Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 2, May 2025, pp. 1383∼1391
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i2.pp1383-1391

Chebyshev distance-embedded twin support vector machine for skewed classification problems

Sai Lakshmi Balasubramanian, Gajendran Ganesan
Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Chengalpattu, India

Article Info

Article history:

Received Apr 26, 2024
Revised Nov 7, 2024
Accepted Nov 24, 2024

Keywords:

Benchmark dataset
Chebyshev distance
Rectangular kernel
Support vector machine
Twin support vector machine

ABSTRACT

Support vector machine (SVM) is a pivotal classification algorithm, and its evolutionary counterpart, the twin SVM (TWSVM), has gained acclaim for its advanced generalization capabilities, particularly in handling imbalanced data. TWSVMs achieve swift training by explicitly exploring a pair of non-parallel hyperplanes, yet selecting numerical values for hyperparameters poses a challenge due to the uncertainty introduced by random preferences. This paper presents a novel approach, the Chebyshev distance-based TWSVM, specifically designed for hyperparameter tuning in imbalanced binary classification. This innovative model mitigates the uncertainty of hyperparameter selection by leveraging Chebyshev distance, thereby enhancing the generalization capabilities of the TWSVM. To evaluate its efficacy, computational tests were conducted on publicly accessible real-world benchmark datasets across various domains, including non-linear cases. The results demonstrate that the Chebyshev distance-based TWSVM outperforms several existing methods, achieving superior performance with reduced computational time and setting a new benchmark in the field.

Corresponding Author:

Gajendran Ganesan
Department of Mathematics, College of Engineering and Technology
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu, 603203, Tamil Nadu, India
Email: gajendrg@srmist.edu.in

1. INTRODUCTION

Support vector machine (SVM) is a robust binary classification method introduced by Cortes and Vapnik in 1995. SVM is based on Vapnik-Chervonenkis dimensional theory and structural risk minimization. It utilizes decision hyperplanes, known as support hyperplanes, to separate data into two classes. The distance between these hyperplanes is maximized through quadratic programming, making SVM effective even in non-linearly separable cases with the kernel trick. The SRM principle enhances generalization and reduces training errors. SVM finds widespread applications in text classification, face recognition, finance, bio-medicine, and more. Robust SVM, with a new loss function, addresses noise for better generalization. The basic idea involves finding supporting hyperplanes with the largest distance between them, achieved by solving a convex quadratic minimization problem. Modifications like the least squares SVM simplify hyperplane determination through linear equations, offering advantages in specific scenarios [1].

In non-linear models, a feature map $\phi : \mathbb{R}^n \to \mathbb{H}$ maps data into a higher-dimensional Hilbert space $\mathbb{H}$, where the search for separating hyperplanes occurs, and the associated kernel function is denoted as $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ [2]. The decision function for linear models is of the form:

$$f(x) = sgn(\langle w, z\rangle + b) \tag{1}$$

and the decision function for non-linear models will be,

$$f(x) = sgn(\sum_{i=1}^{i=I} \alpha_i y_i K(x_i, x) + b) \ \ (x \in R^n) \tag{2}$$

The primary aim of SVM (2) and (3) is to minimize a higher bound on generalization error. In linearly separable cases, SVM finds the hyperplane that is farthest from the closest learning samples. For non-separable data, SVM maximizes the margin while keeping the error low. However, traditional SVM faces challenges due to complex quadratic programming problem (QPP) calculations and difficulties handling non-linearly independent data, which may lead to lower accuracy [3].

Despite its superior performance compared to other systems, SVM struggles with complex data due to the high computational cost of solving QPPs. To tackle this, Platt [4] introduced sequential minimal optimization (SMO), a specialized algorithm for solving QPPs during SVM training. Various enhancements, including advanced kernel methods [5], have improved SVM accuracy over the past decade. In 2001, Fung and Mangasarian [6] proposed proximal SVM (PSVM), using systems of linear equations instead of quadratic or linear programming problems, and classifying data by assigning points to the closest of two parallel planes. In 2006, Mangasarian and Wild [7] introduced generalized eigen value proximal support vector machine (GEPSVM), which classifies non-parallel hyperplanes using a generalized eigenvalue approach and avoids solving QPPs. Ye and Ye [8] proposed an improved approach in 2009, laying the groundwork for twin SVM (TWSVM).

Jayadeva *et al.* [9] introduced TWSVM, a binary SVM classifier, to enhance GEPSVM's generalization ability. TWSVM determines two non-parallel hyperplanes and solves two small QPPs, reducing computational cost. TWSVM offers faster performance and better generalization than SVM and GEPSVM, using kernels in non-linear classification while requiring matrix inversions. In 2017, Ma *et al.* [10] proposed an iterative algorithm based on L_2p norm distance to enhance TWSVM for two-class data classification. Although numerous methods exist in the literature to enhance the performance of the TWSVM model, there is still no appropriate method to tune the hyperparameters, which are considered key parameters in the TWSVM model. To address this issue, in 2023 Balasubramanian and Gajendran [11] introduced the P-distance based regularized TWSVM. This method aims to improve the generalization capacity of TWSVM and reduce the randomness in hyperparameter tuning. To be precise, the proposed method specifically addresses the issue of randomly tuning hyperparameters by using Chebyshev distance. This study investigates the effect of a method designed to improve the generalization capacity of TWSVM and reduce the randomness in hyperparameter tuning.

2. TWIN SUPPORT VECTOR MACHINE

Consider all data samples in the positive class of a matrix $A \in \mathbb{R}_1^m \times n$, where $i^{th} row\ A_i \in \mathbb{R}^n$ and the matrix $B \in \mathbb{R}_2^m \times n$ represents the negative class. From (6), the linear TWSVM determines two non-parallel hyperplanes, where $w_1 \in \mathbb{R}^m, w_2 \in \mathbb{R}^m,\ b_1 \in \mathbb{R}\ and\ b_2 \in \mathbb{R}$. The hyperplane of TWSVM is positioned close to the data of one class [1] or $[-1]$ and distanced from the data of another class [1] or $[-1]$ as much as possible.

The formulation of TWSVM can be expressed as follows:

$$\min_{(w_1,b_1,q_1)\in R^{(2n+1)}} \parallel Aw_1 + e_1 b_1 \parallel^2 + c_1 e_2^T q_1,$$

$$such\ that\ -(Bw_1 + e_2 b_1) + q_1 \geq e_2, q_1 \geq 0 \tag{3}$$

$$\min_{(w_2,b_2,q_2)\in R^{(2n+1)}} \parallel Bw_2 + e_2 b_2 \parallel^2 + c_2 e_1^T q_2,$$

$$such\ that\ (Aw_2 + e_1 b_2) + q_2 \geq e_1, q_2 \geq 0 \tag{4}$$

In TWSVM, $c_1, c_2 > 0$ are hyperparameters, e_1, e_2 are vectors of ones of appropriate dimensions and q_1, q_2 are slack variables. Apparently, the goal of TWSVM is to solve the QPP's (7) and (8).

The dual problems can be obtained as follows:

$$\max_{\alpha} e_2^T \alpha - \frac{1}{2}\alpha^T G(H^T H)^{-1} G^T \alpha,\ such\ that\ 0 \leq \alpha \leq c_1 \tag{5}$$

$$\max_{\gamma} e_1^T \gamma - \frac{1}{2}\gamma^T H(G^T G)^{-1} H^T \gamma,\ such\ that\ 0 \leq \gamma \leq c_2 \tag{6}$$

where $\alpha\ and\ \gamma$ are Lagrange multipliers, $H = [Ae_1]$ and $G = [Be_2]$. The solutions of equations utilized to find two hyper planes by,

$$[w_1 b_1]^T = -(H^T H)^{-1} G^T \alpha \text{ and } [w_2 b_2]^T = -(G^T G)^{-1} H^T \gamma$$

Although the matrices $G^T G$ or $H^T H$ are positive semi definite and singular. They introduced a regularization parameter and the inverse matrices $(G^T G)^{-1}$and $(H^T H)^{-1}$ are replaced by $(G^T G + \delta I)^{-1}$ and $(H^T H + \delta I)^{-1}$ respectively, where δ and I are positive scalar and unit matrix respectively [11]. The nonlinear TWSVM are given by,

$$\min_{(w_1,b_1,q_1)\in R^{2n+1}} \parallel K(A,C^T)w_1 + e_1 b_1 \parallel^2 + c_1 e_2^T q_1,$$

$$such\ that - (K(B,C^T)w_1 + e_2 b_1) + q_1 \geq e_2, q_1 \geq 0 \tag{7}$$

$$\min_{(w_2,b_2,q_2)\in R^{2n+1}} \parallel K\left(B,C^T\right) w_2 + e_2 b_2 \parallel^2 + c_2 e_1^T q_2,$$

$$such\ that - \left(K(A,C^T)w_2 + e_1 b_2\right) + q_2 \geq e_1, q_2 \geq 0 \tag{8}$$

The performance of TWSVM exceeds that of classical SVM and GEPSVM on specific benchmark datasets. During simulation, a key challenge is selecting appropriate parameters for model construction. Despite the existence of various parameter selection methods for TWSVM, directly choosing hyperparameters remains a laborious exhaustive search process. To address these issues, the proposed Chebyshev-based (CB-TWSVM) formulation proves beneficial in tuning hyperparameters without relying on randomness.

3. CHEBYSHEV DISTANCE BASED TWSVM

Hyperparameter sensitivity significantly affects the performance, convergence, and generalization of machine learning models, including TWSVM, which are particularly sensitive to the regularization parameter c_i. Smaller values of c_i lead to wider margins but allow more misclassifications, while larger values result in tighter margins with fewer errors. Traditional TWSVM models often lack systematic methods for tuning c_i, prompting the development of CB-TWSVM, a new approach that employs Chebyshev distance to improve hyperparameter selection without relying on random values where numerous literature reviews suggest that distance measures can be employed in various steps of TWSVM algorithm [12]. Additionally, to enhance performance, we use a fixed rectangular kernel. The Chebyshev distance, also known as the maximum or L_∞ metric, measures the largest difference between any coordinate dimension of two points [13]. The mathematical foundation of Chebyshev distance is derived from Minkowski distance. For variables x and x', Minkowski distance is defined as,

$$\mathrm{d}_p(x,x') = \parallel x - x' \parallel_p = (\sum_{i=1}^{d} \mid x_i - x_i' \mid^p)^{\frac{1}{p}}$$

where $p \geq 1$. When $p \rightarrow \infty$, Chebyshev metric can be derived as,

$$D_{che}(x,x') = \parallel x - x' \parallel_\infty = \max_i \mid x_i - x_i' \mid \tag{9}$$

For two points (x_1, y_1) and (x_2, y_2) in a two-dimensional space, the Chebyshev distance is defined as:

$$D_\infty = max(\mid x_2 - x_1 \mid, \mid y_2 - y_1 \mid)$$

general, for n-dimensional space, the Chebyshev distance between two points $(x_1, y_1, \ldots, z_1)$ and $(x_2, y_2, \ldots, z_2)$ is given by:

$$D_\infty = max(\mid x_2 - x_1 \mid, \mid y_2 - y_1 \mid, \ldots, \mid z_2 - z_1 \mid)$$

Chebyshev distance essentially computes the maximum absolute difference along any dimension. Consequently, Chebyshev distance serves as a measure of the maximum separation along any axis between two points in the learning set [14]. Applying (7) to (9) in place of c_1 and c_2 for the positive and negative training sample sets, respectively, we obtain the following: for the positive training sample set, we use (9) in place of c_1 in (7), and for the negative training sample set, we use (9) in place of c_2 in (8). This substitution allows us to utilize the relationship defined in (9) to express c_1 and c_2 for the respective positive and negative sample sets [15]. Proposed CB-TWSVM shown in Algoritm 1.

Algorithm 1. Proposed CB-TWSVM

```
1. $X = \{\bar{x_i} \mid \bar{x_i} = (x_{i1}, x_{i2}, \ldots)\} \, \forall \, x_{i1} \in [0, 1]$ where, $X_1 \subseteq X$ and $X_2 \subseteq X$ are two classes of X are labeled as class I and class II.
2. Set hyperparameters $c_I$,
$c_I \leftarrow c_1, c_2$
for each $c_I$, From (3.), $D_{che}(x, x') = \parallel x - x' \parallel_\infty = \max_i \mid x_i - y'_i \mid$
3. Compute $c_I$
$c_I = \sigma[D_{che}(x_i, x_j)/i \neq j]$
$\forall \; I = 1, 2$ for $x_i, x_j \in X_I$
4. Apply $c_I \leftarrow c_1, c_2$, validating the model
5. Update and fixing kernel parameter as rectangular kernel $k(A, C^T) = k(B, C^T) = \frac{1}{2}$,
6. Generate hyperplane,
(a) $[w_1 b_1]^T = -(H^T H)^{-1} G^T \alpha$
(b) $[w_2 b_2]^T = -(G^T G)^{-1} H^T \gamma$
end
```

4. SIMULATION RESULTS AND DISCUSSIONS

After testing various pairwise distance measures, Chebyshev distance proved effective for tuning c_i values. The performance of the CB-TWSVM was tested on seven benchmark datasets [16], which were normalized to [0,1] scale using Z-score normalization. MATLAB 2023 was used for performance evaluations, with gaussian kernels. The effectiveness of CB-TWSVM was also examined under conditions of data imbalance, a common issue in classification tasks characterized by unequal class distributions, quantified by the imbalance ratio (IR).

4.1. Dataset description

Breast cancer 699: breast cancer, a leading cause of female cancer deaths, has driven research into prevention. Data mining techniques are utilized to extract insights for prediction and classification. The WBCD dataset, containing 699 samples (with 16 missing values as of July 15, 1992), includes 10 attributes. Among these, 241 cases (65.5%) are malignant and 458 (34.5%) are benign. The class attribute uses integers 2 and 4 to denote benign and malignant cases, respectively [17]. Breast cancer 569: breast cancer ranks as the second leading cause of death among women worldwide, underscoring the importance of early detection. To improve predictive accuracy, researchers increasingly rely on advanced machine learning algorithms. A key resource in this effort is the Wisconsin breast cancer dataset (WBCD) from the UCI repository. This dataset contains 569 instances and 30 features, with 357 benign and 212 malignant cases, providing a comprehensive and robust foundation for assessing the effectiveness of machine learning models in predicting breast cancer outcomes [18]. Fertility: infertility carries a societal stigma, with 30% of cases attributed to male factors. Despite this, male infertility remains under diagnosed and under represented. The World Health Organization (WHO) links the decline in male fertility primarily to lifestyle and environmental changes. The dataset includes 9 input features related to these factors, following WHO criteria. It contains 100 samples, with 88 samples indicating normal seminal quality and 12 indicating altered seminal quality [19].

Immunotherapy: human papillomavirus (HPV) infection can cause warts, particularly on the hands and feet, complicating treatment at later stages. A literature review shows varied proposed treatments, with a key challenge being the need for personalized patient care. This variability complicates the identification of optimal treatments for faster recovery. In their study, the authors used immunotherapy datasets from UCI, which contain detailed information on 8 attributes across 90 patient instances. These features are crucial for analysing and understanding immunotherapeutic responses to HPV [20].

German statlog: the German statlog dataset, used for binary classification, contains 1,000 instances with 20 features related to financial history. It's commonly used to predict credit risk (good or bad) and to compare the performance of classification algorithms in machine learning research [21]. Ionosphere: the Ionosphere dataset is used for binary classification tasks in machine learning. It contains 351 instances with 34 features, which are radar signal attributes used to classify whether signals pass through the ionosphere (labelled "good") or are blocked (labelled "bad"). This dataset is often used to evaluate classification algorithms in research [22]. Sonar: the Sonar dataset is used for binary classification in machine learning, consisting of 208 instances with 60 features. Each feature represents the energy of sound waves reflected off objects at various angles, and the task is to classify whether an object is a rock or a metal cylinder (mine). This dataset is commonly used to test and compare the performance of classification algorithms [23].

The purpose of our modelling is to demonstrate the stability of our model across different non-linear datasets. Benchmark datasets from various domains such as biomedical, financial, and other sectors were taken from UCI. To demonstrate the stability of the proposed model, their accuracy, F-mean, recall, and elapsed time were compared with those of other existing SVM models. Table 1 shows the measure of imbalance ratio and their descriptions about dataset.

Table 1. Class imbalance factor

Dataset	Instances	Attributes	IR
Breast cancer	699	10	1.90041
Breast cancer	569	32	1.68396
Fertility	100	10	7.3333
Immunotherapy	90	8	3.5
German statlog	1,000	20	2.3333
Ionosphere	351	34	1.7857
Sonar	208	61	1.1224

Figure 1 depicts the scatter plots which are useful for visualizing the relationship or correlation between two continuous variables. Each Figures 1(a) to 1(g) represents the dispersion of the respective dataset along the x and y axes, respectively. It allow us to observe patterns, trends, and potential outliers in the data. The position of each point on the plot provides information about the values of both variables x_1, x_2 for that specific data point.

Performance of proposed method is computed by the following metrics as sensitivity, accuracy, F1 score, and its elapsed time. Numerical analysis carry through linear $k(x_i, x_j) = x_i^t x_j$ and the non-linear kernel such as gaussian kernel $k(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{2\mu^2})$ with μ where the hyperparameter c_1 and c_2 are tuned using Chebyshev distance whereas fixing rectangular kernel for more precise results.

Performance of the individual benchmark dataset and experimental results signifies the robustness and advantage of CB-TWSVM approach. Table 2 shows the performance of non-linear CB-TWSVM of benchmark breast cancer datasets and it is compared with various SVM's like PSVM, CPSVM, CPSVM, TWSVM, LSTSVM, CPTWSVM, CB-TWSVM with their elapsed time [2].

Table 3 shows the performance of non-linear CB-TWSVM of fertility and immunotherapy datasets. Also, the comparison with other evolved SVM models, evaluation metrics and computational time are discussed [24]. Table 4 shows the performance of non-linear CB-TWSVM of various benchmark datasets. Also, the comparison with other evolved SVM models like TWSVM, TBSVM, F-TWSVM, γ-TWSVM, WLSTSVM, and LSTSVM's evaluation metrics and computational time are discussed [25], [26].

The proposed CB-TWSVM outperforms all the other existing SVM models in accuracy as well as other metrics for the benchmark datasets taken for numerical experiments. Figure 2 with 2(a) to 2(g) in Appendix shows the three dimensional surface plot for representing data with multiple components or for visualizing uncertainties or error. The benchmark dataset is analysed with the target values represented along the x-axis, actual values along the y-axis, and the corresponding errors depicted along the z-axis. The proposed method is further applied to large datasets, including those for image and text recognition.

(a) (b)

(c) (d)

(e) (f)

(g)

Figure 1. Dataset dispersion of: (a) breast cancer (699), (b) breast cancer (569), (c) fertility, (d) immunotherapy, (e) German statlog, (f) ionosphere, and (g) sonar

Table 2. Performance analysis of various SVMs for breast cancer dataset in non-linear case

Dataset	SVM	PSVM	CPSVM	TWSVM	LSTSVM	CPTSVM	CB-TWSVM
	Acc	Acc	Acc	Acc	Acc	Acc	Acc
	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean
	Recall	Recall	Recall	Recall	Recall	Recall	Recall
	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)
BC (683)	97.06	96.88	96.54	96.91	95.98	97.69	99.51
	-	-	-	-	-	-	0.9976
	-	-	-	-	-	-	0.9951
	0.0036	0.1852	0.7261	0.0233	0.0126	0.2346	0.3745
BC (569)	97.93	98.01	97.97	96.34	96.38	97.56	99.12
	-	-	-	-	-	-	0.9956
	-	-	-	-	-	-	0.9912
	0.18	0.1054	0.4479	0.0539	0.0086	0.3759	0.5193

Table 3. Performance analysis of various SVMs for the fertility and immunotherapy in non-linear case

Dataset	FSSVM	FS-CTBSVM	Chi-TBSVM	TBSVM	TWSVM	NN	CB-TWSVM
	Acc	Acc	Acc	Acc	Acc	Acc	Acc
	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean
	Recall	Recall	Recall	Recall	Recall	Recall	Recall
	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)
Fertility	91.89	88.89	89.88	88.98	88.14	87.00	96.67
	-	-	-	-	-	-	0.9831
	-	-	-	-	-	-	0.9667
	0.02	0.004	0.009	0.005	0.74	0.48	0.04283
Immunotherapy	87.89	87.89	86.77	83.41	78.93	87.80	96.30
	-	-	-	-	-	-	0.9811
	-	-	-	-	-	-	0.963
	0.04	0.06	0.03	0.03	0.41	0.56	0.02086

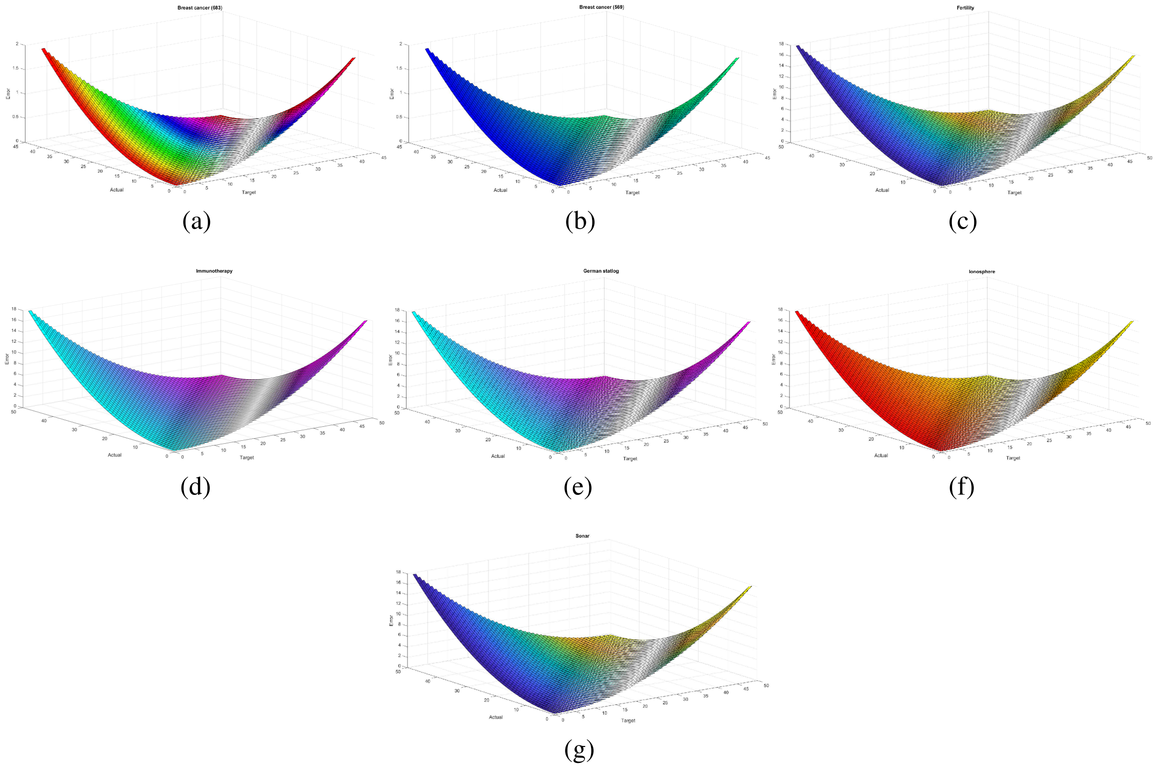
Table 4. Performance analysis of various SVMs for the benchmark datasets in non-linear case

Dataset	TWSVM	TBSVM	F-TWSVM	γ-TWSVM	WLSTSVM	LSTSVM	CB-TWSVM
	Acc	Acc	Acc	Acc	Acc	Acc	Acc
	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean	F-mean
	Recall	Recall	Recall	Recall	Recall	Recall	Recall
	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)	Time(sec)
German Statlog	-	-	78.20	76.25	76.08	74.96	98
	-	-	-	-	-	-	0.9899
	-	-	-	-	-	0-	0.98
	-	-	14.98	13.89	10.713	9.8172	2.4528
Ionosphere	87.46	87.75	-	94.72	95.68	91.82	94.03
	-	-	-	-	-	-	0.9692
	-	-	-	-	-	-	0.9403
	0.0064	0.088	-	3.24	0.6141	0.5438	0.14326
Sonar	89.64	90.0	-	91.71	-	-	95.16
	-	-	-	-	-	-	0.9752
	-	-	-	-	-	-	0.9516
	0.014	0.008	-	0.52	-	-	0.0615

5. CONCLUSION

This paper introduces a novel extension to TWSVM called the CB-TWSVM, designed to enhance the performance of the traditional TWSVM. CB-TWSVM incorporates Chebyshev distance as regularized hyperparameters, leveraging the distance between positive and negative classes in the training data to construct non-parallel hyperplanes. This is achieved by integrating the standard deviation of Chebyshev distances into the model. One of the key advantages of CB-TWSVM is its ability to naturally generalize from TWSVM, strategically considering the distance between binary classes rather than relying on random selection. This approach is particularly beneficial in real-world scenarios where the distance measures between positive and negative classes differ significantly. Experimental results demonstrate that CB-TWSVM offers superior generalization performance compared to other classifiers, outperforming existing models in terms of accuracy and other statistical measures. Specifically, CB-TWSVM achieves the highest accuracy rates of 99.51%, 99.12%, 96.67%, 96.29%, 98%, 94.03%, and 95.16% on the breast cancer 699, breast cancer 569, fertility, immunotherapy, German statlog, ionosphere, and sonar datasets respectively.

APPENDIX



(a) (b) (c)

(d) (e) (f)

(g)

Figure 2. Surface plot for misclassification: (a) breast cancer (699), (b) breast cancer (569), (c) fertility, (d)immunotherapy, (e) German statlog(f) ionosphere, and (g) sonar

REFERENCES

[1] S. Haykin, *Neural networks: a comprehensive foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[2] Y.-H. Shao, X.-J. Lv, L.-W. Huang, and L. Bai, “Twin SVM for conditional probability estimation in binary and multiclass classification,” *Pattern Recognition*, vol. 136, p. 109253, Apr. 2023, doi: 10.1016/j.patcog.2022.109253.

[3] J. Namgyal and E. Schulz, “A novel decision method for multi-class twin support vector machines by average distance,” in *Proceedings of the 1st International Annual Meeting in Mathematics*, 2002, pp. 58–77.

[4] J. Platt, “Sequential minimal optimization: a fast algorithm for training support vector machines,” 1998.

[5] T. Joachims, “Making large-scale support vector machine learning practical,” in *Advances in kernel methods: support vector learning*, 1999, pp. 169–184.

[6] G. Fung and O. L. Mangasarian, “Proximal support vector machine classifiers,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2001, pp. 77–86, doi: 10.1145/502512.502527.

[7] O. L. Mangasarian and E. W. Wild, “Multisurface proximal support vector machine classification via generalized eigenvalues,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 69–74, Jan. 2006, doi: 10.1109/TPAMI.2006.17.

[8] Q. Ye and N. Ye, “Improved proximal support vector machine via generalized eigenvalues,” in *2009 International Joint Conference on Computational Sciences and Optimization*, Apr. 2009, pp. 705–709, doi: 10.1109/CSO.2009.295.

[9] Jayadeva, R. Khemchandani, and S. Chandra, “Twin support vector machines for pattern classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, May 2007, doi: 10.1109/TPAMI.2007.1068.

[10] X. Ma, Q. Ye, and H. Yan, “L2P-norm distance twin support vector machine,” *IEEE Access*, vol. 5, pp. 23473–23483, 2017, doi: 10.1109/ACCESS.2017.2761125.

[11] S. L. Balasubramanian and G. Gajendran, “P-dist based regularized twin support vector machine on imbalanced binary dataset,” *Mathematics and Statistics*, vol. 11, no. 2, pp. 325–334, Mar. 2023, doi: 10.13189/ms.2023.110211.

[12] A. Dax, “The distance between two convex sets,” *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 184–213, Jul. 2006, doi: 10.1016/j.laa.2006.03.022.

[13] A. Elen and E. Avuçlu, “Standardized variable distances: a distance-based machine learning method,” *Applied Soft Computing*, vol. 98, p. 106855, Jan. 2021, doi: 10.1016/j.asoc.2020.106855.

[14] L. V. Utkin, M. S. Kovalev, and E. M. Kasimov, “An explanation method for black-box machine learning survival models using the chebyshev distance,” in *Conference on Artificial Intelligence and Natural Language*, 2020, pp. 62–74, doi: 10.1007/978-3-030-59082-6_5.

[15] I. Iswanto, T. Tulus, and P. Sihombing, “Comparison of distance models on K-nearest neighbor algorithm in stroke disease detection,” *Applied Technology and Computing Science Journal*, vol. 4, no. 1, pp. 63–68, Jul. 2021, doi: 10.33086/atcsj.v4i1.2097.

[16] D. Dua and C. Graff, “UCI machine learning repository,” 2017. http://archive.ics.uci.edu/ml.

[17] Y. Li and Z. Chen, “Performance evaluation of machine learning methods for breast cancer prediction,” *Applied and Computational Mathematics*, vol. 7, no. 4, pp. 212–216, 2018, doi: 10.11648/j.acm.20180704.15.

[18] K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, “Machine learning-based diagnosis of breast cancer utilizing feature optimization technique,” *Computer Methods and Programs in Biomedicine Update*, vol. 3, p. 100098, 2023, doi: 10.1016/j.cmpbup.2023.100098.

[19] D. GhoshRoy, P. Alvi, and K. Santosh, “Unboxing industry-standard AI models for male fertility prediction with SHAP,” *Healthcare*, vol. 11, no. 7, p. 929, Mar. 2023, doi: 10.3390/healthcare11070929.

[20] A. J. Guimarães, P. V. C. Souza, V. J. S. Araújo, T. S. Rezende, and V. S. Araújo, “Pruning fuzzy neural network applied to the construction of expert systems to aid in the diagnosis of the treatment of cryotherapy and immunotherapy,” *Big Data and Cognitive Computing*, vol. 3, no. 2, p. 22, Apr. 2019, doi: 10.3390/bdcc3020022.

[21] G. Zeng, “A comprehensive study of coefficient signs in weighted logistic regression,” *Heliyon*, vol. 10, no. 15, p. e35040, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35040.

[22] D. Bilitza, M. Pezzopane, V. Truhlik, D. Altadill, B. W. Reinisch, and A. Pignalberi, “The international reference ionosphere model: a review and description of an ionospheric benchmark,” *Reviews of Geophysics*, vol. 60, no. 4, Dec. 2022, doi: 10.1029/2022RG000792.

[23] A. Preciado-Grijalva, B. Wehbe, M. B. Firvida, and M. Valdenegro-Toro, “Self-supervised learning for sonar image classification,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, vol. 2022-June, pp. 1498–1507, doi: 10.1109/CVPRW56347.2022.00156.

[24] M. D. de Lima, J. de Oliveira Roque e Lima, and R. M. Barbosa, “Medical data set classification using a new feature selection algorithm combined with twin-bounded support vector machine,” *Medical & Biological Engineering & Computing*, vol. 58, no. 3, pp. 519–528, Mar. 2020, doi: 10.1007/s11517-019-02100-z.

[25] D. Tomar and S. Agarwal, “Twin support vector machine: a review from 2007 to 2014,” *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 55–69, Mar. 2015, doi: 10.1016/j.eij.2014.12.003.

[26] B. S. Lakshmi and G. Gajendran, “Healthcare prognosis: gradilearn-driven elitist genetic algorithm for disease predictions,” *Communications in Mathematical Biology and Neuroscience*, 2024, doi: 10.28919/cmbn/8397.

BIOGRAPHIES OF AUTHORS

Sai Lakshmi Balasubramanian earned her M.Sc. degree in Mathematics from DG Vaishnava College, Chennai, Tamil Nadu, India, in 2011, followed by an M.Phil. degree from PRIST University, Thanjavur, Tamil Nadu, India, in 2013. She is currently pursuing her Ph.D. in the Department of Mathematics at the College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India. Her research interests include neural networks, support vector machines, genetic algorithms, and convolution neural networks. She can be contacted at email: sb7831@srmist.edu.in.

Gajendran Ganesan is an assistant professor (Sr.Gr) at the Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India. He is more than 30 years experience in teaching as well as in the research. His interests include machine learning: artificial neural networks with back propagation, dynamic neural networks, radial basis function and support vector machine. He can be contacted at email: gajendrg@srmist.edu.in.