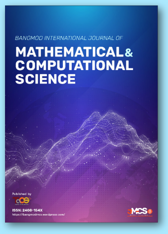
JMCS
BANGMOD INTERNATIONAL JOURNAL OF MATHEMATICAL & COMPUTATIONAL SCIENCE
Volume. 7 Number. 1 & 2 (2021)
Pages 136–154

BANGMOD INTERNATIONAL JOURNAL OF
MATHEMATICAL & COMPUTATIONAL SCIENCE
ISSN: 2408-154X
https://bangmodmcs.wordpress.com/



IMPROVED TWIN SUPPORT VECTOR MACHINE WITH GENERALIZED PINBALL AND APPLICATION ON HUMAN ACTIVITY RECOGNITION

Wanida Panup[1], **Rabian Wangkeeree**[1,2,*]
[1] *Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand*
E-mails: wanidap56@nu.ac.th
[2] *Research center for Academic Excellence in Mathematics, Naresuan University, Phitsanulok 65000, Thailand*
E-mails: rabianw@nu.ac.th

**Corresponding author.*

Received: 25 November 2021 / Accepted: 25 December 2021

Abstract In this paper, we propose a new classifier termed as an improved version of twin support vector machine with generalized pinball (GPin-ITSVM). The primary advantage of GPin-ITSVM over GPin-TSVM is that avoids the singularity problem when solving the dual problems of GPin-TSVM. Motivated by the need to address this issue, we modify the GPin-TSVM by adding an extra regularization term to the objective function in primal problems of GPin-TSVM. Numerical experiments are carried out on 12 UCI benchmark datasets to investigate the validity of our proposed algorithm. The results show that the our GPin-ITSVM is superior to existing classifiers in classification accuracy. In addition, the use of this approach in Weizmann activity recognition applications is investigated, and the automatic feature extractor makes use of 5 types of Convolution Neural Network (CNN) models which are ResNet50, ResNet152V2, InceptionV3, InceptionResNetV2 and Xception.

MSC: 47H09

Keywords: Twin support vector machine; Twin bounded support vector machine; structural risk minimization; Generalized pinball loss; Weizmann activity recognition; Convolution Neural Network

Published online: 31 December 2021

Published by Center of Excellence in Theoretical and Computational Science (TaCS-CoE)

Please cite this article as: W. Panup et al., Improved twin support vector machine with generalized pinball and application on human activity recognition, Bangmod Int. J. Math. & Comp. Sci., Vol. 7 No. 1 & 2 (2021) 136–154.

1. Introduction

The support vector machine (SVM) [1, 2] produces state-of-the-art results in a wide range of applications, with theoretical guarantees. SVM solves a quadratic programming problem (QPP), ensuring that the optimal solution achieved is really the unique global solution. So, SVM establishes a more reasonable classifier and it has a wide range of applications [3–7]. SVM determines an optimal separating hyperplane that maximizes the margin between two classes of data samples. The structural risk minimization (SRM) idea is used in SVM to improve generalization ability, but it has a high computational complexity because it requires solving a single large size QPP.

To accelerate the computational complexity of SVM, Jayadeva [8] developed a method called a novel twin support vector machine (TSVM) for binary classification. TSVM solving two smaller QPPs separately which makes the approach four times faster than the SVM that solves one large QPP. In the last decade, TSVM has been extensively studied and improved, such as twin parametric margin SVM (TPMSVM) [9], twin bounded SVM (TBSVM) [10], weighted Lagrangian TSVM (WLTSVM) [11], least squares TSVM (LSTSVM) [12–14], large scale TSVM [15], Sparse pinball TSVM [16], and so on. To implement the SRM idea in TSVM and avoid the singularity problem, some researchers [17, 18] added an extra regularization term to the objective function in TSVM primal problems. Recently some classifiers based on TSVM such as the GPin-TSVM have been proposed by Panup [19] which lower sensitivity to noise and to handle losing sparsity. The optimal solutions of GPin-TSVM are obtained by solving a pair of quadratic programming problems (QPPs) and the matrics appearing in the formulation of GPin-TSVM are positive semi-definite. It is possible that these matrics mat not be well condition in some situations. To take care of possible ill-conditioning of these matrics, we improve the classification performance of GPin-TSVM, called improved version of twin support vector machine with generalized pinball (GPin-ITSVM) by adding an extra regularization term to the objective function of GPin-TSVM in primal problems. The primary advantage of GPin-ITSVM over GPin-TSVM is that the matrices in the dual formulation of the proposed GPin-ITSVM are positive definite. Our GPin-ITSVM has the following appealing advantages:

- For pattern classification, we adding an extra regularization term to the objective function of GPin-TSVM in primal problems, called improved TSVM with generalized pinball (GPin-ITSVM) which ensures the quality yields high testing accuracy.
- The UCI benchmark datasets is chosen to show the performance of our proposed GPin-ITSVM, which is compared to the state-of-the-art TSVM, TBSVM, and GPin-TSVM. Numerical testing on datasets from UCI benchmarks illustrate that the prediction accuracy of our proposed GPin-ITSVM is superior to that of existing classifiers.
- We determine the applicability of our proposed GPin-ITSVM to Weizmann activity recognition applications compare with TSVM, TBSVM and generalized pinball TSVM (GPin-TSVM). Moreover, we have used the automatic feature extractor by use of 5 types of Convolution Neural Network (CNN) models which are ResNet50, ResNet152V2, InceptionV3, InceptionResNetV2 and Xception and

the softmax layer of CNN is replaced by binary classifier (TSVM, TBSVM, GPin-TSVM and GPin-ITSVM).

Section 2 briefly reviews TSVM, TBSVM, and generalized pinball TSVM. We present an GPin-ITSVM in Section 3. Section 4 show the efficiency of our proposed GPin-TSVM by using UCI Machine Learning Repository is compared to TSVM, TBSVM and GPin-TSVM and the application of the proposed GPin-ITSVM algorithms in human activity recognition applications. Section 5 concludes the paper.

2. Preliminaries

In this section, we present a brief description of TSVM [8], TBSVM [?] and GPin-TSVM [19] formulations.

2.1. Twin Support Vector Machine

Consider a binary classification problem of m_1 positive class and m_2 negative class. Further, let $A \in \mathbb{R}^{m_1 \times n}$ and $B \in \mathbb{R}^{m_2 \times n}$ be the matrices containing the feature vectors of the samples of class $+1$ and -1, respectively. TSVM [8] need to find two non-parallel hyperplanes defined as follows:

$$x^\top w_1 + b_1 = 0 \text{ and } x^\top w_2 + b_2 = 0 \tag{2.1}$$

where $w_1, w_2 \in \mathbb{R}^n$ and $b_1, b_2 \in \mathbb{R}$. Here, each hyperplane is closer to one of the two classes and is at least one distance from the other. Hence, the twin support vector machine (TSVM) based on hinge loss function leads to the following pair of QPPs:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Aw_1 + e_1 b_1\|^2 + c_1 e_2^\top \xi \\ \text{s.t.} \quad & -(Bw_1 + e_2 b_1) + \xi \geq e_2, \;\; \xi \geq 0, \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} \min_{w_2,b_2,\xi} \quad & \frac{1}{2}\|Bw_2 + e_2 b_2\|^2 + c_2 e_1^\top \xi \\ \text{s.t.} \quad & (Aw_2 + e_1 b_2) + \xi \geq e_1, \;\; \xi \geq 0, \end{aligned} \tag{2.3}$$

where c_1 and c_2 are positive penalty parameters, ξ is a slack variable and e_1 and e_2 are vectors of ones of appropriate dimensions. By introducing the Lagrangian multipliers α and β, the dual of QPPs (2.2) and (2.3) can be represented respectively as follows:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\alpha^\top G(H^\top H)G^\top \alpha - e_1^\top \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 e_1, \end{aligned} \tag{2.4}$$

where $G = \begin{bmatrix} B & e_1 \end{bmatrix}$, $H = \begin{bmatrix} A & e_2 \end{bmatrix}$, and

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2}\beta^\top H(G^\top G)H^\top \beta - e_2^\top \beta \\ \text{s.t.} \quad & 0 \leq \beta \leq c_2 e_2. \end{aligned} \tag{2.5}$$

A new sample point $x \in \mathbb{R}^n$ is assigned to class $i(i = +1 \text{ or } -1)$ by

$$\text{class}(i) = \arg\min_{i=1,2} \frac{|x^\top w_i + b_i|}{\|w_i\|}. \tag{2.6}$$

2.2. Twin bounded Support Vector Machine

Shao et al. [10] proposed an improved TSVM also named as twin bounded support vector machine and also used regularization term into the TSVM. The primal problems of the TBSVM is given as follows:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Aw_1+e_1b_1\|^2+c_1e_2^\top\xi+\frac{c_3}{2}\left(\|w_1\|^2+b_1^2\right) \\ \text{s.t.} \quad & -(Bw_1+e_2b_1)+\xi\geq e_2,\ \ \xi\geq 0, \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} \min_{w_2,b_2,\xi} \quad & \frac{1}{2}\|Bw_2+e_2b_2\|^2+c_2e_1^\top\xi+\frac{c_4}{2}\left(\|w_2\|^2+b_2^2\right) \\ \text{s.t.} \quad & (Aw_2+e_1b_2)+\xi\geq e_1,\ \ \xi\geq 0, \end{aligned} \tag{2.8}$$

where c_1, c_2, c_3 and c_4 are positive penalty parameters, ξ is a slack variable, and e_1 and e_2 are vectors of ones of appropriate dimensions. $\frac{1}{2}\left(\|w_1\|+b_1^2\right)$ and $\frac{1}{2}\left(\|w_2\|+b_2^2\right)$ are the extra regularization terms which minimize the training error. By introducing the Lagrangian multipliers, the dual of QPPs (2.7) and (2.8) can be represented respectively as follows:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\alpha^\top H(G^\top G+c_3I)^{-1}H^\top\alpha-e_2^\top\alpha \\ \text{s.t.} \quad & 0\leq\alpha\leq c_1e_2 \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2}\beta^\top G(H^\top H+c_4I)^{-1}G^\top\beta-e_1^\top\beta \\ \text{s.t.} \quad & 0\leq\beta\leq c_2e_1. \end{aligned} \tag{2.10}$$

Once we obtain the solutions of problems (2.9) and (2.10), a new sample point $x\in\mathbb{R}^n$ is assigned to class $i(i=+1 \text{ or } -1)$ by using (2.6) similar to the TSVM.

2.3. Twin support vector machine with generalized pinball loss (GPin-TSVM)

Panup et al. [19] develpoed TSVM using the generalized pinball loss to address the shortcomings of hinge loss. GPin-TSVM is noise insensitive, sparse and performs prediction. They obtain the following pair of QPPs:

$$\begin{aligned}
\min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Aw_1+e_1b_1\|^2+c_1e_2^\top\xi \\
\text{s.t.} \quad & -(Bw_1+e_2b_1)\geq e_2-\frac{1}{\tau_1}(\xi+e_2\epsilon_1), \\
& -(Bw_1+e_2b_1)\leq e_2+\frac{1}{\tau_2}(\xi+e_2\epsilon_2), \\
& \xi\geq 0,
\end{aligned} \tag{2.11}$$

and

$$\begin{aligned}
\min_{w_2,b_2,\xi} \quad & \frac{1}{2}\|Bw_2+e_2b_2\|^2+c_2e_1^\top\xi \\
\text{s.t.} \quad & Aw_2+e_1b_2\geq e_1-\frac{1}{\tau_1}(\xi+e_1\epsilon_1), \\
& Aw_2+e_1b_2\leq e_1+\frac{1}{\tau_2}(\xi+e_1\epsilon_2), \\
& \xi\geq 0,
\end{aligned} \tag{2.12}$$

where τ_1,τ_2,ϵ_1 and ϵ_2 are non-negative parameters, c_1 and c_2 are positive penalty parameters, ξ is a slack variable, and e_1 and e_2 are vectors of ones of appropriate dimensions. By introducing the Lagrange multipliers, the dual of QPPs (2.11) and (2.12) can be derived respectively as follows:

$$\begin{aligned}
\min_{\alpha,\lambda} \quad & \frac{1}{2}\lambda^\top Q(P^\top P)^{-1}Q^\top\lambda-\lambda^\top e_2(1+\frac{\epsilon_2}{\tau_2})+\alpha^\top e_2\left(\frac{\epsilon_1}{\tau_1}+\frac{\epsilon_2}{\tau_2}\right) \\
\text{s.t.} \quad & 0\leq\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)\alpha-\frac{\lambda}{\tau_2}\leq c_1e_2, \\
& \alpha\geq 0, \alpha-\lambda\geq 0.
\end{aligned} \tag{2.13}$$

$$\begin{aligned}
\min_{\omega,\mu} \quad & \frac{1}{2}\mu^\top P(Q^\top Q)^{-1}P^\top\mu-\mu^\top e_1(1+\frac{\epsilon_4}{\tau_4})+\omega^\top e_1\left(\frac{\epsilon_3}{\tau_3}+\frac{\epsilon_4}{\tau_4}\right) \\
\text{s.t.} \quad & 0\leq\left(\frac{1}{\tau_3}+\frac{1}{\tau_4}\right)\omega-\frac{\mu}{\tau_4}\leq c_2e_1, \\
& \omega\geq 0, \mu\geq 0.
\end{aligned} \tag{2.14}$$

Similarly, a new sample point $x\in\mathbb{R}^n$ is assigned to class $i(i=+1 \text{ or } -1)$ by using (2.6).

3. Proposed an Improved twin support vector machine with generalized pinball loss (GPin-ITSVM)

In this section, we propose an GPin-ITSVM base on GPin-TSVM. We improve the performance of GPin-TSVM by including a regularization term into the objective function. Our proposed GPin-ITSVM implements the structural risk minimization principle that also generates optimal results, while minimizing error and maximizing the the generalization ability of the classifier. We discuss GPin-ITSVM in both the linear and nonlinear cases.

3.1. Linear case

The GPin-ITSVM can be represented as follows:

$$\begin{aligned}
\min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Aw_1+e_1b_1\|^2 + c_1 e_2^\top \xi + \frac{c_3}{2}\big(\|w_1\|^2 + b_1^2\big) \\
\text{s.t.} \quad & -(Bw_1+e_2b_1) \geq e_2 - \frac{1}{\tau_1}(\xi + e_2\epsilon_1), \\
& -(Bw_1+e_2b_1) \leq e_2 + \frac{1}{\tau_2}(\xi + e_2\epsilon_2), \\
& \xi \geq 0,
\end{aligned} \tag{3.1}$$

and

$$\begin{aligned}
\min_{w_2,b_2,\xi} \quad & \frac{1}{2}\|Bw_2+e_2b_2\|^2 + c_2 e_1^\top \xi + \frac{c_4}{2}\big(\|w_2\|^2 + b_2^2\big) \\
\text{s.t.} \quad & Aw_2+e_1b_2 \geq e_1 - \frac{1}{\tau_1}(\xi + e_1\epsilon_1), \\
& Aw_2+e_1b_2 \leq e_1 + \frac{1}{\tau_2}(\xi + e_1\epsilon_2), \\
& \xi \geq 0
\end{aligned} \tag{3.2}$$

where $\tau_1, \tau_2, \epsilon_1$ and ϵ_2 are non-negative parameters, c_1, c_2, c_3 and c_4 are positive penalty parameters, ξ is a slack variable, and e_1 and e_2 are vectors of ones of appropriate dimensions. To obtain the solutions of (3.1) and (3.2), we convert them to the dual form. We consider (3.1) for this purpose and introduce the corresponding Lagrange function with Lagrange multipliers $\alpha \geq 0, \beta \geq 0$ and $\gamma \geq 0$ as follows:

$$\begin{aligned}
\mathcal{L}(w_1,b_1,\xi,\alpha,\beta,\gamma) =& \frac{1}{2}(Aw_1+e_1b_1)^\top(Aw_1+e_1b_1) + c_1e_2^\top\xi + \frac{c_3}{2}\big(\|w_1\|^2+b_1^2\big) \\
& - \alpha^\top\Big(-(Bw_1+e_2b_1) - e_2 + \frac{1}{\tau_1}(\xi+e_1\epsilon_1)\Big) - \beta^\top\xi \\
& - \gamma^\top\Big((Bw_1+e_2b_1) + e_2 + \frac{1}{\tau_2}(\xi+e_2\epsilon_2)\Big).
\end{aligned} \tag{3.3}$$

Applying the the KarushKuhnTucker (KKT) optimality conditions [20] on (3.3), we get:

$$\frac{\partial \mathcal{L}}{\partial w_1} = A^\top(Aw_1 + e_1 b_1) + c_3 w_1 + \alpha^\top B - \gamma^\top B = 0, \tag{3.4}$$

$$\frac{\partial \mathcal{L}}{\partial b_1} = e_1^\top(Aw_1 + e_1 b_1) + c_3 b_1 + \alpha^\top e_2 - \gamma^\top e_2 = 0, \tag{3.5}$$

$$\frac{\partial \mathcal{L}}{\partial \xi} = c_1 e_2 - \frac{\alpha}{\tau_1} - \beta - \frac{\gamma}{\tau_2} = 0, \tag{3.6}$$

$$\alpha^\top\big(-(Bw_1 + e_2 b_1) - e_2 + \frac{1}{\tau_1}(\xi + e_1\epsilon_1)\big) = 0, \tag{3.7}$$

$$\beta^\top \xi = 0, \tag{3.8}$$

$$\gamma^\top\big((Bw_1 + e_2 b_1) + e_2 + \frac{1}{\tau_2}(\xi + e_2\epsilon_2)\big) = 0. \tag{3.9}$$

By combining equations (3.4) and (3.5), we can obtain

$$\begin{bmatrix} A^\top \\ e_1^\top \end{bmatrix} \begin{bmatrix} A & e_1 \end{bmatrix} \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + c_3 \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} B^\top \\ e_2^\top \end{bmatrix} (\alpha - \gamma) = 0. \tag{3.10}$$

Define $\lambda = \alpha - \gamma, P = \begin{bmatrix} A & e_1 \end{bmatrix}$ and $Q = \begin{bmatrix} B & e_2 \end{bmatrix}$. With these notations, the equation (3.10) can be rewritten as follows:

$$P^\top P \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + c_3 \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + Q^\top \lambda = 0, \text{ i.e., } \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(P^\top P + c_2 I)^{-1} Q^\top \lambda. \tag{3.11}$$

Using equation (3.3) and the above KKT optimality conditions, we can obtain the dual of (3.1) as follows:

$$\begin{aligned} \min_{\alpha,\lambda}\ & \frac{1}{2}\lambda^\top Q(P^\top P + c_3 I)^{-1} Q^\top \lambda - \lambda^\top e_2(1 + \frac{\epsilon_2}{\tau_2}) + \alpha^\top e_2\left(\frac{\epsilon_1}{\tau_1} + \frac{\epsilon_2}{\tau_2}\right) \\ \text{s.t. }\ & c_1 e_2 - \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)\alpha - \beta - \frac{\lambda}{\tau_2} = 0, \\ & \alpha \geq 0, \beta \geq 0, \alpha - \lambda \geq 0. \end{aligned} \tag{3.12}$$

In a similar manner, we can obtain the dual problem of (3.2) as follows:

$$\begin{aligned} \min_{\omega,\mu}\ & \frac{1}{2}\mu^\top P(Q^\top Q + c_4 I)^{-1} P^\top \mu - \mu^\top e_1(1 + \frac{\epsilon_2}{\tau_2}) + \omega^\top e_1\left(\frac{\epsilon_1}{\tau_1} + \frac{\epsilon_2}{\tau_2}\right) \\ \text{s.t. }\ & c_2 e_1 - \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)\omega - \eta - \frac{\mu}{\tau_2} = 0, \\ & \omega \geq 0, \eta \geq 0, \omega - \mu \geq 0, \end{aligned} \tag{3.13}$$

where $\omega \geq 0, \eta \geq 0$ and $\mu \geq 0$ are Lagrange multipliers. After the solution of the QPPs (3.12) and (3.13) are found, the optimal separating hyperplanes are given by:

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(P^\top P + c_3 I)^{-1} Q^\top \lambda$$

and

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (Q^\top Q + c_4 I)^{-1} P^\top \mu. \tag{3.14}$$

A new sample $x \in \mathbb{R}^n$ is assigned to class $i(i = +1$ or $-1)$ depending on which of the two hyperplanes in (2.1) is closer to x, i.e.,

$$\text{class}(i) = \arg\min_{i=1,2} \frac{|x^\top w_i + b_i|}{\|w_i\|}. \tag{3.15}$$

3.2. Nonlinear case

We use a kernel trick [21] to extend the linear GPin-ITSVM to the nonlinear case. Some nonlinear kernel functions map the input data into a high-dimensional feature space. If $K(\cdot,\cdot)$ is the predefined kernel function, then the nonparallel hyperplanes in the kernel-generated space can be written as follows:

$$K(x^\top, D^\top)u_1 + b_1 = 0 \text{ and } K(x^\top, D^\top)u_2 + b_2 = 0, \tag{3.16}$$

where $D = \begin{bmatrix} A_{m_1\times n} \\ B_{m_1\times n}\end{bmatrix}, u_1, u_2 \in \mathbb{R}^m$ and K is an arbitrary kernel function. The corresponding problems for the nonlinear case of the problem (3.1) and (3.2) are

$$\begin{aligned}
\min_{u_1,b_1,\xi} \quad & \frac{1}{2}\|K(A,D^\top)u_1 + e_1b_1\| + c_1e_2^\top\xi + \frac{c_3}{2}\big(\|w_1\|^2 + b_1^2\big)\\
\text{s.t.} \quad & -(K(B,D^\top)u_1 + e_2b_1) \ge e_2 - \frac{1}{\tau_1}(\xi + e_2\epsilon_1),\\
& -(K(B,D^\top)u_1 + e_2b_1) \le e_2 + \frac{1}{\tau_2}(\xi + e_2\epsilon_2),\\
& \xi \ge 0,
\end{aligned} \tag{3.17}$$

and

$$\begin{aligned}
\min_{u_2,b_2,\xi} \quad & \frac{1}{2}\|K(B,D^\top)u_2 + e_2b_2\| + c_2e_1^\top\xi + \frac{c_4}{2}\big(\|w_2\|^2 + b_2^2\big)\\
\text{s.t.} \quad & K(A,D^\top)u_2 + e_1b_2 \ge e_1 - \frac{1}{\tau_1}(\xi + e_1\epsilon_1),\\
& K(A,D^\top)u_2 + e_1b_2 \le e_1 + \frac{1}{\tau_2}(\xi + e_1\epsilon_2),\\
& \xi \ge 0.
\end{aligned} \tag{3.18}$$

By introducing the Lagrange function in QPP (3.17), we get that

$$\begin{aligned}
\mathcal{L}(u_1, b_1,\xi, \alpha, \beta, \gamma) = {} & \frac{1}{2}(K(A,D^\top)u_1 + e_1b_1)^\top(K(A,D^\top)u_1 + e_1b_1) + c_1e_2^\top\xi\\
& + \frac{c_3}{2}\big(\|u_1\|^2 + b_1^2\big) - \alpha^\top\Big(-(K(B,D^\top)u_1 + e_2b_1) - e_2 + \frac{1}{\tau_1}(\xi + e_1\epsilon_1)\Big)\\
& - \beta^\top\xi - \gamma^\top\Big((K(B,D^\top)u_1 + e_2b_1) + e_2 + \frac{1}{\tau_2}(\xi + e_2\epsilon_2)\Big).
\end{aligned} \tag{3.19}$$

Applying the KKT optimality conditions on (3.19), we get:

$$\frac{\partial \mathcal{L}}{\partial u_1} = K(A,D^\top)^\top(K(A,D^\top)u_1 + e_1b_1) + c_3u_1 + \alpha^\top K(B,D^\top) - \gamma^\top K(B,D^\top) = 0, \tag{3.20}$$

$$\frac{\partial \mathcal{L}}{\partial b_1} = e_1^\top(K(A,D^\top)u_1 + e_1b_1) + c_3b_1 + \alpha^\top e_2 - \gamma^\top e_2 = 0, \tag{3.21}$$

$$\frac{\partial \mathcal{L}}{\partial \xi} = c_1e_2 - \frac{\alpha}{\tau_1} - \beta - \frac{\gamma}{\tau_2} = 0, \tag{3.22}$$

$$\alpha^\top\big(-(K(B,D^\top)u_1 + e_2b_1) - e_2 + \frac{1}{\tau_1}(\xi + e_1\epsilon_1)\big) = 0, \tag{3.23}$$

$$\beta^\top\xi = 0, \tag{3.24}$$

$$\gamma^\top\big((K(B,D^\top)u_1 + e_2b_1) + e_2 + \frac{1}{\tau_2}(\xi + e_2\epsilon_2)\big) = 0. \tag{3.25}$$

By combining equations (3.20) and (3.21), we can obtain

$$\begin{bmatrix} K(A,D^\top)^\top \\ e_1^\top \end{bmatrix} \begin{bmatrix} K(A,D^\top) & e_1 \end{bmatrix} \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} + c_3 \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} K(B,D^\top)^\top \\ e_2^\top \end{bmatrix}(\alpha - \gamma) = 0. \tag{3.26}$$

Define $H = \begin{bmatrix} K(A,D^\top) & e_1 \end{bmatrix}$ and $G = \begin{bmatrix} K(B,D^\top) & e_2 \end{bmatrix}$. With these notations, the equation (3.26) can be rewritten as follows:

$$H^\top H \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} + c_3 \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} + G^\top(\alpha - \gamma) = 0, \text{ i.e., } \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(H^\top H + c_3I)^{-1}G^\top(\alpha - \gamma).$$

Finally, the dual of QPP (3.17) can be derived as:

$$\begin{aligned} \min_{\alpha,\gamma}\ & \frac{1}{2}(\alpha-\gamma)^\top G(H^\top H + c_3I)^{-1}G^\top(\alpha-\gamma) - (\alpha-\gamma)^\top e_2(1+\frac{\epsilon_2}{\tau_2}) + \alpha^\top e_2\Big(\frac{\epsilon_1}{\tau_1} + \frac{\epsilon_2}{\tau_2}\Big) \\ \text{s.t. } & \frac{\alpha}{\tau_1} + \frac{\gamma}{\tau_2} \le c_1e_1, \\ & \alpha \ge 0, \alpha - \gamma \ge 0. \end{aligned} \tag{3.27}$$

Similarly, we can obtain the dual of QPP (3.18) as follows:

$$\begin{aligned} \min_{\omega,\mu}\ & \frac{1}{2}(\omega-\mu)^\top H(G^\top H + c_4I)^{-1}G^\top(\omega-\mu) - (\omega-\mu)^\top e_1(1+\frac{\epsilon_2}{\tau_2}) + \omega^\top e_1\Big(\frac{\epsilon_1}{\tau_1} + \frac{\epsilon_2}{\tau_2}\Big) \\ \text{s.t. } & \frac{\omega}{\tau_1} + \frac{\mu}{\tau_2} \le c_2e_1, \\ & \omega \ge 0, \omega - \mu \ge 0. \end{aligned} \tag{3.28}$$

After the solutions of QPPs (3.27) and (3.28) are found, the optimal separating hyperplanes are given by:

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(H^\top H + c_3I)^{-1}G^\top(\alpha - \gamma)$$

and

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (G^\top G + c_4 I)^{-1} H^\top (\omega - \mu). \tag{3.29}$$

Once we obtain the solutions (3.29), a new sample point $x \in \mathbb{R}^n$ is assigned to class $i(i = +1 \text{ or } -1)$ by

$$\text{class}(i) = \arg\min_{i=1,2} \frac{|K(x, D^\top)^\top u_i + b_i|}{\|u_i\|}. \tag{3.30}$$

.

Algorithm 1 Our GPin-ITSVM can be obtained by performing the following steps:

Input: Parameter values for $c_1, c_2, c_3, c_4, \tau_1, \tau_2, \epsilon_1, \epsilon_2$ and σ.

1: Choose a kernel function K.
2: Define $H = \begin{bmatrix} K(A, D^\top) & e_1 \end{bmatrix}$ and $G = \begin{bmatrix} K(B, D^\top) & e_2 \end{bmatrix}$.
3: Solve the QPPs (3.27) and (3.28), and obtain their solutions.
4: Calculate the perpensicular distances $\frac{|K(x,D^\top)^\top u_1+b_1|}{\|u_1\|}$ and $\frac{|K(x,D^\top)^\top u_2+b_2|}{\|u_2\|}$ for a new sample point.
5: Assign the new sample point to class $i(i = +1 \text{ or } -1)$ by (3.30).

---

4. Numerical experiments

In this section, classification performance of the proposed method in terms of accuracy is compared with other related methods, viz. hinge loss twin support vector machine (TSVM), hinge loss twin bounded support vector machine (TBSVM) and generalized pinball loss TSVM (GPin-TSVM) on UCI Machine Learning Repository [22] and Human Activity Recognition applications have been proposed. We have used 10-fold cross validation for all experiments.

All experiments are implemented in Python 3.9.5. on Windows 8 running on a 1.9 GHz laptop with 4 GB RAM with system configuration Intel Core i5+ Duo CPU E7500 (2.93 GHz) with 4 GB of RAM. To derive the nonlinear case, we use the radial basis function kernel $K(x, y) = \exp\{-\frac{\|x-y\|^2}{2\sigma}\}$.

4.1. Parameters Selection

In each algorithm, we used the grid search technique [24] to optimize the tradeoff parameters and kernel parameter. We have selected values of parameter $\sigma, c_i, i = 1, \ldots, 4$ from the set $\{10^i | i = -2, -1, 0, 1, 2\}$. The values of τ_1 and τ_2 are taken from $\{0.5, 0.75, 1\}$, The value of ϵ_1 and ϵ_2 in the experiments are chosen from $\{0.1, 0.25, 0.5, 0.75, 1\}$ and the experimental results are the average accuracy and standard deviation, and each experiments time consists of testing time, and the unit of time is seconds. The bold values indicate best mean of accuracy (in %).

TABLE 1. Description of UCI datasets.

Datasets	#Feature	#Sample	Imbalance ratio
Breast	10	110	1.23
Sonar	60	208	1.14
Appendicitis	7	106	4.05
Spectf Heart	44	267	3.73
Monk-2	6	432	1.12
Monk-3	6	432	1.12
Australian	14	690	1.25
Heart Statlog	13	270	1.25
Ionosphere	33	351	1.79
Bupa	6	345	1.38
Diabetes	8	768	1.87
Pima-Indian	8	768	1.87

4.2. UCI DATASETS

We also perform experiments on 12 benchmark datasets available at the UCI machine learning repository [22]. In classification problems, imbalanced datasets lead to erroneous classification. Imbalance ratio [23] defined as the ratio of the number of data points in the majority class to the number of samples in the minority class.

$$\text{Imbalance ratio} = \frac{\text{number of data points on 'majority class'}}{\text{number of data points on 'minority class'}}.$$

The descriptions of the datasets are given in Table 1. The optimal parameters using in linear and nonlinear cases are shown in Tables 3 and Table 6, respectively.

TABLE 2. Accuracy obtained on the UCI datasets by 10-fold cross validation using linear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
	Time (s)	Time (s)	Time (s)	Time (s)
Breast	69.02±8.61	69.02±8.61	70.83±9.78	**72.65±12.86**
	0.0145	0.0162	0.0465	0.0451
Sonar	67.76±6.16	78.33±7.27	74.95±9.15	78.33±6.72
	0.0389	0.0360	0.1782	0.1769
Appendicitis	85.73±16.79	85.09±10.11	87.64±14.07	**88.82±12.45**
	0.0160	0.0165	0.0614	0.0510
Spectf Heart	78.03±8.53	80.30±8.82	80.32±8.08	**80.76±9.36**
	0.0649	0.0375	0.3931	0.3732
Monk-2	75.92±8.30	76.38±8.78	81.73±5.47	81.73±5.64
	0.1743	0.1688	1.5150	1.6041
Monk-3	76.35±5.68	76.35±5.68	80.29±5.49	**80.53±6.39**
	0.1138	0.0887	0.9924	1.2302
Australian	86.09±3.56	86.23±3.38	86.81±3.69	86.81±3.69
	0.5241	0.5180	6.1842	6.0457
Average Accuracy	76.98	78.81	80.36	81.37
Average Time	0.1352	0.1259	1.3386	1.3608

TABLE 3. The optimal parameters in linear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
	c_1, c_2	c_1, c_2, c_3, c_4	$c_1, c_2,$ $\tau_1, \tau_2, \epsilon_1, \epsilon_2$	$c_1, c_2, c_3, c_4,$ $\tau_1, \tau_2, \epsilon_1, \epsilon_2$
Breast	0.01,0.01	0.01, 0.01, 0.1,0.1	0.01,0.01, 0.75,0.5,0.25,0.25	1,10,0.1,0.1, 1,1,0.1,0.25
Sonar	0.1,0.01	0.01, 0.01, 10,1	0.1,1, 0.75,1,0.25,0.25	0.1,0.1,10,1, 1,0.75,0.5,0.75
Appendicitis	0.01,0.01	1,1,100,1	0.1,0.1, 1,1,0.75,0.5	0.1,0.1,10,1, 1,0.75,0.5,0.5
Spectf Heart	0.01,0.1	10,10,100,0.01	0.1,0.1, 1,0.5,0.5,0.5	0.1,0.1,10,0.01, 1,0.75,0.5,0.25
Monk-2	0.1,0.1	0.1,0.1,10,1	0.1,0.1, 0.75,1,0.5,0.25	0.1,0.1,10,1, 0.75,1,0.5,0.25
Monk-3	0.01,0.01	0.01,0.01,0.01,0.01	0.01,0.01, 0.75,0.5,0.25,0.25	1,10,0.1,0.1, 1,1,0.1,0.5
Australian	1,0.1	1,0.1,0.01,0.01	1,0.1, 1,0.5,0.1,0.1	1,0.1,0.01,0.1, 1,0.5,0.1,0.1

TABLE 4. Average ranks of different algorithms on UCI dataset with linear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
Breast	3.5	3.5	2	1
Sonar	4	1.5	3	1.5
Appendicitis	3	4	2	1
Spectf Heart	4	3	2	1
Monk-2	4	3	1.5	1.5
Monk-3	3.5	3.5	2	1
Australian	4	3	1.5	1.5
Average Rank	3.71	3.07	2.00	1.21

Table 2 shows the mean and standard deviation of evaluation accuracy for a linear kernel on 7 distinct UCI datasets using TSVM, TBSVM, GPin-TSVM, and our proposed GPin-ITSVM. We can see that the our proposed GPin-ITSVM slightly outperforms GPin-TSVM in terms of accuracy. As can be seen from Table 2, TSVM and TBSVM have much faster computational complexity than other models, while our proposed GPin-ITSVM and GPin-TSVM similar computational complexity. Results for a nonlinear kernel on 6 distinct UCI datasets are given in Table 5. For the nonlinear kernel, the accuracy of the GPin-ITSVM classifier for 6 distinct UCI datasets is also greater than the others classifier. At the same time, TBSVM takes the least computational complexity on all UCI datasets in both lineat and nonlinear case.

Table 5. Accuracy obtained on the UCI datasets by 10-fold cross validation using nonlinear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
	Time (s)	Time (s)	Time (s)	Time (s)
Appendicitis	86.09±13.93	87.00±12.68	87.73±12.31	**8**7.91±11.21
	0.0310	0.0259	0.0741	0.0640
Heart Statlog	84.44±7.73	83.70±9.69	84.81±7.67	84.81±7.11
	0.0730	0.0709	0.3804	0.4901
Ionosphere	96.02±2.58	96.02±2.61	95.15±4.05	95.16±3.39
	0.1375	0.1484	1.0246	1.0515
Bupa	70.43±8.43	71.88±8.92	71.59±9.97	**72.75±8.48**
	0.1474	0.1006	0.5793	0.6299
Diabetes	76.06±5.46	76.97±4.48	76.84±5.87	**77.23±5.24**
	1.1095	1.0586	9.6623	11.7026
Pima	76.70±3.31	77.09±3.57	76.83±3.54	**77.22±3.70**
	1.1001	1.1887	11.0425	10.3058
Average Accuracy	81.62	82.11	82.16	82.51
Average Time	0.4331	0.4322	3.8696	4.3081

Table 6. The optimal parameters in nonlinear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
	c_1, c_2, σ	$c_1, c_2, c_3, c_4, \sigma$	c_1, c_2, σ, $\tau_1, \tau_2, \epsilon_1, \epsilon_2$	$c_1, c_2, c_3, c_4, \sigma$, $\tau_1, \tau_2, \epsilon_1, \epsilon_2$
Appendicitis	0.01,0.01,0.1	0.1,0.1,10,1,0.1	1,1,0.1, 1,0.75,0.5,1	0.1,0.1,10,1,0.1, 1,0.75,0.5,0.5
Heart Statlog	1,1,0.1	0.1,0.1,10,1,0.1	1,10,0.01, 1,0.5,0.1,0.5	0.1,0.1,10,1,0.01, 0.75,0.5,0.5,0.25
Ionosphere	0.1,1,0.1	0.1,1,0.01,0.1,0.1	0.1,0.1,0.01, 1,1,0.5,0.5	0.1,0.1,0.1,0.1,0.01, 1,1,0.5,0.25
Bupa	0.1,0.1,0.01	0.1,0.1,0.01,0.1,0.01	0.1,0.1,0.01, 1,0.75,0.5,0.25	0.1,0.1,0.01,0.1,0.01, 1,1,0.5,0.25
Diabetes	0.1,0.1,0.01	0.1,0.1,0.01,0.1,0.1	0.1,0.1,0.01, 1,0.75,0.5,0.25	0.1,0.1,0.01,0.1,0.01, 1,1,0.5,0.25
Pima	0.1,0.1,0.1	0.1,0.1,0.01,0.01,0.01	0.1,0.1,0.01, 1,0.75,0.5,0.25	0.1,0.1,0.01,0.1,0.01, 1,1,0.5,0.25

Table 7. Average ranks of different algorithms on UCI dataset with nonlinear case.

Datasets	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
Appendicitis	4	3	2	1
Heart Statlog	3	4	1.5	1.5
Ionosphere	1.5	1.5	4	3
Bupa	4	2	3	1
Diabetes	4	2	3	1
Pima	4	2	3	1
Average Rank	3.42	2.42	2.75	1.58

FIGURE 1. Samples from Weizmann activity recognition dataset.

4.3. Statistical Analysis

The Friedman test with post hoc tests [25] is primarily used in the following section to evaluate the statistical significance of the proposed GPin-ITSVM in comparison to TSVM, TBSVM, and GPin-TSVM. The Friedman test is a statistical test method that employs ranks the algorithms differently for each data set, with the best method receiving the smallest rank value. The accuracy of the related classifiers on each dataset is ranked, and the classifier with the highest accuracy has the smallest rank r_i. The Friedman statistic is

$$\mathcal{X}_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right],$$

$$F_F = \frac{(N-1)\mathcal{X}_F^2}{N(k-1) - \mathcal{X}_F^2},$$

where $R_j = \frac{1}{N}\sum_{j=1}^{N} r_j$, F_F is F-distribution with a degree of freedom $(k-1)(N-1)$, k is the number of algorithms and N number of datasets.

4.3.1. Linear case

Under the null hypothesis, which states that all the algorithms are equivalent. According to the Table 4, we obtain $\mathcal{X}_F^2 = 15.29$ and $F_F = 16.07$. The critical value of $F(3, 18)$ for a significance level of 0.05 is 3.16, and $16.07 > 3.16$, so the null hypothesis is rejected. That is, there significant difference among the 4 classifiers. Furthermore, as shown in Table 4, the proposed Improve Pin-GTSVM achieved a lower average rank. On the UCI datasets, the classification performance of the proposed GPin-ITSVM outperforms the other classifiers.

4.3.2. Nonlinear case

From Table 7, we have $\mathcal{X}_F^2 = 9.43$ and $F_F = 5.50$. We see that the critical value of $F(3, 15)$ for a significance level of 0.05 is 3.29 and $5.50 > 3.29$, so the null hypothesis is rejected. This imply that there are significant difference among the four algorithms. From Table 7, it is seen that the proposed Improve Pin-GTSVM achieved a lower average rank, that is it outperforms the other classifiers.

4.4. Human Activity Recognition

In this section, we discuss the application of the proposed GPin-ITSVM to the human activity recognition. Human activity recognition is a hot topic in computer vision research. We conduct experiments on the well-known Weizmann activity recognition dataset [26]. This data consists of 84 videos that correspond to ten actions, which were performed by nine different people. The ten actions are walk (walking), run (running), jump (jumping), side (striding sideways), bend (bending), jack (jumping-back), skipping, pjump (jumping in place) wave1 (waving with one hand) and wave2 (waving with both hands). Figure 1 illustrates some examples of actions from the Weizmann dataset. We compare the performance between TSVM, TBSVM, GPin-TSVM, and GPin-ITSVM on a pair of activity classes.

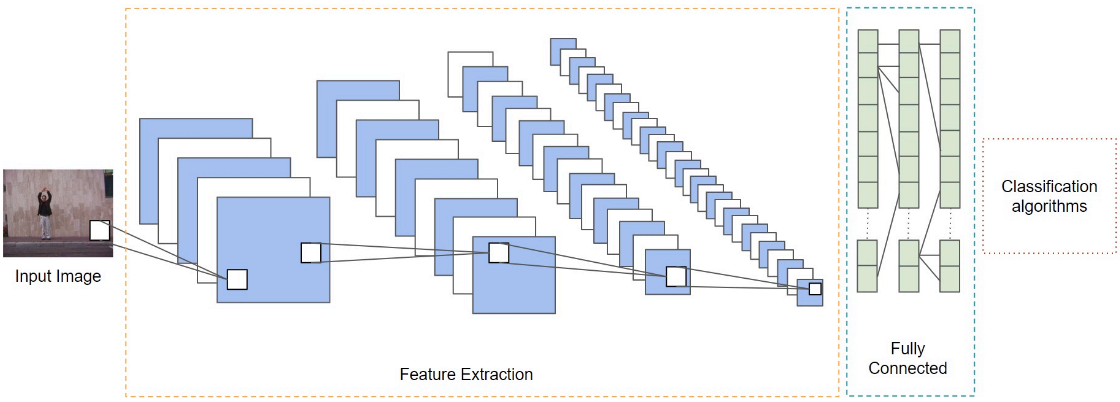
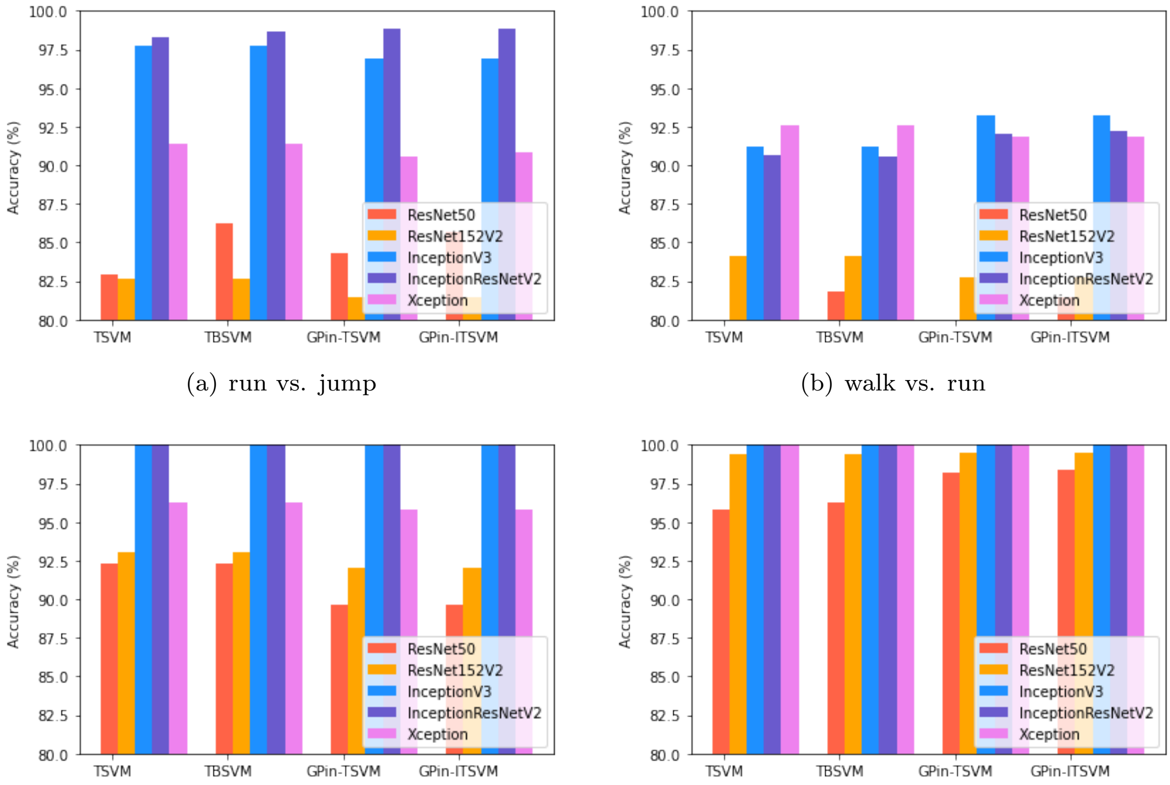


Figure 2. Illustration of CNN model.

One of the most important aspects of our cognition system success is feature extraction. Traditional feature extraction by hand is a tedious and time-consuming procedure that does not work with raw images, but features can be recovered directly from raw images using automatic extraction algorithms. We have used an automatic extraction algorithm called Convolutional Neural Networks (CNN) for feature extraction. CNN utilize local feature detectors applied to the entire image to measure the correspondence between individual image patches and signature patterns in the training set. In this subsection, we combines a powerful CNN with proposed algorithm for human activity recognition using the Weizmann activity recognition dataset, where TSVM, TBSVM, GPin-TSVM, and proposed GPin-ITSVM are a binary classifier. In addition, we considered 5 types of CNN models which are ResNet50, ResNet152V2, InceptionV3, InceptionResNetV2 and Xception. According to the Keras website, these proposed 5 CNN Models are highly accurate. Keras is the most widely used Python-based deep learning framework. Utilize the TensorFlow platform's full deployment capabilities. Moreover, we compare the performance between TSVM, TBSVM, GPin-TSVM, and proposed GPin-ITSVM on Weizmann activity recognition dataset. An illustration of CNN model is shown in Figure 2. The influence of optimal bianary classifier parameters on the classification results of Weizmann activity recognition dataset has been presented in Table 8.



(a) run vs. jump

(b) walk vs. run

(c) jump vs. pjump

(d) wave1 vs. wave2

Figure 3. The classification performance of TSVM, Pin-SVM, GPin-TSVM and GPin-ITSVM on Weizmann activity recognition datasets.

The experimental results in Figure 3 demonstrate that the recognition accuracy of 100% over jump vs. pjump and wave1 vs. wave2 datasets are InceptionV3 and Inception-ResNetV2 CNN combined with our GPin-ITSVM, TSVM, TBSVM and GPin-TSVM. However, InceptionResNetV2 combined with binary classification models have higher recognition accuracy than other CNN models for run vs. jump dataset, i.e., Inception-ResNetV2 CNN combined with our GPin-ITSVM and GPin-TSVM provided recognition accuracy of 98.80% which is greater than the recognition accuracy of 98.32% and 98.68% of InceptionResNetV2 CNN combined with TSVM and TBSVM classifier, respectively. Finally, the maximum recognition accuracy for walk vs. run dataset is 93.24% which is given by InceptionV3 CNN combined with our GPin-ITSVM and GPin-TSVM, while Xception CNN have higher recognition accuracy than other CNN models for TSVM and TBSVM, i.e., 92.55%. After analysing the results, we can observe that the CNN models combined with the our GPin-ITSVM presented a better performance when compared with the others, achieving higher recognition accuracy. It can be observed in this situation that the CNN combined with our GPin-ITSVM is more useful from the others when compared with the TSVM and TBSVM.

TABLE 8. The optimal parameters for Weizmann activity recognition datasets.

Datasets	CNN models	TSVM	TBSVM	GPin-TSVM	GPin-ITSVM
		c_1, c_2	c_1, c_2, c_3, c_4	$c_1, c_2,$ $\tau_1, \tau_2, \epsilon_1, \epsilon_2$	$c_1, c_2, c_3, c_4,$ $\tau_1, \tau_2, \epsilon_1, \epsilon_2$
run vs. jump	ResNet50	0.01,0.01	0.01, 0.01, 1,1	0.1,0.1, 1,0.75,0.1,0.1	0.01,0.01,1,1, 1,0.5,0.1,0.1
	ResNet152V2	0.01,0.01	0.01, 0.01, 0.01,0.01	0.01,0.01, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,0.75,0.1,0.1
	InceptionV3	0.1,0.1	0.01, 0.01, 0.01,0.01	0.1,0.1, 1,0.75,0.1,0.1	0.01,0.01,1,1, 1,0.5,0.1,0.1
	InceptionResNetV2	0.1,0.1	0.01, 0.01, 1,1	0.1,0.1, 1,0.75,0.1,0.1	0.01,0.01,1,1, 1,0.5,0.1,0.1
	Xception	0.01,0.01	0.01, 0.01, 0.01,0.01	0.1,0.1, 1,0.75,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
walk vs. run	ResNet50	0.1,1	1, 1, 100,100	1,1, 1,1,0.1,0.1	1,1,100,100, 1,1,0.1,0.1
	ResNet152V2	0.1,0.1	1, 1, 0.01,0.01	0.01,0.01, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
	InceptionV3	1,1	0.1, 0.01, 0.01,0.01	0.01,0.01, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
	InceptionResNetV2	0.1,0.1	0.1, 0.1, 0.01,0.01	0.1,0.1, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.25,0.25
	Xception	0.1,0.1	1, 1, 0.01,0.01	0.1,0.1, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
jump vs. pjump	ResNet50	0.01,0.01	0.01, 0.01, 0.1,0.1	1,1, 1,1,0.1,0.1	0.01,0.01,0.1,0.1, 1,1,0.1,0.1
	ResNet152V2	1,1	1,1, 1,1	1,1, 1,1,0.1,0.1	0.01,0.01,0.1,0.1, 1,1,0.1,0.1
	InceptionV3	0.1,0.1	0.01, 0.01, 1,1	0.01,0.01, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
	InceptionResNetV2	0.01,0.01	0.01, 0.01, 1,1	0.01,0.01, 1,1,0.1,0.1	0.01,0.01,1,1, 1,1,0.1,0.1
	Xception	0.01,0.01	0.1, 0.1, 1,1	1,1, 1,1,0.1,0.1	0.01,0.01,1,1, 1,1,0.1,0.1
wave1 vs. wave2	ResNet50	0.1,0.1	1, 1, 1,1	0.1,0.1, 1,1,0.75,0.75	0.1,0.1,1,1, 1,1,0.75,0.75
	ResNet152V2	0.1,0.1	0.1, 0.1, 0.1,0.1	0.1,0.1, 1,1,0.1,0.1	0.1,0.1,1,1, 1,1,0.1,0.1
	InceptionV3	1,1	0.1, 0.01, 0.01,0.01	0.01,0.01, 1,1,0.1,0.1	0.1,0.1,0.01,0.01, 1,1,0.1,0.1
	InceptionResNetV2	0.1,0.1	0.1, 0.1, 1,1	0.1,0.1, 1,1,0.1,0.1	0.1,0.1,1,1, 1,1,0.1,0.1
	Xception	0.1,0.1	1, 1, 0.01,0.01	0.1,0.1, 1,1,0.1,0.1	0.1,0.1,1,1, 1,1,0.1,0.1

5. Conclusions

In this paper, an improved version of twin support vector machine with generalized pinball (GPin-ITSVM) is proposed. By using a new improved version of twin support vector machine with generalized pinball, we reformulated the GPin-ITSVM problems as a quadratic programming problems. The primary advantage of GPin-ITSVM is that avoids the singularity problem when solving the dual quadratic programming problems. Numerical experiments are carried out on 12 UCI benchmark datasets to investigate the validity of our proposed algorithm. The results show that the our GPin-ITSVM outperforms others. Furthermore, this novel method can be used in a variety of fields, including human activity recognition. In addition, the use of this approach in Weizmann activity recognition applications is investigated, and the automatic feature extractor makes use of 5 types of Convolution Neural Network (CNN) models which are ResNet50, ResNet152V2, InceptionV3, InceptionResNetV2 and Xception. The results demonstrated that the our GPin-ITSVM is beneficial for the recognition accuracy.

References

[1] C. Cortes, V. Vapnik, Support-vector networks, Mach Learn 20(1995), 273-297.

[2] V. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10(5)(1999) 988–999.

[3] K. G. Manosha Chathuramali, R. Rodrigo, Faster human activity recognition with SVM, Proc of Advances in ICT for Emerging Regions (ICTer) (2012), 197-203. Available from: https://doi.org/10.1109/ICTer.2012.6421415

[4] A. Mohammad, T. Alwadan, O. Almomani, Arabic Text Categorization Using Support vector machine, Nave Bayes and Neural Network, GSTF Journal on Computing (JoC), 5(2016). Available from: https://doi.org/10.7603/s40601-016-0016-9

[5] Z. Mehmood, T. Mahmood, Dr. Javid, Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine, Applied Intelligence, 48(1)(2018) 166–181.

[6] B. Richhariya, M. Tanveer, EEG signal classification using universum support vector machine, Expert Syst Appl. 106(2018) 169–182.

[7] V. Parameswari, S. Pushpalatha, Human Activity Recognition using SVM and Deep Learning, European Journal of Molecular & Clinical Medicine 7(4)(2020) 1984–1990.

[8] Jayadeva, R. Khemchandani, S. Chandra, Twin support vector machines for pattern classification, IEEE Trans Pattern Anal Mach Intell 29(5)(2007) 905–910.

[9] X. Peng, TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition, Pattern Recognit. 44(10)(2011) 2678–2692.

[10] Y. Shao, C. Zhang, X. Wang, N. Deng, Improvements on twin support vector machines, IEEE Trans. Neural Networks 22(6)(2011) 962–968.

[11] Y. Shao, W. Chen, C. Zhang, X. Wang, N. Deng, An efficient weighted lagrangian twin support vector machine for imbalanced data classification, Pattern Recognit. 47(9)(2014) 3158–3167.

[12] M.A. Kumar, M. Gopal, Application of smoothing technique on twin support vector machines, Pattern Recognit. Lett. 29(13)(2008) 1842–1848.

[13] M. A. Kumar, R. Khemchandani, M. Gopal, S. Chandra, Knowledge based least squares twin support vector machines, Inform. Sci. 180(23)(2010) 4606–4618.

[14] M.A. Ganaie, M. Tanveer, LSTSVM classifier with enhanced features from pre-trained functional link network, Applied Soft Computing Journal 93(2020). Available from: https://doi.org/10.1016/j.asoc.2020.106305

[15] Y. Tian, Y. Ping, Large-scale linear nonparallel support vector machine solver, Neural Netw. 50(2014) 166–174.

[16] M. Tanveer, A. Tiwari, R. Choudhary, S. Jalan, Sparse pinball twin support vector machines, Applied Soft Computing Journal 78(2019) 164–175.

[17] H. Wang, Z. Zhou, Y. Xu, An improved v-twin bounded support vector machine, Appl Intell. 48(3)(2018) 1–13.

[18] M. Tanveer, T. Rajani, M. A. Ganaie, Improved sparse pinball twin SVM, Proc. of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy. October 6-9, 2019. Available from: https://doi.org/10.1109/SMC.2019.8914642

[19] W. Panup, W. Ratiphaphongthon, W. Wangkeeree, A Novel Twin Support Vector Machine with Generalized Pinball Loss Function for Pattern Classification, Symmetry 14(2)(2022) 289.

[20] O. L. Mangasarian, Nonlinear Programming; Society for Industrial and Applied Mathematics, (1994).
[21] R. Khemchandani, Jayadeva, S. Chandra, Optimal kernel selection in twin support vector machines, Optimization Letters 3(1)(2009) 77–88.
[22] D. Dua, C. Graff, UCI machine learning repository, University of California, School ofInformation and Computer Science: Irvine, CA, USA, 2019. Available from: http://archive.ics.uci.edu/ml
[23] V. Garc, J.S. Sanche, R.A. Mollineda, On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, Knowledge Based Systems 25(1)(2012) 13–21.
[24] C. W. Hsu, C. C. Chang, C. J. Lina, A Practical Guide to Support Vector Classification, Nat. Taiwan Univ., Taipei, Taiwa, (2003) 1–12.
[25] S. Garcia, A. Fernandez, J. Luengo, F. Herrera, Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining experimental analysis of power, Information Sciences 180(10)(2010) 2044–2064.
[26] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, Proc. of the 10th IEEE Int. Conf. Comput. Vis. (ICCV) 1(2005) 1395–1402.