4), 1063–1071.

Xu, Y., Yang, Z., & Pan, X. (2016c). A novel twin support-vector machine with pinball loss. *IEEE Transactions on Neural Networks and Learning Systems, 28*(2), 359–370.

Xu, Y., Yang, Z., Zhang, Y., Pan, X., & Wang, L. (2016d). A maximum margin and minimum volume hyper-spheres machine with pinball loss for imbalanced data classification. *Knowledge-Based Systems, 95*, 75–85.

Xu, Y., Yu, J., & Zhang, Y. (2014b). KNN-based weighted rough ν-twin support vector machine. *Knowledge-Based Systems, 71*, 303–313.

Xu, Y., Zhang, Y., Yang, Z., Pan, X., & Li, G. (2015). Imbalanced and semi-supervised classification for prognosis of ACLF. *Journal of Intelligent & Fuzzy Systems, 28*(2), 737–745.

Yang, C., & Wu, Z. (2009). Study to multi-twin support vector machines and its applications in speaker recognition. In *2009 International conference on computational intelligence and software engineering* (pp. 1–4). IEEE.

Yang, H., Huang, K., King, I., & Lyu, M. R. (2009). Localized support vector regression for time series prediction. *Neurocomputing, 72*(10–12), 2659–2669.

Yang, H.-Y., Wang, X.-Y., Niu, P.-P., & Liu, Y.-C. (2014). Image denoising using nonsubsampled Shearlet transform and twin support vector machines. *Neural Networks, 57*, 152–165.

Yang, Z., Pan, X., & Xu, Y. (2018). Piecewise linear solution path for pinball twin support vector machine. *Knowledge-Based Systems, 160*, 311–324.

Yang, Z., & Xu, Y. (2016). Laplacian twin parametric-margin support vector machine for semi-supervised classification. *Neurocomputing, 171*, 325–334.

Yang, Z., & Xu, Y. (2018). A safe sample screening rule for Laplacian twin parametric-margin support vector machine. *Pattern Recognition, 84*, 1–12.

Yang, Z.-M., Wu, H.-J., Li, C.-N., & Shao, Y.-H. (2016). Least squares recursive projection twin support vector machine for multi-class classification. *International Journal of Machine Learning and Cybernetics, 7*(3), 411–426.

Yang, Z.-X., Shao, Y.-H., & Zhang, X.-S. (2013). Multiple birth support vector machine for multi-class classification. *Neural Computing and Applications, 22*(1), 153–161.

Yan, H., Ye, Q.-L., & Yu, D.-J. (2019). Efficient and robust TWSVM classification via a minimum l_1-norm distance metric criterion. *Machine Learning, 108*(6), 993–1018.

Ye, Q., Ye, N., & Gao, S. (2012a). Density-based weighting multi-surface least squares classification with its applications. *Knowledge and Information Systems, 33*(2), 289–308.

Ye, Q., Zhao, C., Gao, S., & Zheng, H. (2012b). Weighted twin support vector machines with local information and its application. *Neural Networks, 35*, 31–39.

Ye, Q., Zhao, C., Ye, N., & Chen, X. (2011). Localized twin SVM via convex minimization. *Neurocomputing, 74*(4), 580–587.

Ye, Q., Zhao, H., Li, Z., Yang, X., Gao, S., Yin, T., & Ye, N. (2018). L1-norm distance minimization-based fast robust twin support vector k-plane clustering. *IEEE Transactions on Neural Networks and Learning Systems, 29*(9), 4494–4503.

Ye, Y., Shao, Y., & Chen, W. (2013a). Comparing inflation forecasts using an ε-wavelet twin support vector regression. *Journal of Information & Computational Science, 10*(7), 2041–2049.

Ye, Y.-F., Bai, L., Hua, X.-Y., Shao, Y.-H., Wang, Z., & Deng, N.-Y. (2016). Weighted lagrange ε-twin support vector regression. *Neurocomputing, 197*, 53–68.

Ye, Y.-F., Cao, H., Bai, L., Wang, Z., & Shao, Y.-H. (2013b). Exploring determinants of inflation in china based on $l_1 - \epsilon$-twin support vector regression. *Procedia Computer Science, 17*, 514–522.

Yuan, C., & Yang, L. (2021). Capped $l_{2,p}$-norm metric based robust least squares twin support vector machine for pattern classification. *Neural Networks, 142*, 457–478.

Yuan, M., & Xu, Y. (2021). Bound estimation-based safe acceleration for maximum margin of twin spheres machine with pinball loss. *Pattern Recognition, 114*, 107860.

Zhang, H., & Li, H. (2019). Fuzzy twin support vector machine based on intra-class hyperplane. *Journal of Physics Conference Series, 1302*, 032016.

Zhang, X. (2009). Boosting twin support vector machine approach for MCs detection. In *2009 Asia-Pacific conference on information processing* (Vol. 1, pp. 149–152). IEEE.

Zhang, Y., & Wang, S. (2015). Detection of Alzheimer disease by displacement field and machine learning. *PeerJ, 3*, 1251.

Zhang, Z., Zhen, L., Deng, N., & Tan, J. (2014). Sparse least square twin support vector machine with adaptive norm. *Applied Intelligence, 41*(4), 1097–1107.

Zhao, J., Xu, Y., & Fujita, H. (2019). An improved non-parallel universum support vector machine and its safe sample screening rule. *Knowledge-Based Systems, 170*, 79–88.

Zhao, Y.-P., Zhao, J., & Zhao, M. (2013). Twin least squares support vector regression. *Neurocomputing, 118*, 225–236.

Zhong, P., Xu, Y., & Zhao, Y. (2012). Training twin support vector regression via linear programming. *Neural Computing and Applications, 21*(2), 399–407.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.