Research Reports on Computer Science
https://ojs.wiserpub.com/index.php/RRCS/

UNIVERSAL WISER PUBLISHER

Research Article

Efficient Learning of Pinball TWSVM using Privileged Information and its Applications

Reshma Rastogi (nee. Khemchandani)[1*], Aman Pal[1], Suresh Chandra[2]

[1]Department of Computer Science, South Asian University, New Delhi, India
[2]Department of Mathematics, Indian Institute of Technology, New Delhi, India
E-mail: reshma.khemchandani@sau.ac.in

Received: 10 February 2022; **Revised:** 2 May 2022; **Accepted:** 18 May 2022

Abstract: Expert knowledge plays a vital role in any learning framework. However, in the field of machine learning, an expert's knowledge is rarely utilised. Furthermore, most machine learning methods (support vector machine, SVM-based) use a hinge loss function that is sensitive to noise. Thus, in order to benefit from expert knowledge while reducing noise sensitivity, we propose in this paper a fast and novel Twin Support Vector Machine classifier based on privileged information with a pinball loss function, dubbed Pin-TWSVMPI, where expert knowledge is in the form of privileged information. The proposed Pin-TWSVMPI incorporates privileged information into two nonparallel decision hyperplanes by employing a correction function. Furthermore, we employ the Sequential Minimal Optimization (SMO) technique to get the classifier in order to make computations more efficient and faster, and we have demonstrated its applicability for pedestrian detection and handwritten digit recognition. Furthermore, for UCI datasets, we first construct a process that retrieves privileged information from the dataset's features, which is then used by Pin-TWSVMPI, resulting in improved classification accuracy with a reduced computing time.

Keywords: Twin Support Vector Machine, privileged information, pinball loss function, pedestrian detection, handwritten digit recognition, Sequential Minimal Optimization

1. Introduction

Classification has been extensively worked upon in machine learning. Due to the limitation of training data, learning is still a demanding work [1]. Therefore, training data when combined with an expert's opinion has the power to tremendously improve the performance of the algorithm. In the classical supervised learning paradigm, by using training data the learner aims to obtain a decision function having minimum generalization error on the unfamiliar test data. Practically expert's opinion is minimally used. In literature, human learning is performed by using expert's opinion in the form of descriptions, remarks, comparisons, etc. In [2], Vapnik et al. termed this additional prior information as the privileged information, which is available only at the training stage but not available while testing. Vapnik et al. [2] introduced Learning Using Privileged Information (LUPI) also known as Support Vector Machine with Privilege Information (SVMPI), which can boost the convergence rate of learning especially when the learning problem itself is NP(nondeterministic polynomial time)-hard.

In recent times, Support Vector Machine (SVM) has become a significant algorithm for classification problem [3-5]. Motivated by the statistical learning theory and the maximum margin principle, SVM [6, 7] tries to find an optimal

DOI: https://doi.org/10.37256/rrcs.1120221325

separating hyperplane by solving a quadratic programming problem (QPP). However, solving QPP is expensive, especially for large-scale datasets. To improve the performance of SVM, in [8], Jayadeva et al. proposed twin SVM (TWSVM) for binary classification in the spirit of the Generalized Eigenvalues Proximal Support Vector Machine (GEPSVM) [9]. TWSVM aims to identify two nonparallel hyperplanes such that each hyperplane is proximal to its own class and away from the samples of other class and it's obtained by solving a pair of smaller-sized QPPs. The training speed of the TWSVM is faster than the standard SVM because it solves two smaller sized QPPs rather than a single larger-sized QPP. Further Least Square-TWSVM [10] and Wavelet Kernel TWSVM [11] improve the training speed of the TWSVM classifier.

Recently, Huang et al. [12], proposed SVM with the pinball loss function (Pin-SVM) where hinge loss function is replaced with pinball loss function, which brought noise insensitivity to the classifier [12]. Huang et al. [12] have also shown that for noise-corrupted data, Pin-SVM is theoretically and empirically better than SVM. Recently, in [13], Mehrkanoon et al. have presented non-parallel support vector classifiers with different loss functions. They have also shown the superiority of pinball loss function over hinge loss function in TWSVM.

In recent times, machine-human interaction has gained prominence and is widely being researched and developed. The role of machine learning has become indispensable while designing tools to detect and track people [14]. In recent years, pedestrian detection is become a popular topic of research due to its several applications [15]. Practical applications [16] in pedestrian detection include robotics, surveillance, content-based indexing and automatic driver-assistance systems in vehicles. In this paper, we perform the experiment on pedestrian detection to demonstrate the effectiveness of privileged information with the proposed formulation. It is essential to note that in computer vision research problems, time plays a crucial role as it usually involves nearly real-time interaction with the environment which leads to a large number of samples. Thus, our proposed formulation shows a significant deduction concerning training time while maintaining the overall performance of the method.

In the past few years, extracting privileged/expert information from the benchmark datasets has become a challenging task [17, 18]. Recently, in [19], Qi et al. have proposed a fast version of TWSVM based privileged information termed as FTWSVMPI where the authors present the privileged information as the solution of two additional QPP model along with two more QPP of TWSVM. This leads to a major disadvantage of Qi et al. [19] model as it lengthens the training time of the classifier. To overcome this drawback of Qi et al. model, in our previous work [20], we have introduced a novel method to extract privileged information from the datasets themselves which avoids solving two additional QPPs. Further, Qi et al. [19] and our previous model [20] have used hinge loss function which leads to noise sensitivity. Therefore, in this paper, we propose a novel twin support vector machine with the pinball loss function based on privileged information, which brings back the noise sensitivity to the classifier. The highlights of this work are the following:

- We introduce TWSVM with pinball loss function termed as Pin-TWSVM with privileged information by replacing the error variable with privileged information variable in the optimization problem of Pin-TWSVM [13]. The privileged information variable is obtained by using correction space which defines correcting function [2].
- Firstly, in order to take care of the structural risk component, a regularization term in the objective function is introduced.
- The dual problems of Twin Support Vector Machine with Pinball Loss based on Privilege Information (Pin-TWSVMPI) have a similar formulation to that of standard SVM, and hence they are solved efficiently by using Sequential Minimal Optimization (SMO) [21] which makes computation faster as compared to FTWSVMPI.
- We also utilize a novel method proposed by Aman and Reshma [20], which extracts privileged information from the features of datasets where expert's knowledge is not available, e.g., in UCI datasets.
- Experimental results on multiple benchmark UCI datasets [22] specify that the introduced formulation has better classification accuracy in comparison to the formulations discussed in this paper and with considerably decreased computational time for linear and nonlinear formulations.
- To establish the efficacy of the proposed formulation with privileged information, we have shown its application for pedestrian detection over INRIA dataset [23] and handwritten digit recognition over MNIST dataset.

The rest of the paper is organized as follows. Section 2 discusses the Pin-TWSVM. The Pin-TWSVMPI is proposed in Section 3, which includes both the linear and nonlinear cases. Section 4 discusses the privileged information used for both the applications such as pedestrian detection and handwritten digit recognition. Experiments on UCI benchmark, INRIA and MNIST datasets are conducted to verify the effectiveness of proposed Pin-TWSVMPI in Section 5. Section 5.1 discusses the method used to extract privileged information from the dataset. Section 6 provides the concluding remarks.

2. Related works

We consider a binary classification problem with a training set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$ and privileged information $T^* = \{(x_1^*, y_1), (x_2^*, y_2), \ldots, (x_l^*, y_l)\}$, where $X_l = \{x_i : i = 1, 2, ..., l\}$ and $X_l^* = \{x_i : i = 1, 2, ..., l\}$ are the l data points in d and d^* dimensions with corresponding class labels $Y_l = \{y_i \in [1, -1] : i = 1, 2, ..., l\}$ respectively.

Data points belonging to class +1 and −1 are represented by matrices A and B each with number of patterns m_1 and m_2, respectively. Therefore, the size of matrices A and B are $(m_1 \times d)$ and $(m_2 \times d)$, respectively. Here, d is the dimension of the feature space. Let $A_i(i = 1, 2, ..., m_1)$ and $B_i(i = 1, 2, ..., m_2)$ are the row vectors in d-dimensional real space $\mathbb{R}^d$ that represents feature vector of class +1 and class −1 data samples respectively.

Privilege information belonging to class +1 and −1 are represented by matrices A^* and B^* each with number of patterns m_1^* and m_2^*, respectively. Therefore, the size of matrices A^* and B^* are $(m_1^* \times d^*)$ and $(m_2^* \times d^*)$, respectively. Here, d^* is the dimension of privileged information feature space. Let $A_i^*(i = 1, 2, ..., m_1^*)$ and $B_i^*(i = 1, 2, ..., m_2^*)$ are the row vectors in d^*-dimensional correcting space $\mathbb{R}^{d^*}$ that represents privileged information of class +1 and class −1 respectively.

The privileged information paradigm can be described as follows: given a set of triplets (training data)

$$(x_1, x_1^*, y_l), \ldots, (x_l, x_l^*, y_l)\}, x_i \in X_l, x_1^* \in X_l^*, y_i \in \{-1, 1\}$$

where X^* is the correcting space which is obtained by expert knowledge. The classical paradigm of supervised machine learning is explained as follows: given a set of pairs

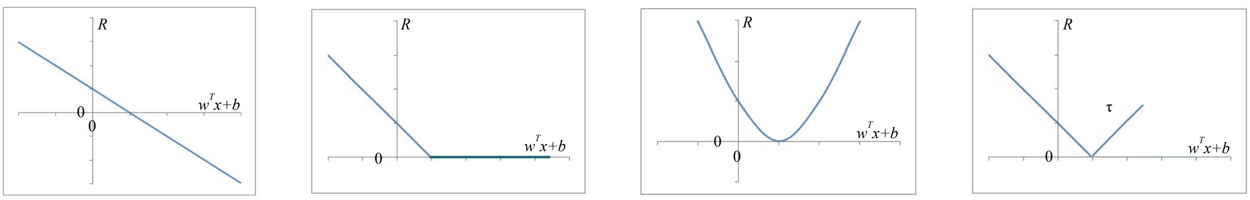
$$(x_1, y_1), \ldots, (x_l, y_l)\}, x_i \in X, y_i \in \{-1, 1\}.$$

2.1 *Pin-TWSVM*

In [12], Huang et al. proposed a SVM classifier with pinball loss function, named as Pin-SVM. With the help of pinball loss function, the authors use the quantile distance to measure the margin and propose the corresponding classifier to maximize the quantile distance. The pinball loss function is given as follows:

$$L_\tau(x, y, f(x)) = \begin{cases} 1 - y f(x) & : \text{ if } 1 - y f(x) \geq 0, \\ -\tau(1 - y f(x)) & : \text{ if } 1 - y f(x) < 0, \end{cases} \tag{1}$$

which can be regarded as generalized L_1-loss function. Figure 1 shows the geometric interpretation of different types of loss functions. Figure 1(d) represents pinball loss function.



(a) Linear loss (b) Hinge loss (c) Quadratic loss (d) Pinball loss

Figure 1. Geometrical interpretation of different loss functions

Taking motivation from [8] and [12], in [13], Mehrkanoon et al. have proposed TWSVM with pinball loss function termed as Pin-TWSVM. The corresponding model is formulated as the following pair of QPPs:

$$\begin{aligned} &\underset{w_1, b_1, \xi_2}{\text{Min}} \ \frac{1}{2}\left\|Aw_1 + e_1 b_1\right\|_2^2 + c_1 e_2^T \xi_2 \\ &\text{s.t.} \ -(Bw_1 + e_2 b_1) \geq e_2 - \xi_2, \\ &\quad -(Bw_1 + e_2 b_1) \leq e_2 + \frac{1}{\tau}\xi_2, \end{aligned} \tag{2}$$

and

$$\begin{aligned} &\underset{w_2, b_2, \xi_1}{\text{Min}} \ \frac{1}{2}\left\|Bw_2 + e_2 b_2\right\|_2^2 + c_2 e_1^T \xi_1 \\ &\text{s.t.} \ -(Aw_2 + e_1 b_2) \geq e_1 - \xi_1, \\ &\quad -(Aw_1 + e_1 b_2) \leq e_1 + \frac{1}{\tau}\xi_1, \end{aligned} \tag{3}$$

where c_1, $c_2 > 0$ are trade-off parameters and e_1 and e_2 are vectors of ones of appropriate dimensions. Notice that, similar to Pin-SVM, when $\tau = 0$, the second constraint of both the problems become $\xi_1 \geq 0$, $\xi_2 \geq 0$ and equation (2) and equation (3) respectively reduce to the traditional TWSVM.

Using Wolfe's dual method [24] and Karush-Kuhn-Tucker (KKT) conditions [24], we obtain the dual of equation (2) as follows:

$$\begin{aligned} &\underset{(\alpha-\beta)}{\text{Max}} \ e_2^T(\alpha-\beta) - \frac{1}{2}(\alpha-\beta)^T G^T (H^T H)^{-1} G(\alpha-\beta) \\ &\text{s.t.} \qquad c_1 - \alpha - \frac{\beta}{\tau} = 0, \\ &\qquad\qquad \alpha \geq 0, \beta \geq 0, \end{aligned} \tag{4}$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{l_1})^T$ and $\beta = (\beta_1, \beta_2, ..., \beta_{l_2})^T$ are the vectors of Lagrange multipliers. Here, $H = [A \ e_1]$ and $G = [B \ e_2]$. Similarly, the dual problem of equation (3) can be derived as follows:

$$\begin{aligned} &\underset{(\gamma-\rho)}{\text{Max}} \ e_1^T(\gamma-\rho) - \frac{1}{2}(\gamma-\rho)^T H^T (G^T G)^{-1} H(\gamma-\rho) \\ &\text{s.t.} \qquad c_2 - \gamma - \frac{\rho}{\tau} = 0, \\ &\qquad\qquad \gamma \geq 0, \rho \geq 0, \end{aligned} \tag{5}$$

Here, the augmented vectors $u = [w_1, b_1]$ and $v = [w_2, b_2]$ are given by:

$$u = (H^T H)^{-1} G^T (\alpha - \beta), \tag{6}$$

$$v = (G^T G)^{-1} H^T (\gamma - \rho). \tag{7}$$

Once vectors u and v are known from equations (7) and (8), the separating planes

$$x^T w_1 + b_1 = 0, \qquad\qquad x^T w_2 + b_2 = 0, \tag{8}$$

are obtained. A new data sample $x_{\text{new}} \in \mathbb{R}^n$ is assigned to the class 1 or class −1, based on which of the two hyperplanes it lies closest to, i.e.

$$f(x_{\text{new}}) = \text{argmin}\{d_{\pm}(x_{\text{new}})\}, \tag{9}$$

where

$$d_{\pm} = \left| \frac{x_{\text{new}}^{T} w_{\pm} + b_{\pm}}{\|w_{\pm}\|_2^2} \right| \tag{10}$$

where $|\cdot|$ is the absolute value of the distance of point x_{new} from the plane.

3. Pin-TWSVMPI

In this section, we introduce a novel approach of classification termed as Pin-TWSVMPI. In [19], Qi et al. have solved two additional QPPs model to incorporate privileged information in TWSVM which further leads to the higher computational cost of the model. In order to avoid, solving two additional QPPs model, in this paper, we utilize the novel method proposed in which extract privileged information from the dataset and then incorporate the same to the proposed Pin-TWSVMPI model. This leads to the lesser computational cost of the proposed Pin-TWSVMPI as compared to FTWSVMPI.

The process for generating privileged information from the different experts is shown in Figure 2 which is further used in training the proposed Pin-TWSVMPI classifier. This figure also shows that an expert always supplies privileged information x^* in the correcting space of X^* with the admissible set of the correcting functions [2].

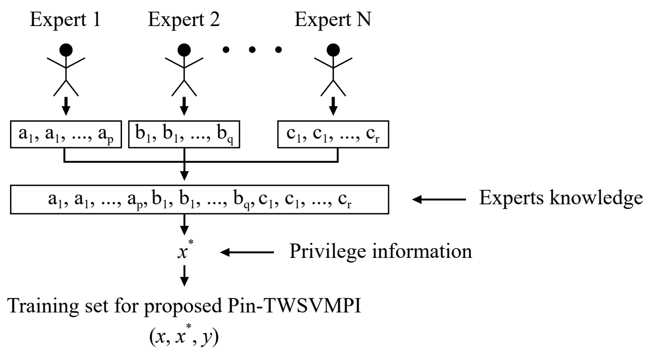


Figure 2. Process to generate privileged information from different experts knowledge

3.1 *Linear Pin-TWSVMPI*

Once we have generated privileged information from different experts, we map input feature vector x of our training triplets (x, x^*, y) into space X and privileged information vector x^* into X^* where we define our decision rule as $(w^T X + b)$ and correcting function as $\xi = (w^{*T} x^* + b^*)$ in proposed Pin-TWSVMPI model. To calculate the functions, we minimize the given QPPs:

$$\begin{aligned} \min_{w_1, b_1, w_1^*, b_1^*} \quad & \frac{1}{2}\|w_1\|_2^2 + \frac{\gamma}{2}\|w_1^*\|_2^2 + \frac{1}{2}\|\eta_1\|_2^2 + \frac{\gamma}{2}\|\eta_1^*\|_2^2 + c_1 e_2^{*T}(B^* w_1^* + e_2^* b_1^*) \\ \text{s.t.} \quad & A w_1 + e_1 b_1 = \eta_1, \\ & A^* w_1^* + e_1^* b_1^* = \eta_1^*, \\ & -(B w_1 + e_2 b_1) \geq e_2 - (B^* w_1^* + e_2^* b_1^*), \\ & -(B w_1 + e_2 b_1) \leq e_2 + \frac{1}{\tau}(B^* w_1^* + e_2^* b_1^*), \end{aligned} \tag{11}$$

and

$$\begin{aligned}
\underset{w_2, b_2, w_2^*, b_2^*}{\text{Min}} \quad & \frac{1}{2}\|w_2\|_2^2 + \frac{\gamma}{2}\|w_2^*\|_2^2 + \frac{1}{2}\|\eta_2\|_2^2 + \frac{\gamma}{2}\|\eta_2^*\|_2^2 + c_2 e_1^{*T}(A^* w_2^* + e_1^* b_2^*) \\
\text{s.t.} \quad & Bw_2 + e_2 b_2 = \eta_2, \\
& B^* w_2^* + e_2^* b_2^* = \eta_2^*, \\
& -(Aw_2 + e_1 b_2) \geq e_1 - (A^* w_2^* + e_1^* b_2^*), \\
& -(Aw_2 + e_1 b_2) \leq e_1 + \frac{1}{\tau}(A^* w_2^* + e_1^* b_2^*),
\end{aligned} \tag{12}$$

where $\gamma > 0$ is a parameter.

The proposed Pin-TWSVMPI implements the Structured Risk Minimization (SRM) principle [25] represented by the first two terms of the objective function (11) in feature space and correcting space respectively, which ensure good generalization ability. The third and fourth terms of the objective function (11) minimize the projection of the class A samples in feature space and correcting space respectively, from the desired hyperplane, i.e., it tries to drive the samples of class A close to the desired hyperplane which is a representation of class A. Since the proposed Pin-TWSVMPI is formulated as a soft-margin classifier, it permits violations of constraints represented by the fifth term of the objective function of (11). The proposed Pin-TWSVMPI takes into consideration the principle of empirical risk minimization (ERM). The last constraint of (11) implements the pinball loss function. A similar illustration can be drawn for (12).

The dual of the explained primal problem can be calculated by using Wolfe's dual method. The Lagrangian function corresponding to the problem equation (11) is given by

$$\begin{aligned}
& L(w_1, b_1, w_1^*, b_1^*, \eta_1, \eta_1^*, \alpha, \beta) \\
&= \frac{1}{2}\|w_1\|_2^2 + \frac{\gamma}{2}\|w_1^*\|_2^2 + \frac{1}{2}\|\eta_1\|_2^2 + \frac{\gamma}{2}\|\eta_1^*\|_2^2 + c_1 e_2^{*T}(B^* w_1^* + e_2^* b_1^*) + \alpha_1^T[Aw_1 + e_1 b_1 - \eta_1] + \alpha_2^T[Aw_1^* + e_1^* b_1^* - \eta_1^*] \\
&\quad + \alpha_3^T[-(Bw_1 + e_2 b_1) - e_2 + (B^* w_1^* + e_2^* b_1^*)] + \alpha_4^T[-(Bw_1 + e_2 b_1) - e_2 + \frac{1}{\tau}(B^* w_1^* + e_2^* b_1^*)],
\end{aligned} \tag{13}$$

where α_1, α_2, α_3, α_4 are vectors of Lagrange multipliers with length l. The KKT conditions [24] for equation (13) are described as:

$$w_1 + A^T\alpha_1 + B^T\alpha_3 - B^T\alpha_4 = 0, \tag{14}$$

$$e_1^T\alpha_1 + e_2^T\alpha_3 - e_2\alpha_4 = 0, \tag{15}$$

$$\gamma w_1^* + c_1 B^{*T} e_2^* + A^{*T}\alpha_2 + B^{*T}\alpha_3 - \frac{1}{\tau}B^{*T}\alpha_4 = 0, \tag{16}$$

$$c_1 e_2^{*T} e_2^* + e_1^{*T}\alpha_2 - e_2^{*T}\alpha_3 - \frac{1}{\tau}e_2^{*T}\alpha_4 = 0, \tag{17}$$

$$\eta_1 - \alpha_1 = 0, \tag{18}$$

$$\eta_1^* - \alpha_2 = 0, \tag{19}$$

$$Aw_1 + e_1 b_1 = \eta_1, \tag{20}$$

$$A^* w_1^* + e_1^* b_1^* = \eta_1^*, \tag{21}$$

$$-(Bw_1 + e_2 b_1) + (B^* w_1^* + e_2^* b_1^*) \geq e_2, \tag{22}$$

$$e_2 + (Bw_1 + e_2 b_1) - \frac{1}{\tau}(B^* w_1^* + e_2^* b_1^*) \geq 0, \tag{23}$$

$$\alpha_1[Aw_1 + e_1 b_1 - \eta_1] = 0, \tag{24}$$

$$\alpha_2[A^* w_1^* + e_1^* b_1^* - \eta_1^*] = 0, \tag{25}$$

$$\alpha_3[-(Bw_1 + e_2 b_1) - e_2 + (B^* w_1^* + e_2^* b_1^*)] = 0, \tag{26}$$

$$-\alpha_4[-(Bw_1 + e_2 b_1) - e_2 + \frac{1}{\tau}(B^* w_1^* + e_2^* b_1^*)] = 0, \tag{27}$$

$$\alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_3 \geq 0, \alpha_4 \geq 0. \tag{28}$$

Now, from (14) and (16), we get

$$w_1 = B^T(\alpha_4 - \alpha_3) - A^T\alpha_1, \tag{29}$$

and

$$w_1^* = \frac{1}{\gamma}\left[B^{*T}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*) - A^{*T}\alpha_2 \right], \tag{30}$$

respectively. Using equation (13) and the above KKT conditions, we calculate the Wolfe dual of equation (11) as given:

$$\begin{aligned} \underset{\alpha_1,\alpha_2,\alpha_3,\alpha_4}{\text{Min}} \; & \frac{1}{2}(\alpha_4-\alpha_3)^T BB^T(\alpha_4-\alpha_3) + \frac{1}{2}\alpha_1 AA^T\alpha_1 + (\alpha_4-\alpha_3)BA^T\alpha_1 + \frac{1}{2\gamma}(\alpha_3+\frac{1}{\tau}\alpha_4 - c_1e_2^*)^T B^* B^{*T}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*) \\ & + \frac{1}{2\gamma}\alpha_2^T A^* A^{*T}\alpha_2 - \frac{1}{\gamma}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*)^T B^* A^{*T}\alpha_2 + e_2^T(\alpha_3 - \alpha_2) \\ \text{s.t.} \; & e_1^T\alpha_1 - e_2^T(\alpha_4 - \alpha_3) = 0, \\ & e_1^{*T}\alpha_2 - (\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*)^T e_2^* = 0, \\ & \alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0. \end{aligned} \tag{31}$$

Next, from (18) and (20), we get

$$b_1 = \frac{1}{l}(\alpha_1 - Aw_1). \tag{32}$$

Similarly, from (19) and (21), we get

$$b_1^* = \frac{1}{l}(\alpha_2 - A^* w_1^*). \tag{33}$$

Furthermore, let

$$Q = \begin{bmatrix} AA^T & 0 & -AB^T & AB^T \\ 0 & \frac{1}{\gamma}A^*A^{*T} & -\frac{1}{\gamma}A^*B^{*T} & -\frac{1}{\gamma\tau}A^*B^{*T} \\ -BA^T & -\frac{1}{\gamma}B^*A^{*T} & BB^T + \frac{1}{\gamma}B^*B^{*T} & -BB^T + \frac{1}{\gamma\tau}B^*B^{*T} \\ BA^T & -\frac{1}{\gamma\tau}B^*A^{*T} & -BB^T + \frac{1}{\gamma\tau}B^*B^{*T} & BB^T + \frac{1}{\gamma\tau}B^*B^{*T} \end{bmatrix},$$

$$f = \left[0 \quad (c_1 e_2^{*T} B^* A^{*T} - e_2) \quad (e_2 - c_1\frac{2}{\gamma}B^*B^{*T}) \quad c_1\frac{2}{\gamma\tau}e_2^{*T}B^*B^{*T} \right],$$

$$C=\begin{bmatrix} e_1 & 0 & e_2 & -e_2 \\ 0 & e_1^* & -e_2^* & -\frac{1}{\tau}e_2^* \end{bmatrix}, \quad D=\begin{bmatrix} 0 & c_1 \end{bmatrix}^T,$$

and

$$X=[\alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4]^T.$$

Thus, the problem (31) can be reduced to

$$\begin{aligned} &\min_X \frac{1}{2} X^T Q X + f^T X \\ &CX = D, \\ &X \geq 0. \end{aligned} \tag{34}$$

The problem (34) is a convex QPP and has a formulation similar to the dual of SVM [7]. Therefore, the well known SMO [21] could be used to solve the aforementioned problem.

On similar lines, the dual of (12) can also be calculated as discussed below:

$$\begin{aligned} \underset{\beta_1,\beta_2,\beta_3,\beta_4}{\text{Min}} \quad & \frac{1}{2}(\beta_4-\beta_3)^T AA^T(\beta_4-\beta_3)+\frac{1}{2}\beta_1 BB^T\beta_1+(\beta_4-\beta_3)AB^T\beta_1+\frac{1}{2\gamma}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T A^*A^{*T}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*) \\ & +\frac{1}{2\gamma}\beta_2^T B^*B^{*T}\beta_2-\frac{1}{\gamma}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T A^*B^{*T}\beta_2+e_1^T(\beta_3-\beta_2) \\ \text{s.t.} \quad & e_2^T\beta_1-e_1^T(\beta_4-\beta_3)=0, \\ & e_2^{*T}\beta_2-(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T e_1^*=0, \\ & \beta_1,\beta_2,\beta_3,\beta_4\geq 0. \end{aligned} \tag{35}$$

Here, (w_2, b_2) and (w_2^*, b_2^*) can be calculated as

$$\begin{aligned} w_2 &= A^T(\beta_4-\beta_3)-B^T B_1, \\ b_2 &= \frac{1}{l}(\beta_1-Bw_2). \end{aligned}$$

and

$$\begin{aligned} w_2^* &= \frac{1}{\gamma}\left[A^{*T}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)-B^{*T}\beta_2\right], \\ b_2^* &= \frac{1}{l}(\beta_2-B^*w_2^*). \end{aligned} \tag{36}$$

Similar to (31), SMO can also be used to solve (35).

Once vectors w_1, w_2, w_1^*, w_2^* and scalars b_1, b_2, b_1^*, b_2^* are obtained from above, the separating hyperplanes

$$x^T w_1 + b_1 = 0 \quad \text{and} \quad x^T w_2 + b_2 = 0, \tag{37}$$

and subsequent correcting functions

$$x^{*T} w_1^* + b_1^* = 0 \quad \text{and} \quad x^{*T} w_2^* + b_2^* = 0, \tag{38}$$

are obtained. Due to the availability of privilege information at the time of training only, so a new data sample $x_{\text{new}} \in \mathbb{R}^n$ is assigned to the class 1 or to class −1, depending on its proximity to the two hyperplanes, i.e.

$$f(x_{\text{new}}) = \text{argmin}\{d_{\pm}(x_{\text{new}})\}, \tag{39}$$

where

$$d_{\pm}(x_{\text{new}}) = \left| \frac{x_{\text{new}}^T w_{\pm} + b_{\pm}}{\|w_{\pm}\|_2^2} \right|, \tag{40}$$

where $|\cdot|$ is the absolute value of the distance of point x_{new} from the plane.

3.2 *Nonlinear Pin-TWSVMPI*

The nonlinear Pin-TWSVMPI can be demonstrated by incorporating kernel generated hyperplanes by

$$K(A, X^T)\mu_1 + e_1 v_1 = 0 \quad \text{and} \quad K(B, X^T)\mu_2 + e_2 v_2 = 0, \tag{41}$$

where $X = A \cup B$ and $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j))$ is an appropriate chosen kernel. Now, the kernel correcting function is defined by $\xi_1 = (K(A^*, X^*)^* \mu_1^* + v_1^*)$ and $\xi_2 = (K(B^*, X^*)^* \mu_2^* + v_2^*)$. Next, the optimization problems for nonlinear Pin-TWSVMPI are given as:

$$\begin{aligned}
\underset{\mu_1, v_1, \mu_1^*, v_1^*}{\text{Min}} \quad & \frac{1}{2}\|\mu_1\|_2^2 + \frac{\gamma}{2}\|\mu_1^*\|_2^2 + \frac{1}{2}\|\lambda_1\|_2^2 + \frac{\gamma}{2}\|\lambda_1^*\|_2^2 + c_1 e_2^{*T}(K(B^*, X^{*T})\mu_1^* + e_2^* v_1^*) \\
\text{s.t.} \quad & K(A, X^T)\mu_1 + e_1 v_1 = \lambda_1, \\
& K(A^*, X^{*T})\mu_1^* + e_1^* v_1^* = \lambda_1^*, \\
& -(K(B, X^T)\mu_1 + e_2 v_1) \ge e_2 - (K(B^*, X^{*T})\mu_1^* + e_2^* v_1^*), \\
& -(K(B, X^T)\mu_1 + e_2 v_1) \le e_2 + \frac{1}{\tau}(K(B^*, X^{*T})\mu_1^* + e_2^* v_1^*),
\end{aligned} \tag{42}$$

and

$$\begin{aligned}
\underset{\mu_2, v_2, \mu_2^*, v_2^*}{\text{Min}} \quad & \frac{1}{2}\|\mu_2\|_2^2 + \frac{\gamma}{2}\|\mu_2^*\|_2^2 + \frac{1}{2}\|\lambda_2\|_2^2 + \frac{\gamma}{2}\|\lambda_2^*\|_2^2 + c_2 e_1^{*T}(K(A^*, X^{*T})\mu_2^* + e_1^* v_2^*) \\
\text{s.t.} \quad & K(B, X^T)\mu_2 + e_2 v_2 = \lambda_2, \\
& K(B^*, X^{*T})\mu_2^* + e_2^* v_2^* = \lambda_2^*, \\
& (K(A, X^T)\mu_2 + e_1 v_2) \ge e_1 - (K(A^*, X^{*T})\mu_2^* + e_1^* v_2^*), \\
& (K(A, X^T)\mu_2 + e_1 v_2) \le e_1 + \frac{1}{\tau}(K(A^*, X^{*T})\mu_2^* + e_1^* v_2^*).
\end{aligned} \tag{43}$$

Similar to the linear case, the solution of equation (42) and equation (43) can be formulated from Wolfe's dual method as given:

$$\begin{aligned}
\underset{\alpha_1, \alpha_2, \alpha_3, \alpha_4}{\text{Min}} \quad & \frac{1}{2}(\alpha_4 - \alpha_3)^T NN^T(\alpha_4 - \alpha_3) + \frac{1}{2}\alpha_1 MM^T \alpha_1 + (\alpha_4 - \alpha_3)NM^T\alpha_1 + \frac{1}{2\gamma}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*)^T N^* N^{*T}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*) \\
& + \frac{1}{2\gamma}\alpha_2^T M^* M^{*T}\alpha_2 - \frac{1}{\gamma}(\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*)^T N^* M^{*T}\alpha_2 + e_2^T(\alpha_3 - \alpha_2) \\
\text{s.t.} \quad & e_1^T\alpha_1 - e_2^T(\alpha_4 - \alpha_3) = 0, \\
& e_1^{*T}\alpha_2 - (\alpha_3 + \frac{1}{\tau}\alpha_4 - c_1 e_2^*)^T e_2^* = 0, \\
& \alpha_1, \alpha_2, \alpha_3, \alpha_4 \ge 0.
\end{aligned} \tag{44}$$

and

$$\underset{\beta_1,\beta_2,\beta_3,\beta_4}{\text{Min}}\ \frac{1}{2}(\beta_4-\beta_3)^T MM^T(\beta_4-\beta_3)+\frac{1}{2}\beta_1 NN^T\beta_1+(\beta_4-\beta_3)MN^T\beta_1+\frac{1}{2\gamma}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T M^*M^{*T}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)$$

$$+\frac{1}{2\gamma}\beta_2^T N^*N^{*T}\beta_2-\frac{1}{\gamma}(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T M^*N^{*T}\beta_2+e_1^T(\beta_3-\beta_2)$$

$$\text{s.t.}\quad e_2^T\beta_1-e_1^T(\beta_4-\beta_3)=0,$$

$$e_2^{*T}\beta_2-(\beta_3+\frac{1}{\tau}\beta_4-c_2e_1^*)^T e_1^*=0,$$

$$\beta_1,\beta_2,\beta_3,\beta_4\geq 0. \tag{45}$$

where $M=[K(A,X^T)\ \ e_1]$, $N=K(B,X^T)\ \ e_2]$, $M^*=[K(A^*,X^{*T})\ \ e_1^*]$, and $N^*=[K(B^*,X^{*T})\ \ e_2^*]$.

Similar to the linear case, SMO can be used to solve equation (44) and equation (45) respectively. Once the vector of dual variables α_1, α_2, α_3, α_4 and β_1, β_2, β_3, β_4 are determined, the vectors of primal variables are calculated as below:

$$\mu_1=N^T(\alpha_4-\alpha_3)-M^T\alpha_1,$$
$$v_1=\frac{1}{l}(\alpha_1-M\mu_1),$$
$$\mu_1^*=\frac{1}{\gamma}\left[\alpha_3+\frac{1}{\tau}\alpha_4-c_1N^{*T}1e_2^*-M^{*T}\alpha_2\right],$$
$$v_1^*=\frac{1}{l}(\alpha_2-A^*w_1^*),$$

and

$$\mu_2=M^T(\beta_4-\beta_3)-N^T\beta_1,$$
$$v_2=\frac{1}{l}(\beta_1-N\mu_2),$$
$$\mu_2^*=\frac{1}{\gamma}\left[\beta_3+\frac{1}{\tau}\beta_4-c_2M^{*T}e_1^*-N^{*T}\beta_2\right],$$
$$v_2^*=\frac{1}{l}(\beta_2-N^*\mu_1^*).$$

A new data sample $x_{\text{new}}\in\mathbb{R}^n$ is assigned to the class 1 or to class −1, depending on its proximity to the two hyperplanes, i.e.

$$f(x_{\text{new}})=\text{argmin}\{d_{1,2}(x_{\text{new}})\}, \tag{46}$$

where

$$d_{1,2}(x_{\text{new}})=\left|\frac{K(x_{\text{new}},X^T)^T\mu_{1,2}+v_{1,2}}{\|\mu_{1,2}\|_2^2}\right| \tag{47}$$

where $|\cdot|$ is the absolute value of distance of point x_{new} from the plane.

4. Two applications of privileged information

In this section, we discuss privileged information for pedestrian detection and handwritten digit recognition.

4.1 *Pedestrian detection*

Pedestrian detection is a key problem in the computer vision field [16]. In the last few years, diverse efforts have been made to improve the performance of pedestrian detection in [26-29]. In this paper, we also propose a fast and novel framework for pedestrian detection problem and explain each step involved therein. We use Pin-TWSVMPI which helps to enhance the performance of the proposed pedestrian detection framework. The training dataset in pedestrian detection includes positive and negative images. A positive image means at least one pedestrian is present in the image whereas a negative image means pedestrian is not present in the image. In this framework, we extract histogram of oriented gradients (HOG) [23], histogram of optical flow (HOF) [30] and gray level co-occurrence matrix [31] features from the images. Here, each feature vector gives a different independent view of the same data, e.g., HOG represents training data view, HOF and gray level co-occurrence represents privileged information view. From another perspective, we may also term this privileged information view as different 'expert's knowledge' which is extracted in the correcting space. Note that, expert's knowledge is only available at the time of training. Thus, for the testing image, we extract only HOG feature as the testing feature. The proposed framework would follow the following steps to detect pedestrians:

- **Training phase**: Input the training dataset (although this framework could be extended in general for any type of dataset), k-fold parameter, kernel parameter, τ, γ, c_1 and c_2 parameters of Pin-TWSVMPI.
- **Step 1**: Use Viola Jones method [26] to identify the detection window for every pedestrian present in the positive image.
- **Step 2**: Each detection window is expressed by overlapping blocks, and we extract HOG in feature space and HOF and gray level co-occurrence matrix features in correcting space for each block.
- **Step 3**: Train Pin-TWSVMPI with the combination of extracted features and privileged information via k-fold cross validation strategy and obtain the optimal classifier.
- **Testing phase:** Identify the pedestrian in a testing image i.e., either the pedestrian is presented or not in the given image with the help of trained Pin-TWSVMPI.

Figure 3 represents the graphical interpretation of the proposed framework to detect the pedestrian. In Figure 3, we have shown the result for both presence and absence of pedestrian/object under consideration. This figure also shows that if the testing image does not have any pedestrian present in the image, then the classifier will not identify the same. Note that, in our experiments, we have used only two experts knowledge, but this can be increased as per an expert's availability.

A major highlight of the above-discussed framework is the freedom to incorporate knowledge from different experts without increasing the training time of the proposed Pin-TWSVMPI classifier.

4.2 *Handwritten digit recognition*

Handwritten digit recognition is a current research area in Optical Character Recognition (OCR) applications and pattern classification [32]. The performance of OCR is largely based on the classification/learning scheme [33]. In recent years, SVM based classifiers have attracted the attention of researchers in the field of handwritten digit recognition [34, 35].

To show an application of the proposed Pin-TWSVMPI on handwritten digit recognition, we used MNIST dataset which has become a standard benchmark dataset for learning handwritten digit recognition [36]. In this paper, we classify the images of digits 5 and 8 in the MNIST database. Similar to [2], for every training image of MNIST dataset, we utilized its holistic (poetic) description which is further used as the privileged information for the proposed model. The poetic description of every digit of MNIST dataset is directly taken from [2]. These poetic descriptions for all training images were available in 21-dimensional feature vectors which will act as the privileged information for the proposed model. The MNIST data and its poetic descriptions with corresponding feature vectors can be downloaded from [37].

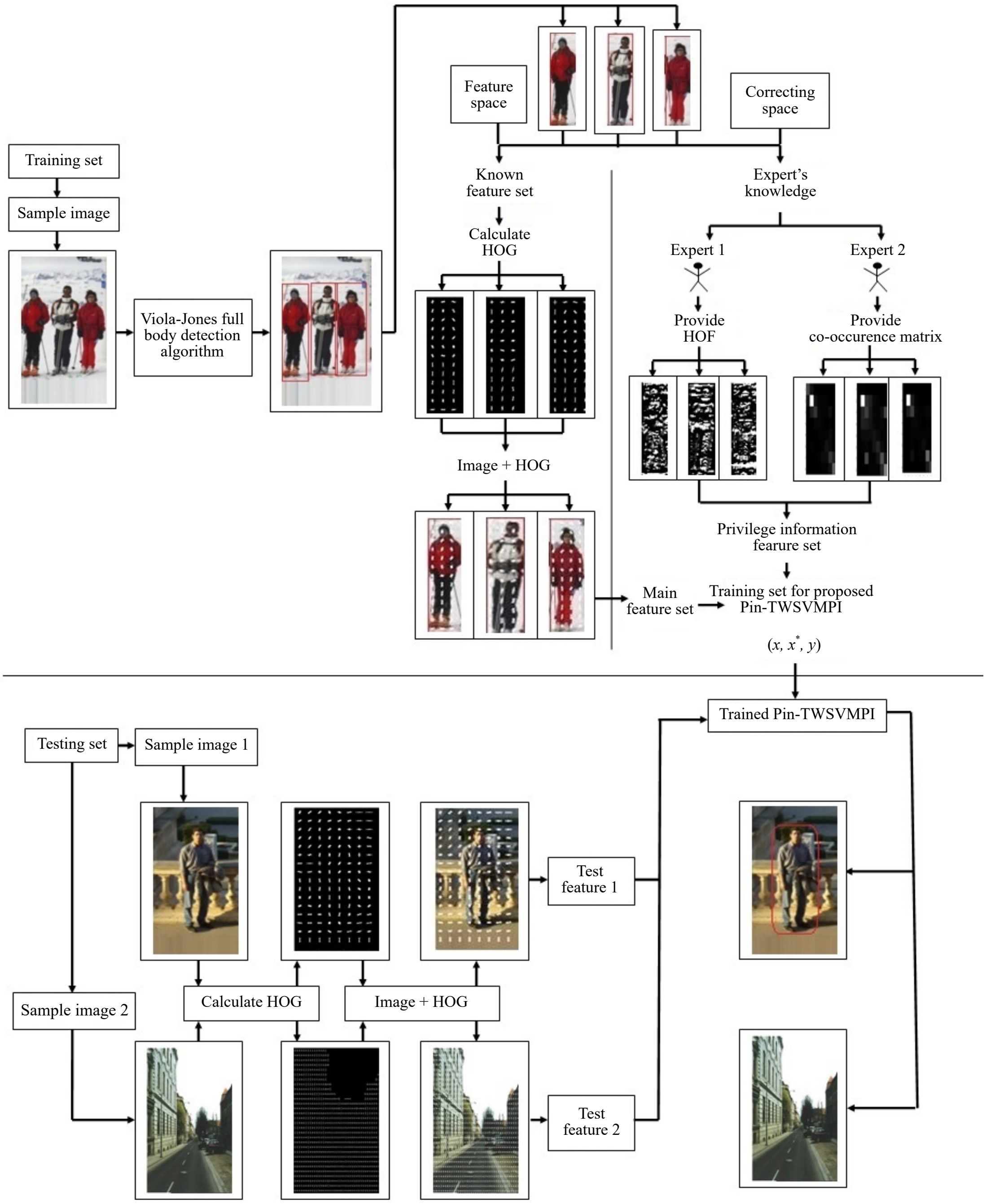


Figure 3. Complete procedure for pedestrian detection with proposed Pin-TWSVMPI

5. Experimental Results

Experiments were performed in MATLAB version 8.1 on a machine with 3.40 GHz CPU and 16 GB RAM under Microsoft Windows 64-bit operating system. To check the performance of proposed Pin-TWSVMPI, where privileged information is available i.e., pedestrian detection and handwritten digit recognition problems, and when privileged information is not available i.e., UCI machine learning repository [22] datasets. Note that the bold values in the experiment tables represent the best value along rows.

5.1 *Performance on UCI database*

In this section, we present the performance of the proposed formulation over UCI datasets (Table 1) where privileged information is not available to establish the classification accuracy of the proposed Pin-TWSVMPI. Privilege information is the additional information about the datasets which further help us to improve the performance of the proposed Pin-TWSVMPI. However, in a few problems, where an expert's knowledge is not available, the following procedure is used to extract privileged information.

Table 1. Summary of UCI datasets

Dataset	No. of instances	No. of features	No. of classes
Iris	150	4	3
Compound	399	2	6
Wine	178	13	3
Ecoli	327	7	5
Dermatology	358	34	6
Zoo	101	17	7
Haberman	303	3	2
Soybean	47	35	4
Page Block	5,473	10	5

5.1.1 *Procedure to extract privilege information*

To obtain privileged information, we have implemented the idea of Aman and Reshma [20] which retrieve notable characteristics from the feature set of data. To do the retrieving, we incorporate Principal Components Analysis (PCA) [38], a statistical method for determining patterns in the data, and presenting the data so that the similarity and differences are aptly highlighted. PCA implements an orthogonal transformation to change a set of conclusions of likely correlated variables into a set of values of linearly uncorrelated variables called principal components [39]. The transformation is characterized so that the first major component has the greatest possible deviation (that is, corresponds to as much variability in the data as possible).Therefore, we use principal components derived from PCA as privileged information in the proposed formulation. Tables 2 and 3 show that acquiring privileged information via PCA and incorporating the same in the proposed formulation is effective for linear and nonlinear kernels respectively. Hence, we implement principal components obtained from PCA as privileged information in the designed formulation. Tables 2 and 3 show that acquiring privileged information via PCA and incorporating the same in the proposed formulation is effective for linear and nonlinear kernels respectively.

The classification performance of the proposed Pin-TWSVMPI has been evaluated through average training accuracy over UCI datasets. It can be calculated as

$$\text{Testing Accuracy} = \frac{\text{Number of correctly classified testing samples}}{\text{Total number of samples in testing set}}. \tag{48}$$

Generalization error was determined by following the standard 5-fold cross-validation methodology [40]. The Average Testing Accuracy of all formulations for 5-fold cross-validation was represented by (mean ± std) in the following tables. Optimal values of c_1, c_2, γ, τ, and kernel parameter σ were obtained by using a tuning set comprising of 10 percent of the data set which is further sent back to the training dataset.

Table 1 summarizes different types of real-world datasets from the UCI machine learning repository [22]. These datasets include Iris, Compound, Wine, Ecoli, Dermatology, Zoo, Haberman, Soybean and Page Block. The missing values in the datasets were replaced by the mean value of their corresponding column

Table 2 and Table 3 analyse the classification accuracy of the proposed classifier Pin-TWSVMPI against SVMPI, TWSVMPI, Pin-SVM, Pin-SVMPI and Pin-TWSVM with linear and nonlinear kernel respectively. Figure 4 highlights the training times achieved by proposed Pin-TWSVMPI as compared to SVMPI, TWSVMPI, Pin-SVM, Pin-SVMPI, and Pin-TWSVM. This figure shows that the proposed Pin-TWSVMPI is faster among all. This is due to the SMO technique which is used to obtain the optimal classifier.

Table 2. Comparison of Average Testing Accuracy with standard deviation of different classifiers on UCI datasets with linear kernel

Dataset	Classifiers				
	SVMPI	**TWSVMPI**	**Pin-SVM**	**Pin-TWSVM**	**Pin-TWSVMPI**
Iris	95.65 ± 4.36	95.66 ± 4.21	95.89 ± 3.01	96.24 ± 3.87	**98.01 ± 2.87**
Compound	81.72 ± 4.07	82.02 ± 5.05	82.31 ± 4.87	82.78 ± 5.41	**84.78 ± 3.66**
Wine	95.45 ± 3.54	96.00 ± 3.58	96.12 ± 5.30	96.45 ± 4.65	**97.54 ± 3.86**
Ecoli	82.28 ± 3.35	82.76 ± 5.87	82.88 ± 4.77	83.04 ± 5.01	**85.68 ± 4.68**
Dermatology	93.01 ± 3.76	94.52 ± 2.64	94.21 ± 2.22	94.04 ± 4.02	**95.98 ± 3.38**
Zoo	91.65 ± 3.43	92.00 ± 3.14	92.41 ± 3.65	93.00 ± 2.06	**94.31 ± 2.11**
Glass	55.69 ± 4.11	56.66 ± 5.07	56.40 ± 4.78	56.68 ± 4.54	**58.62 ± 2.94**
Haberman	70.02 ± 4.12	70.16 ± 4.89	71.67 ± 3.36	71.74 ± 5.04	**73.88 ± 3.03**
Soybean	97.64 ± 1.08	**100.00 ± 0.00**	98.86 ± 2.11	**100.00 ± 0.00**	**100.00 ± 0.00**
Page Block	88.14 ± 1.21	90.47 ± 0.88	89.74 ± 1.34	90.11 ± 1.23	**92.14 ± 0.21**

Table 3. Comparison of Average Testing Accuracy with standard deviation of different classifiers on UCI datasets with nonlinear kernel

Dataset	Classifiers				
	SVMPI	**TWSVMPI**	**Pin-SVM**	**Pin-TWSVM**	**Pin-TWSVMPI**
Iris	96.61 ± 3.36	96.66 ± 5.32	96.78 ± 3.12	97.04 ± 4.01	**98.85 ± 4.08**
Compound	96.41 ± 3.98	97.21 ± 5.74	96.87 ± 5.78	97.85 ± 5.14	**98.93 ± 3.08**
Wine	98.12 ± 3.74	98.85 ± 4.02	98.45 ± 2.65	98.81 ± 3.78	**98.96 ± 4.14**
Ecoli	87.63 ± 4.87	88.00 ± 4.97	87.94 ± 5.47	88.50 ± 4.68	**90.11 ± 4.67**
Dermatology	94.89 ± 3.14	95.71 ± 2.78	95.00 ± 2.74	95.78 ± 3.47	**96.88 ± 2.09**
Zoo	94.00 ± 3.54	95.00 ± 3.28	94.15 ± 3.78	95.45 ± 5.09	**96.79 ± 4.08**
Glass	62.84 ± 3.12	63.33 ± 4.47	63.04 ± 2.68	63.84 ± 4.65	**65.61 ± 3.88**
Haberman	75.16 ± 4.08	75.40 ± 5.21	75.64 ± 4.98	76.14 ± 6.06	**77.78 ± 4.67**
Soybean	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
Page Block	91.76 ± 1.87	92.74 ± 1.44	90.75 ± 2.07	92.74 ± 1.21	**94.45 ± 1.00**

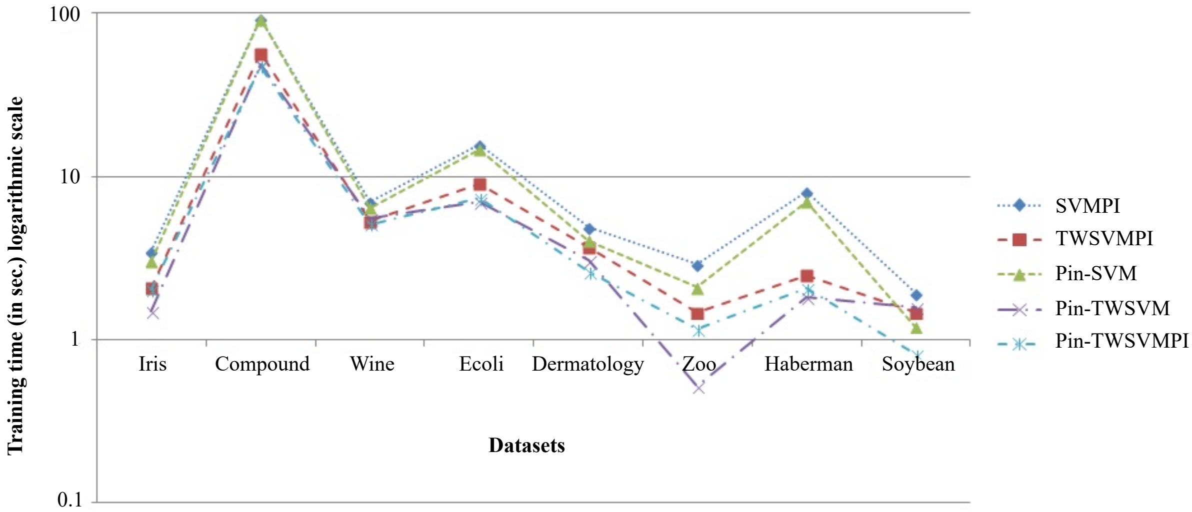


Figure 4. Comparison of training time with various methods

5.2 *Performance on pedestrian detection*

In this section, we used the standard INRIA dataset [23] to demonstrate the effectiveness of the pedestrian detector. INRIA dataset consists of 1,805 pedestrian images with the size of (64 × 128). In the experiment, we choose HOG as the standard input data, and HOF and co-occurrence matrix feature descriptors as the privileged information, respectively. Further, we choose 1,271 pedestrians images as the positive training samples, and negative samples are chosen from the rest of the images, which do not contain pedestrians. Figure 5 shows the sample images from INRIA dataset.

In literature, there are two established methodologies for evaluating pedestrian detection. One of them is per-window performance, and another one is per-image measure. In [41], the authors have shown that in practice per-window measure could fail to predict per-image performance. Therefore, in this paper, to assess the performance of the proposed Pin-TWSVMPI for pedestrian detection, we have incorporated the per-image measure method [41]. Generally, a detection system always has ground truth Bounding Box (BB_g), and it needs to take in an image as input and return a detected Bounding Box (BB_d) and a score or confidence for each detection. If the area of a BB_d and a BB_g overlap sufficiently then, they form a potential match. The overlap area between BB_d and BB_g is calculated as

$$ar_o = \frac{\text{area}(BB_d \cap BB_g)}{\text{area}(BB_d \cup BB_g)} > 0.5. \tag{49}$$

Here, each BB_d and BB_g is matched at most once. The BB_d and BB_g which are not matched to each other are counted as false positive and false negative respectively. To analyze different methods, we plot the miss rate against false positives per-image (FPPI) in log scale (lower curves indicate better performance) as shown in Figure 7. This figure concludes that privileged information plays a crucial role in pedestrian detection like frameworks where privileged information is already available. Figure 8 shows the bounding box obtained by various methods discussed in this paper over testing images.

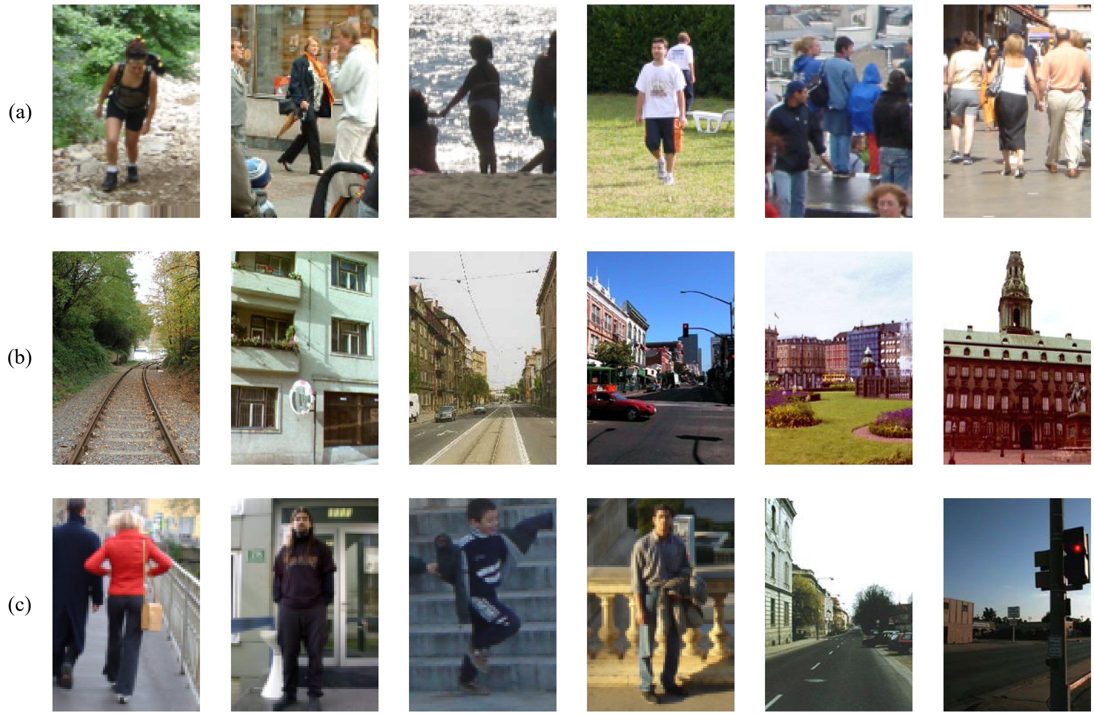


Figure 5. Sample images from benchmark INRIA dataset: (a) pedestrian training samples, (b) nonpedestrain training samples and (c) test images

Figure 6. Pedestrian detection over training set of INRIA dataset with Pin-TWSVMPI: (a) pedestrian training sample images and (b) bounding box obtained by proposed Pin-TWSVMPI where red and blue bounding box represents the positive and negative detection of pedestrian respectively.

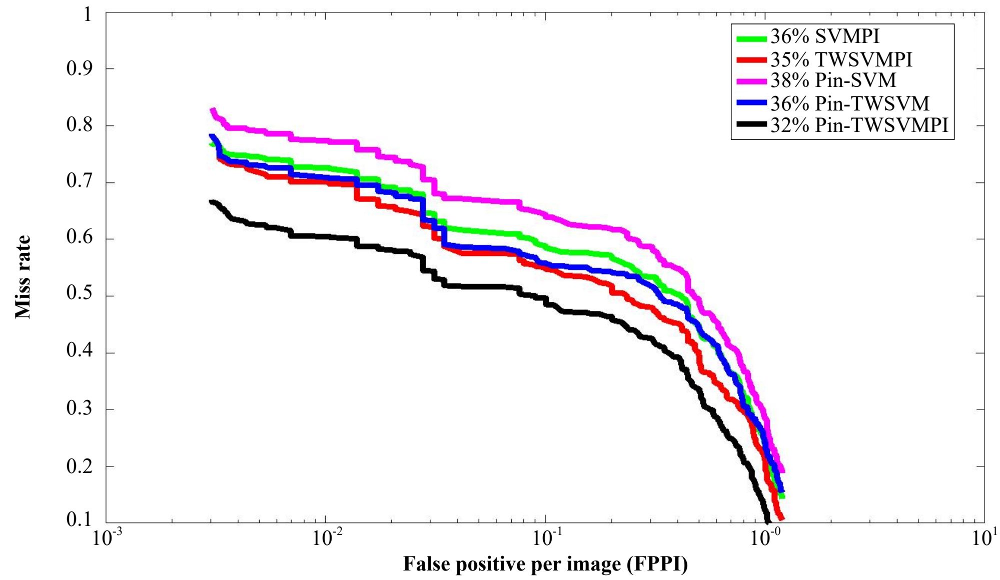


Figure 7. Comparison of pedestrian detection result of proposed Pin-TWSVMPI with SVMPI, TWSVMPI, Pin-SVM and Pin-TWSVM

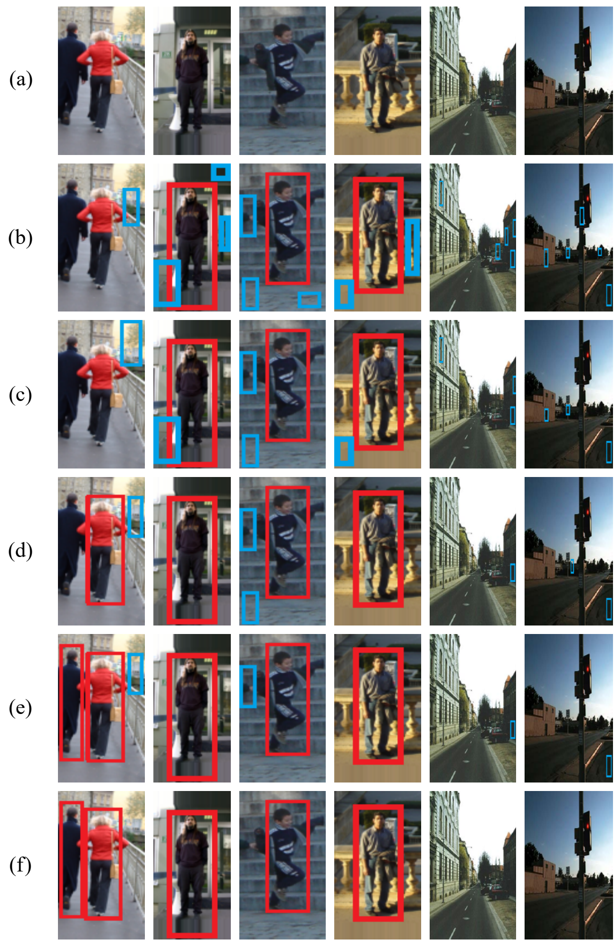


Figure 8. Pedestrian detection over a testing set of INRIA dataset: (a) pedestrian testing sample images, (b) bounding box obtained by Pin-SVM, (c) bounding box obtained by Pin-TWSVM, (d) bounding box obtained by SVMPI, (e) bounding box obtained by TWSVMPI, and (f) bounding box obtained by proposed Pin-TWSVMPI where red and blue bounding box represents the positive and negative detection of pedestrian respectively

5.3 *Performance on handwritten digit recognition*

MNIST contains 5,522 and 5,652 images of 5 and 8, respectively. The size of the images is 28 × 28 pixels. Classifying these two digits is an easy problem. Therefore, following the experimental setup of [2], to make it more difficult, we resized the digits to 10 × 10 pixel images. Figure 9 shows a sample image of the original 28 × 28 image and the corresponding 10 × 10 image. We choose 100 images as a training set, 4,000 images as a validation set (for tuning the parameters) and the rest 1,866 images as the test set.

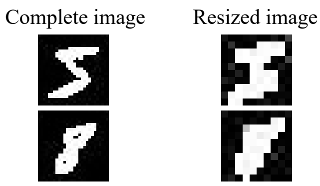


Figure 9. Sample digit image of first 5 and 8 along with the resized image

In Figure 10, we present the results by varying the number of training data. From Figure 10, we can observe that the proposed Pin-TWSVMPI outperforms SVMPI and FTWSVMPI in all cases. The average error rate of Pin-TWSVMPI is 1.92222 and 4.31111 lower than that of FTWSVMPI and SVMPI respectively. These results above show that the classifier decided by pinball loss function is superior to ones by a hinge loss function. In Figure 11, we show the comparison of training time among the discussed algorithms. More importantly, from Figure 11, we can find the training time speed of Pin-TWSVMPI is much faster than FTWSVMPI and SVMPI.

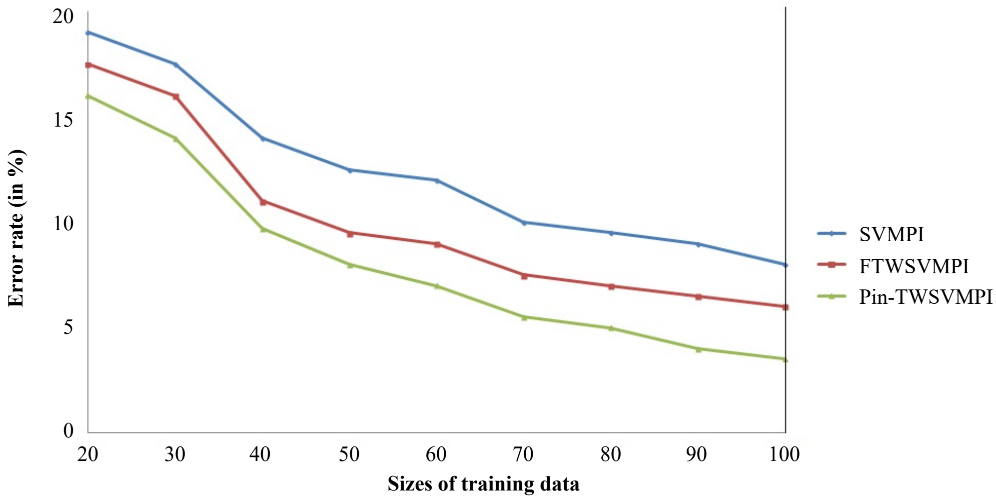


Figure 10. Error rate comparison among SVMPI, FTWSVMPI and Pin-TWSVMPI

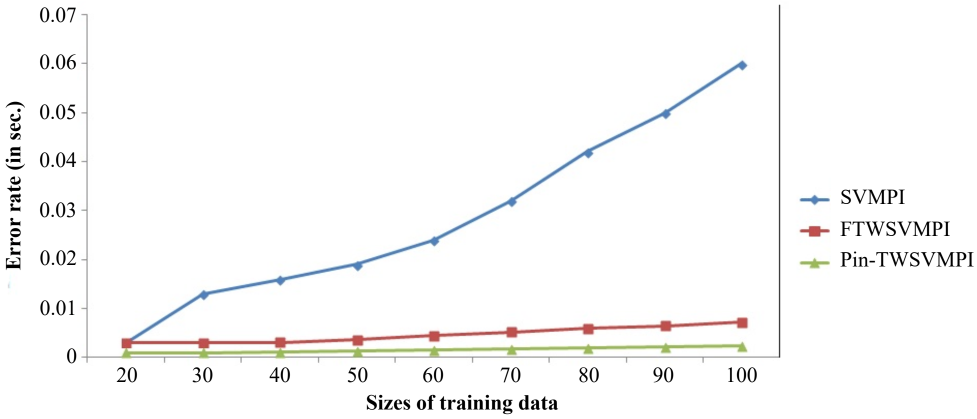


Figure 11. Training time comparison among SVMPI, FTWSVMPI and Pin-TWSVMPI

6. Conclusions

In this paper, we have proposed a fast Twin Support Vector Machine based on privileged information with pinball loss classifier (termed as Pin-TWSVMPI) and have shown their applications over pedestrian detection and handwritten digit recognition. Some major properties of pinball loss, i.e., noise insensitivity and re-sampling ability over other loss functions motivate us to build Pin-TWSVM using privileged information. As far as we know, it is the first time that privileged information has been incorporated to improve the performance of Pin-TWSVM classifier. The privileged information is identified from different individual expert's knowledge, which enhanced the generalization performance of the proposed framework. In order to incorporate the SRM principle along with ERM principle, we introduced a regularization term in the objective function of Pin-TWSVMPI. We have also used SMO to solve the dual problem of the proposed formulation which helps in solving Pin-TWSVMPI efficiently. This framework also provides the freedom to integrate information obtained from multiple expert's knowledge without building a separate new model which also improves the training time of the classifier. We have also implemented the novel method given in [20] which extracts privileged information by avoiding to solve two additional QPPs as in FTWSVMPI where expert's knowledge is not available. The experiment results on several UCI benchmark datasets show that our proposed method achieves better classification accuracy than other algorithms, i.e., Pin-TWSVM and TWSVMPI with considerably lesser computational time. As an application to the proposed Pin-TWSVMPI model, we also did experiments for pedestrian detection over INRIA dataset and handwritten digit recognition over MNIST dataset.

Acknowledgment

We are thankful to the Machine Learning Lab of South Asian University for providing the required infrastructure which helped us to complete the work smoothly.

Conflict of interest

We authors hereby declare that we do not have any conflict of interest with the content of this manuscript.

References

[1] Lapin M, Hein M, Schiele B. Learning using privileged information: SVM+ and weighted SVM. *Neural Networks*. 2014; 53: 95-108. Available from: doi: 10.1016/j.neunet.2014.02.002.

[2] Vapnik V, Vashist A. A new learning paradigm: Learning using privileged information. *Neural Networks*. 2009; 22(5-6): 544-557. Available from: doi: 10.1016/j.neunet.2009.06.042.

[3] Xie Z, Xu Y, Hu Q. Uncertain data classification with additive kernel support vector machine. *Data & Knowledge Engineering*. 2018; 117: 87-97. Available from: doi: 10.1016/j.datak.2018.07.004.

[4] Phan AV, Chau PN, Le Nguyen M, Bui LT. Automatically classifying source code using tree-based approaches. *Data & Knowledge Engineering*. 2018; 114: 12-25. Available from: doi: 10.1016/j.datak.2017.07.003.

[5] Saha S, Ekbal A. Combining multiple classifiers using vote based classifier ensemble technique for named entity recognition. *Data & Knowledge Engineering*. 2013; 85: 15-39. Available from: doi: 10.1016/j.datak.2012.06.003

[6] Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995; 20(3): 273-297. Available from: doi: 10.1007/BF00994018.

[7] Vapnick VN. *Statistical learning theory*. New York: Wiley; 1998.

[8] Jayadeva, Khemchandani R, Chandra S. Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2007; 29(5): 905-910. Available from: doi: 10.1109/TPAMI.2007.1068.

[9] Mangasarian OL, Wild EW. Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005; 28(1): 69-74. Available from: doi: 10.1109/TPAMI.2006.17

[10] Kumar MA, Gopal M. Least squares twin support vector machines for pattern classification. *Expert Systems with Applications*. 2009; 36(4): 7535-7543. Available from: doi: 10.1016/j.eswa.2008.09.066.

[11] Wu Q, Zang BY, Zhang Y, Qi ZX. Wavelet Kernel Twin Support Vector Machine. *Journal of Information Hiding and Multimedia Signal Processing*. 2021; 12(2): 93-101.

[12] Huang X, Shi L, Suykens JAK. Support vector machine classifier with pinball loss. *IEEE transactions on pattern analysis and machine intelligence*. 2014; 36(5): 984-997. Available from: doi: 10.1109/TPAMI.2013.178.

[13] Mehrkanoon S, Huang X, Suykens JAK. Non-parallel support vector classifiers with different loss functions. *Neurocomputing*. 2014; 143: 294-301. Available from: doi: 10.1016/j.neucom.2014.05.063.

[14] Nasiri JA, Charkari NM, Mozafari K. Energy-based model of least squares twin Support Vector Machines for human action recognition. *Signal Processing*. 2014; 104: 248-257. Available from: doi: 10.1016/j.sigpro.2014.04.010.

[15] Zhang S, Benenson R, Omran M, Hosang J, Schiele B. How far are we from solving pedestrian detection? In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2016. p.1259-1267. Available from: doi: 10.1109/CVPR.2016.141.

[16] Min W, Zhang Y, Li J, Xu S. Recognition of pedestrian activity based on dropped-object detection. *Signal Processing*. 2018; 144: 238-252. Available from: doi: 10.1016/j.sigpro.2017.09.024.

[17] Li Y, Sun H, Yan W. Domain adaptive twin support vector machine learning using privileged information. *Neurocomputing*. 2022; 469: 13-27. Available from: doi: 10.1016/j.neucom.2021.10.069.

[18] Xu R, Wang H. Multi-view learning with privileged weighted twin support vector machine. *arXiv* [Preprint] 2022. Available from: doi: 10.48550/arXiv.2201.11306.

[19] Qi Z, Tian Y, Shi Y. A new classification model using privileged information and its application. *Neurocomputing*. 2014; 129:146-152. Available from: doi: 10.1016/j.neucom.2013.09.045.

[20] Pal A, Rastogi (nee Khemchandani) R. Learning TWSVM using Privilege Information. In: *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE; 2018. p.1548-1554. Available from: doi: 10.1109/SSCI.2018.8628645.

[21] Platt J. *Sequential minimal optimization: A fast algorithm for training support vector machines*. Microsoft Research. Report number: MSR-TR-98-14, 1998.

[22] Asuncion A, Newman D. *UCI machine learning repository*. [Dataset] http://archive.ics.uci.edu/ml/index.php. 2007.

[23] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE; 2005. p.886-893 vol. 1. Available from: doi: 10.1109/CVPR.2005.177.

[24] Mangasarian OL. *Nonlinear programming*. USA: Society for Industrial and Applied Mathematics; 1994.

[25] Gunn SR. Support vector machines for classification and regression. *ISIS technical report*. 1998; 14(1): 5-16.

[26] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001. IEEE; 2001. p.1-9. Available from: doi: 10.1109/CVPR.2001.990517.

[27] Zhang S, Bauckhage C, Klein DA, Cremers AB. Exploring human vision driven features for pedestrian detection. *IEEE Transactions on Circuits and Systems for Video Technology*. 2015; 25(10): 1709-1720. Available from: doi: 10.1109/TCSVT.2015.2397199.

[28] Zhang S, Benenson R, Schiele B. Filtered channel features for pedestrian detection. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2015. p.1751-1760. Available from: doi: 10.1109/CVPR.2015.7298784.

[29] Paisitkriangkrai S, Shen C, Hengel AVD. Strengthening the effectiveness of pedestrian detection with spatially pooled features. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T. (eds.) *European Conference on Computer Vision*. Springer, Cham: 2014. p.546-561.

[30] Park D, Zitnick CL, Ramanan D, Dollár P. Exploring weak stabilization for motion feature extraction. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2013. p.2882-2889. Available from: doi: 10.1109/CVPR.2013.371.

[31] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*. 1973; 6:610-621. Available from: doi: 10.1109/TSMC.1973.4309314.

[32] Das N, Sarkar R, Basu S, Kundu M, Nasipuri M, Basu DK. A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application. *Applied Soft Computing*. 2012; 12(5): 1592-1606. Available from: doi: 10.1016/j.asoc.2011.11.030.

[33] Liu CL, Nakashima K, Sako H, Fujisawa H. Handwritten digit recognition: benchmarking of state-of-the-art techniques. *Pattern recognition*. 2003; 36(10): 2271-2285. Available from: doi: 10.1016/S0031-3203(03)00085-2.

[34] Sharma A. Handwritten digit recognition using support vector machine. *arXiv* [Preprint] 2012. Available from: doi: 10.48550/arXiv.1203.3847.

[35] Sadri J, Suen CY, Bui TD. Application of support vector machines for recognition of handwritten Arabic/Persian digits. In: *Proceedings of Second Iranian Conference on Machine Vision and Image Processing*. CIVILICA; 2003. p.300-307 vol. 1.

[36] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998; 86(11): 2278-2324. Available from: doi: 10.1109/5.726791.

[37] LeCun Y, Cortes C, Burges CJC. *The MNIST database of handwritten digits*. http://yann.lecun.com/exdb/mnist/ [Accessed 15th June 2022].

[38] Abdi H, Williams LJ. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*. 2010; 2(4): 433-459. Available from: doi: 10.1002/wics.101.

[39] Smith LI. *A tutorial on Principal Components Analysis (Computer Science Technical Report No. OUCS-2002-12)*. http://hdl.handle.net/10523/7534 [Accessed 23rd May 2022].

[40] Duda RO, Hart PE, Stork DG. *Pattern classification. New York*: Wiley; 1973.

[41] Dollár P, Wojek C, Schiele B, Perona P. Pedestrian detection: A benchmark. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009. p.304-311. Available from: doi: 10.1109/CVPR.2009.5206631.