

Research Article

Twin Support Vector Machine for Multiple Instance Learning Based on Bag Dissimilarities

Divya Tomar and Sonal Agarwal

Indian Institute of Information Technology, Allahabad, Uttar Pradesh 211012, India

Correspondence should be addressed to Divya Tomar; divyatomar26@gmail.com

Received 31 March 2016; Revised 7 June 2016; Accepted 11 July 2016

Academic Editor: Elpida Keravnou

Copyright © 2016 D. Tomar and S. Agarwal. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In multiple instance learning (MIL) framework, an object is represented by a set of instances referred to as bag. A positive class label is assigned to a bag if it contains at least one positive instance; otherwise a bag is labeled with negative class label. Therefore, the task of MIL is to learn a classifier at bag level rather than at instance level. Traditional supervised learning approaches cannot be applied directly in such kind of situation. In this study, we represent each bag by a vector of its dissimilarities to the other existing bags in the training dataset and propose a multiple instance learning based Twin Support Vector Machine (MIL-TWSVM) classifier. We have used different ways to represent the dissimilarity between two bags and performed a comparative analysis of them. The experimental results on ten benchmark MIL datasets demonstrate that the proposed MIL-TWSVM classifier is computationally inexpensive and competitive with state-of-the-art approaches. The significance of the experimental results has been tested by using Friedman statistic and Nemenyi post hoc tests.

1. Introduction

Standard pattern recognition problems consider that the objects are represented as a single feature vector which contains sufficient information for the recognition of these objects. However, some complex objects exist in the real world which are difficult to represent by using a single feature vector; that is, single feature vector representation of an object is not sufficient for its separability, for example, a document with several paragraphs, an image containing many regions, each with different characteristics, and a drug with various conformations of a molecule. Traditional supervised learning techniques handle such kind of problems by representing complex objects using single feature vector. This reduction may lose significant information which further degrades the performance of supervised learning techniques. A set of feature vectors or multiple instances representation can be used for the better understanding of complex object [1, 2]. Multiple instances representation of a complex object can preserve more information about it. MIL is a variation of supervised learning in which a classifier is trained on a set of instances known as bag instead of individual instance. The

objective of MIL approaches is to predict the class label for a bag. A bag may have different number of instances and may belong to positive or negative class label. Positive class label is assigned to a bag if it contains at least one positive instance while negative class label is assigned to a bag when all of its instances are negative [1, 3–5]. Figure 1 illustrates the framework of single instance learning and multiple instance learning.

From Figure 1, it is observed that, in single instance learning, each object is represented by a single feature vector or instance and the classifier learns at instance level by assigning the class label to each instance individually. However, in MIL framework, an object is represented by a set of feature vectors or a bag and the classifier trains at bag level instead of instance level and predicts the class label of a bag instead of an instance.

The term MIL was first used for the drug activity prediction problem (Musk odor prediction) [1]. Later on, it is widely used by the researchers to solve various real world problems like image annotation [6–9], document categorization [6, 10, 11], object detection [12, 13], human action recognition [14], visual tracking [15–18], spam filtering [19], and many other

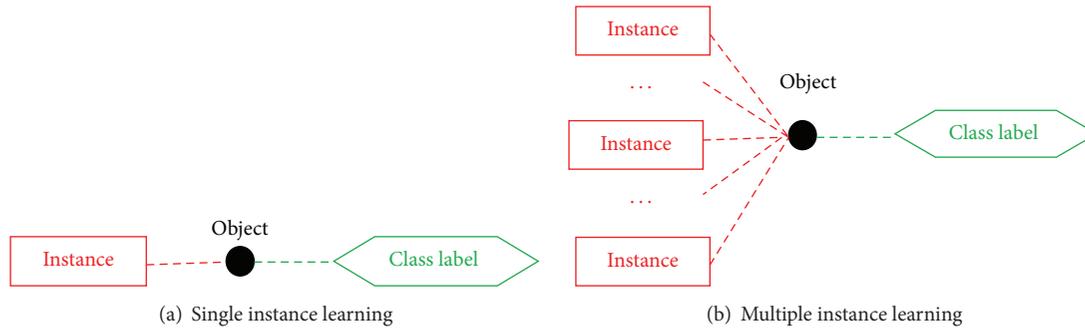


FIGURE 1: Learning frameworks.

problems. Several MIL approaches have been proposed by the researchers which can be broadly categorized into two groups. The first category, also known as bag-based method, works only on the bag label without having any knowledge of each instance label. Thus the bag labels can be predicted by converting a bag into a single instance representation and using supervised algorithms or by defining kernels or distances between bags. MIL approaches belonging to first category include those of Chen et al. [7], Sørensen et al. [20], and Cheplygina et al. [21] that use some dissimilarity measures to represent a bag with a derived feature vector. Wang and Zucker [22] defined nearest neighbors among bags, Gärtner et al. [23] and Wang et al. [24] determined kernels between bags, Zhou et al. [10] generated a graph with instance from a bag, and Zhang et al. [25] incorporated structure information between bags and many more. On the other hand, the second category also known as instance-based methods focuses on the instance label and a bag label is determined by combining the classification of instances. Axis-parallel rectangle method [1] and Diverse Density [2] and its variation [26] are some examples of instance-based approaches. The bag-based methods are widely used by the researchers as they have shown better performance on a wide range of MIL datasets. Therefore, this study has focused on the first category and extended the recently proposed Twin Support Vector Machine (TWSVM) classifier to multiple instance learning scenarios by obtaining summarized information of each bag using different dissimilarity measures [27].

In recent years, many nonparallel hyperplane Support Vector Machine (SVM) classifiers are proposed by the researchers for binary classification [27–29]. For example, Mangasarian and Wild proposed a Generalized Eigenvalue Proximal Support Vector Machine (GEPSSVM), the first nonparallel hyperplane classifier, which aims to find a pair of nonparallel hyperplane in such a way that each hyperplane is nearest to one of the two classes and as far as possible from the other classes [30]. GEPSSVM shows excellent performance with several benchmark datasets especially with the “Cross-Planes” dataset. Later, by utilizing the concept of traditional SVM and GEPSSVM, Jayadeva et al. proposed a nonparallel hyperplane based novel binary classifier, named as TWSVM [27]. The aim of TWSVM classifier is to generate two nonparallel hyperplanes in such a way that each hyperplane lies in close affinity to one of the two classes while maintaining

distant from the data instance of other classes. For this purpose, it solves two SVM-type Quadratic Programming Problems (QPPs), while GEPSSVM solves two generalized eigenvalue problems. Since TWSVM solves two smaller QPPs as opposed to a single complex QPP, the learning of TWSVM classifier is four times faster than that of standard SVM. TWSVM has shown its superiority over the other existing machine learning approaches on several benchmark datasets. Therefore, in this study, we have extended TWSVM to multiple instance learning scenarios. This paper proposes a bag dissimilarity based multiple instance learning TWSVM (MIL-TWSVM) classifier. We have defined the dissimilarity between bags using different approaches. The proposed classifier is trained with the summarized information of instances in each bag where a bag is represented by a feature vector. Feature vector contains the dissimilarity scores of a bag derived from the other bags in the training set. The experiment has been performed on ten MIL benchmark datasets. The results of the proposed approach have been compared with several existing MI learning approaches, such as Diverse Density (DD) [2], Expectation-Maximization Diverse Density (EMDD) [26], Multi-instance Logistic Regression (MILR) [31], Citation k -NN [22], and Multi-Instance Support Vector Machine (MISVM) [6]. The effectiveness of the proposed approach has also been analyzed by using Friedman average rank hypothesis tests [32, 33]. The statistical inferences are made from the observed difference in predictive accuracy. Modified version of Demšar significance diagram has been used to display the output of Friedman test.

The rest of the paper is organized as follows. Section 2 provides a brief overview of multiple instance learning approaches and their applications in the real world. Section 3 includes the formulation of Twin Support Vector Machine classifier. Section 4 describes different approaches used to measure the dissimilarity between bags. The experimental results are discussed in Section 5 and finally the conclusion is drawn in Section 6.

2. Overview of Multiple Instance Learning Approaches and Their Applications

In Multiple Instance Learning (MIL), a bag is used to represent an object as follows:

$$B_i = \{x_{il} \mid l = 1, 2, \dots, n_i\} \in R^d, \quad (1)$$

where n_i is the number of instances or feature vectors in bag B_i and R^d is the d -dimensional feature space. Consider the training dataset contains N bags. The training dataset for MIL is represented as

$$T = \{(B_i, y_i) \mid i = 1, \dots, N\}, \quad (2)$$

where $y_i = \{+1, -1\}$ is the class label corresponding to each bag in the dataset. The bags are labeled with class +1 if and only if it contains at least one positive instance; otherwise it is considered as negative labeled bag. The objective of MIL problem is to learn a model which can determine the class label of the unseen bag. MIL approaches have been widely used by the researchers to solve many real world problems. The drug activity recognition is one of the most popular applications of it. In this problem, for a given chemical molecule, the system must decide if it is useful for drug design or not. A good drug has the characteristic that it is strongly bound to a target “binding site.” A molecule can adopt multiple conformations or shapes and only one or a few of them bind well with the target protein or binding site. Dietterich et al. modeled the MIL framework to deal with the drug activity recognition [1]. They have predicted whether a new molecule was suitable for drug design or not by analyzing a set of known molecules. They have developed a three-axis-parallel rectangle algorithm in which the combination of extracted molecule features was used to determine the axis-parallel rectangles (APR). Zhao et al. proposed an MIL approach based on joint instance and feature selection for drug activity prediction [34]. They have focused on irrelevant and redundant features reduction in order to improve the interpretability of the drug activity recognition model. Diverse Density (DD) has been proposed to measure a region in the feature space which consists of at least one instance from each positive bag while there are no instances from negative bags [2]. DD of a given point p is defined as the ratio of the number of positive bags which have instances near to this point and the sum of distances of negative instances from p . The point p^* at which DD is maximized corresponds to the target concept. DD has suffered from local optimization problem and the best solution could be achieved by many restarts. An algorithm Expectation-Maximization Diverse Density (EMDD) has been proposed by Zhang and Goldman to solve this problem [26]. They have combined expectation-maximization approach with DD to solve the local optimum problem by iteratively updating the previous target point. Expectation step finds the most positive instance from each bag after an initial guess for the target point p . Then, maximization step searches for a new point by maximizing DD on the selected most positive instances. These steps are repeated until the algorithm converges. EMDD performs well on a variety of MIL problems, but it is also very computationally intensive. Several regular supervised classifiers have also been extended to MIL scenarios, for example, Citation k -Nearest Neighbor (Citation k -NN), Bayesian k -NN, ID3-MI, and multiple instance learning SVM. Citation k -NN is an extension of k -NN in which a bag has been labeled by

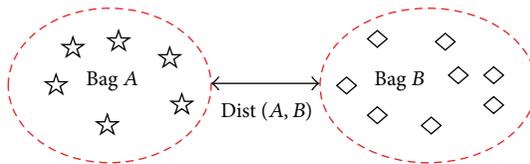


FIGURE 2: Distance between different bags.

analyzing its neighboring bags and the bags that consider the concerned bag as a neighbor [22]. Citation k -NN uses different distance metric (minimal Hausdorff distance) in which the focus has been shifted from instances to the bags; that is, the distance is measured between different bags instead of different instances (see Figure 2). Citation k -NN has shown better performance on Musk datasets. Recently, a variance of Citation k -NN has been proposed by Zhou et al. for web mining task in which the minimum Hausdorff distance has been modified for text features [35]. ID3-MI is a decision tree algorithm which uses multi-instance entropy criterion to split the tree nodes [36].

SVM has been extended to MIL scenario by Andrews et al. [6]. They have used two approaches for the extension of SVM. In the first approach, traditional SVM has been extended to MIL scenarios in which hidden labels of instances are decided under constraints posed by the class labels of bags. In the second approach, the objective was to maximize the bag margin directly. MIL Boost is another example of MIL approach where the weights of instances are updated in each of the boosting rounds [12]. In this approach, Noisy-OR rule is used to determine the bag labels from given instance labels. Logistic Regression and Neural Network are also extended to the MIL framework. Logistic Regression is a popular probabilistic supervised learning approach which has been extended to MIL framework. Fu and Robles-Kelly extended Logistic Regression (LR) to MIL problem domains and combine L_1 and L_2 regularization methods [37]. Xu and Frank also upgraded single instance LR to multi-instance data and showed its effectiveness on artificial and Musk drug activity prediction dataset [31]. They have followed several assumptions to form the bag-level probability from instance-level class probabilities. Ramon and De Raedt have also explored the utility of NN for MIL due to its ability of automatic learning from examples [38]. In another research work, Zhou and Zhang proposed BP-MLP which is the extension of NN to MIL [39]. In this approach, traditional BP algorithm has been extended using a global error function which is defined at bag level rather than at instance level. Image classification and retrieval is another significant application of MIL in which a given image is to be classified into a target category on the basis of its visual content [40–42]. In MIL, an image can be viewed as a bag of local image patches and can be labeled as positive or negative. A positive label image contains a set of image patches or instances in which at least one patch is conceptual to the user while if all the patches are not conceptual to the user then image is considered as negative label image. For example, in beach scene classification, the target class is beach and by

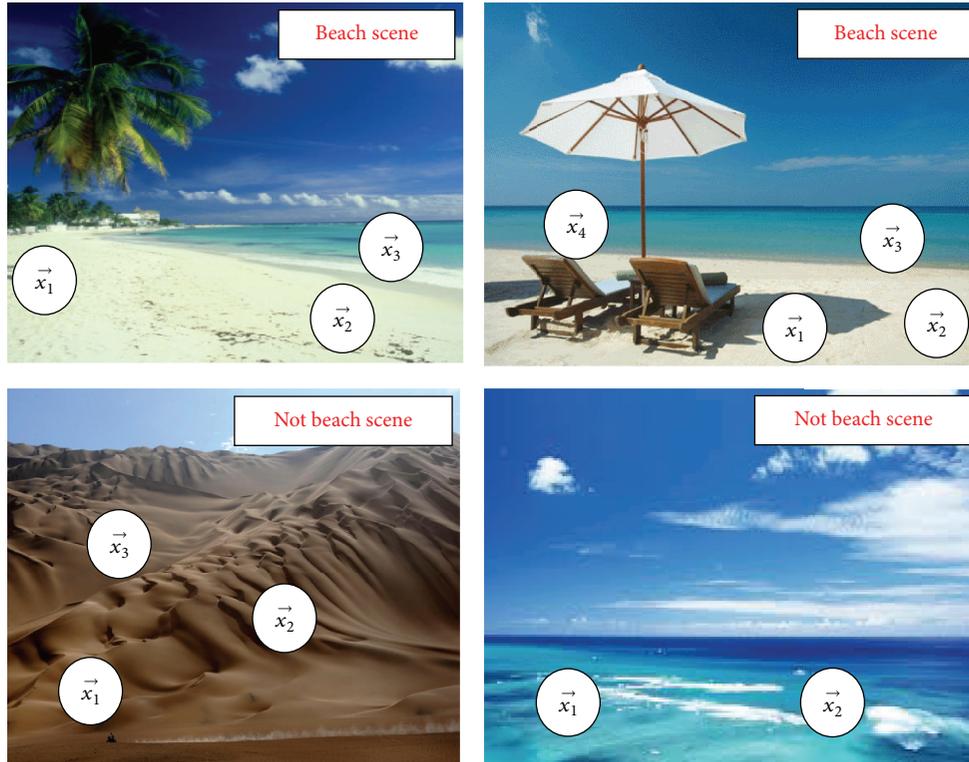


FIGURE 3: Classification of a scene image into beach and not beach scene.

using different regions or contents of the scene image (see Figure 3), the objective is to recognize whether it belongs to beach scene or not. In this case, any visual content that displays a beach may be considered as positive image while negative images show another different visual content.

Xu proposed an MIL extension of Neural Network for the retrieval and classification of images [43]. Maron and Ratan applied DD MIL approach for the classification of natural image scene by using different kinds of bag generators [44]. Bag generators consider each image as a bag and various subregions in the image as instances. They have performed experiment on COREL photo library. Cheng et al. proposed BP-MLP and BP-SVM approaches for automatic image categorization in multiple instance learning scenario and performed experiment on 2000 images obtained from COREL repository [45]. They have extracted frequent patterns from each image category and embedded an image bag into a multidimensional data point which is useful to characterize the similarity between the image and every common pattern of an image category. Gondra and Xu proposed a Relevance Feedback (RF) learning based Content Based Image Retrieval (CBIR) framework in multiple instance learning scenario [8]. In another research work, Pao et al. proposed an EMDD based MIL method for image classification [46]. Sener and Ikitler-Cinbis proposed an ensemble of multiple instance learning approaches for the problem of image reranking [47]. They have constructed bags by using three different approaches: sliding window and dynamic and dynamic-sliding methods. Then the constructed bags have been used

to develop multi-instance classifiers. They have used multiple instance learning with instance selection (MILES) algorithm as MIL-classifiers. Rank was assigned to an image by combining the decision score of MIL-classifiers. Li and Liu used graph based MIL with instance weighting for the retrieval of images [48]. Different weights were assigned to each region in positive images on the basis of learning results and then rank was calculated for each image. Feng et al. proposed a multi-instance semisupervised learning approach on the basis of hierarchical sparse representation for the categorization of images [49]. They have solved the instance confidence value identification problem under the framework of instance-level sparse representation. Several other research works have also focused on the image categorization problem in MIL framework. Xu et al. utilized the concept of deep learning of feature representation with multiple instance learning for colon cancer classification based on histopathology images [50]. The deep learning network is a process of obtaining high level features from low level features. They have proposed a system based on deep learning having a set of linear filters in encoder and decoder and used the last hidden layer of deep learning as fully supervised feature learning, as it represents intrinsic features compared to lower level features. Wu et al. extended deep learning to multiple instance learning framework for image annotation [51]. They have used deep convolutional neural network which contains five convolutional layers, followed by a pooling layer and three fully connected layers for learning visual representation with multiple instance learning. The last hidden layer was

redesigned for multiple instance learning. Kotzias et al. have also combined the concept of deep learning and multi-instance learning for knowledge transfer [52]. MIL has also been utilized in disease diagnosis by analyzing medical images. Ding et al. considered the breast ultrasound image classification task as multiple instance learning tasks and proposed an MIL method based on SVM which classifies the tumors into benign and malignant [53]. They have used self-organizing map (SOM) to map the instance space into concept-space and constructed the bag feature vector by using the distribution of the instances of each bag. Li et al. proposed a novel computer aided diagnosis scheme for the recognition of tumor invasion depth of gastric cancer [54]. They have extracted both bag-level and instance-level features and applied an improved citation k -NN algorithm for the identification of gastric tumor invasion depth. Tong et al. used MIL method for the detection of Alzheimer's disease (AD) and its prodromal stage mild cognitive impairment (MCI) [55]. They have built a graph for each image to identify the relationships among the patches and performed experiment on 834 MRI images taken from ADNI study. In another research work, Quellec et al. proposed an MIL framework for diabetic retinopathy screening [56]. Text categorization is another popular application of MIL. Wang et al. [11] proposed a novel instance specific distance method for the application of MIL text categorization. They have derived this data from Reuters-21578 collection having 2000 bags with 243 features. He and Wang investigated the problem of text categorization from multiple instance view in which each text is considered as a bag and each of its sentences as instance [57]. They have developed an MIL approach for Chinese text classification using k -NN. MIL is also used for web mining or web index recommendation problem in which each web page is considered as a bag and each of its linked pages is considered as bag instances. Viola et al. [12] proposed an algorithm, known as Fretcit k -NN based on minimum Hausdorff dissimilarity measure, and determined the class label of unseen bag by utilizing both references and citers. MIL approaches also have significant contribution to visual tracking [15, 17, 58] and real time video event detection areas [59].

3. Twin Support Vector Machine

Twin Support Vector Machine is a binary classification technique that does classification of data instances by constructing two nonparallel hyperplanes instead of a single hyperplane as in the case of traditional Support Vector Machine. It obtains two nonparallel hyperplanes by solving two QPPs of smaller size as compared to a single complex QPP solved by traditional SVM. TWSVM generates hyperplane for each class in such a way that the data instances of each class lie in close affinity to its corresponding hyperplane and as far as possible from the other hyperplane. The effectiveness of TWSVM over other existing classification approaches has been validated on various benchmark datasets. TWSVM has better generalization ability and faster computational speed due to which it has been applied to several real life applications such as intrusion detection [60, 61], activity

recognition [62], image denoising [63], emotion recognition [64], text classification [65], defect prediction [66, 67], disease diagnosis [68, 69], and speaker identification [70]. Consider a binary classification problem of " m " size. The training dataset for such kind of problem can be represented as

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, \quad (3)$$

where $x_i \in R^n$, $i = 1, 2, \dots, m$, represents input data instances in n -dimensional feature space R and $y_i \in \{+1, -1\}$ indicates corresponding class label. Consider two matrices $A_{+1} \in R^{m_1 \times n}$ and $A_{-1} \in R^{m_2 \times n}$ comprising the data instance of class +1 and class -1, respectively. TWSVM solves the following two QPPs:

$$\min_{w_{+1}, b_{+1}, \xi} \frac{1}{2} \|A_{+1}w_{+1} + e_{+1}b_{+1}\|^2 + c_1 e_{-1}^T \xi, \quad (4)$$

$$\text{s.t.} \quad -(A_{-1}w_{+1} + e_{-1}b_{+1}) + \xi \geq e_{-1}, \quad \xi \geq 0,$$

$$\min_{w_{-1}, b_{-1}, \eta} \frac{1}{2} \|A_{-1}w_{-1} + e_{-1}b_{-1}\|^2 + c_2 e_{+1}^T \eta, \quad (5)$$

$$\text{s.t.} \quad (A_{+1}w_{-1} + e_{+1}b_{-1}) + \eta \geq e_{+1}, \quad \eta \geq 0,$$

and seeks the following two nonparallel hyperplanes in R^n :

$$x^T w_{+1} + b_{+1} = 0, \quad (6)$$

$$x^T w_{-1} + b_{-1} = 0.$$

Here, w_{+1} and w_{-1} are normal vectors to the hyperplanes; b_{+1} and b_{-1} represent the bias terms. $e_{+1} \in R^{m_1}$ and $e_{-1} \in R^{m_2}$ are two vectors of 1's of appropriate dimensions. c_1 and c_2 are two positive trade-off constants. $\xi \in R^{m_1}$ and $\eta \in R^{m_2}$ are slack variables due to the class -1 and class +1, respectively. The first term of (4) or (5) is the sum of squared distances of data instances from their corresponding hyperplane. Minimization of this term keeps the hyperplane closest to the data instances of class +1 or class -1. The second term of (4) or (5) assigns penalty to the data instances of other classes which are misclassified. The constraints require the hyperplane to be maintained at least 1 distance from the data instances of other classes. Slack variable measures the error wherever the hyperplane is closer than the 1 distance. In this way, the hyperplane is kept closer to the data instances of its respective class and as far as possible from the data instances of other classes. The Lagrangian corresponding to (4) is given as follows:

$$\begin{aligned} L(w_{+1}, b_{+1}, \xi, \alpha) &= \frac{1}{2} \|A_{+1}w_{+1} + e_{+1}b_{+1}\|^2 + c_1 e_{-1}^T \xi \\ &\quad - \alpha^T (-(A_{-1}w_{+1} + e_{-1}b_{+1}) + \xi - e_{-1}) - \beta^T \xi, \end{aligned} \quad (7)$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{m_2})^T$ and $\beta = (\beta_1, \beta_2, \dots, \beta_{m_2})^T$ are two vectors of Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions are given by

$$A_{+1}^T (A_{+1} w_{+1} + e_{+1} b_{+1}) + A_{-1}^T \alpha = 0, \quad (8)$$

$$e_{+1}^T (A_{+1} w_{+1} + e_{+1} b_{+1}) + e_{-1}^T \alpha = 0, \quad (9)$$

$$c_1 e_{-1} - \alpha - \beta = 0, \quad (10)$$

$$-(A_{-1} w_{+1} + e_{-1} b_{+1}) + \xi \geq e_{-1}, \quad (11)$$

$$\xi \geq 0,$$

$$\alpha^T (-(A_{-1} w_{+1} + e_{-1} b_{+1}) + \xi - e_{-1}) - \beta^T = 0, \quad (12)$$

$$\beta^T \xi = 0,$$

$$\alpha \geq 0, \quad (13)$$

$$\beta \geq 0.$$

Since $\beta \geq 0$, from (10), we can determine

$$0 \leq \alpha \leq c_1. \quad (14)$$

Equations (8) and (9) lead to

$$\begin{bmatrix} A_{+1}^T \\ e_{+1}^T \end{bmatrix} [A_{+1} \ e_{+1}] \begin{bmatrix} w_{+1} \\ b_{+1} \end{bmatrix} + \begin{bmatrix} A_{-1}^T \\ e_{-1}^T \end{bmatrix} \alpha = 0. \quad (15)$$

Let $B_{+1} = [A_{+1} \ e_{+1}]$, $B_{-1} = [A_{-1} \ e_{-1}]$, and $u_{+1} = \begin{bmatrix} w_{+1} \\ b_{+1} \end{bmatrix}$. The above equation becomes

$$B_{+1}^T B_{+1} u_{+1} + B_{-1}^T \alpha = 0 \quad (16)$$

$$\text{or } u_{+1} = -(B_{+1}^T B_{+1})^{-1} B_{-1}^T \alpha.$$

In similar manner,

$$u_{-1} = (B_{-1}^T B_{-1})^{-1} B_{+1}^T \gamma. \quad (17)$$

From (16) and (17), it is clear that the solution of hyperplane parameters requires the inverse of matrix $B_{+1}^T B_{+1}$ and $B_{-1}^T B_{-1}$. Sometimes, matrix may be ill-conditioned due to which it is difficult to calculate its inverse. To avoid this situation, regularization terms $c_3 I$ and $c_4 I$ are added to the above-mentioned equations as follows:

$$u_{+1} = -(B_{+1}^T B_{+1} + c_3 I)^{-1} B_{-1}^T \alpha, \quad (18)$$

$$u_{-1} = (B_{-1}^T B_{-1} + c_4 I)^{-1} B_{+1}^T \gamma,$$

where $c_3, c_4 > 0$ are user defined parameters having small values and I is an identity matrix of appropriate dimension. Wolfe dual of (4) and (5) can be defined as

$$\begin{aligned} \max_{\alpha} \quad & e_{-1}^T \alpha - \frac{1}{2} \alpha^T B_{-1} (B_{+1}^T B_{+1} + c_3 I)^{-1} B_{-1}^T \alpha, \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1, \\ \max_{\gamma} \quad & e_{+1}^T \gamma - \frac{1}{2} \gamma^T B_{+1} (B_{-1}^T B_{-1} + c_4 I)^{-1} B_{+1}^T \gamma, \\ \text{s.t.} \quad & 0 \leq \gamma \leq c_2. \end{aligned} \quad (19)$$

Using these equations, we can determine Lagrangian multipliers which are further useful to obtain hyperplane parameters. In this way, hyperplane is constructed for each class using (6). A class +1 or -1 is assigned to new data instance depending upon its closeness to the two nonparallel hyperplanes. TWSVM assigns class label to an instance by using the following decision function:

$$f(x) = \arg \min_{i=+1,-1} \frac{|w_i \cdot x + b_i|}{\|w_i\|}, \quad (20)$$

where $|\cdot|$ is the absolute value. TWSVM has also been extended to the nonlinear cases where data instances are not separable by linear class boundaries. For this purpose, it uses kernel trick to transform the data instances into higher-dimensional feature space. Nonlinear TWSVM seeks the following two kernel surfaces instead of planes:

$$\begin{aligned} K(x^T, D^T) \mu_{+1} + b_{+1} &= 0, \\ K(x^T, D^T) \mu_{-1} + b_{-1} &= 0, \end{aligned} \quad (21)$$

where K is any arbitrary kernel function and $D^T = [A_{+1} \ A_{-1}]^T$. The primal QPPs of nonlinear TWSVM corresponding to kernel-generated surfaces (21) are given below:

$$\begin{aligned} \min_{\mu_{+1}, b_{+1}, \xi} \quad & \frac{1}{2} \|K(A_{+1}, D^T) \mu_{+1} + e_{+1} b_{+1}\|^2 + c_1 e_{-1}^T \xi, \\ \text{s.t.} \quad & -(K(A_{-1}, D^T) \mu_{+1} + e_{-1} b_{+1}) + \xi \geq e_{-1}, \\ & \xi \geq 0, \\ \min_{\mu_{-1}, b_{-1}, \eta} \quad & \frac{1}{2} \|K(A_{-1}, D^T) \mu_{-1} + e_{-1} b_{-1}\|^2 + c_2 e_{+1}^T \eta, \\ \text{s.t.} \quad & (K(A_{+1}, D^T) \mu_{-1} + e_{+1} b_{-1}) + \eta \geq e_{+1}, \\ & \eta \geq 0. \end{aligned} \quad (22)$$

Similar to (16) and (17), kernel-generated surface parameters can be determined as

$$\begin{aligned} \begin{bmatrix} \mu_{+1} \\ b_{+1} \end{bmatrix} &= -(G_{+1}^T G_{+1})^{-1} G_{-1}^T \alpha, \\ \begin{bmatrix} \mu_{-1} \\ b_{-1} \end{bmatrix} &= (G_{-1}^T G_{-1})^{-1} G_{+1}^T \gamma. \end{aligned} \quad (23)$$

Here, $G_{+1} = [K(A_{+1}, D^T) e_{+1}]$ and $G_{-1} = [K(A_{-1}, D^T) e_{-1}]$. Similar to the linear case, regularization terms $c_3 I$ and $c_4 I$ are added to (23) to avoid the ill-conditioned matrices. A new data instance is labeled with class +1 or -1 in a similar manner to the linear case.

4. Bag Dissimilarity Representation

In multiple instance learning case, a classifier works at the bag level rather than instance level and takes a bag as an input. Therefore, the objective of MIL is to develop a classifier which generates a decision function for the bag. In the proposed approach, each bag is represented by a vector of its dissimilarities to the other bags in the training set. The dissimilarity of a bag from all other bags represents a feature vector. If there are N bags in the training set and i th bag contains n_i number of instances, then i th bag can be represented as

$$v_i = [d(B_i, B_1), d(B_i, B_2), \dots, d(B_i, B_N)]^T. \quad (24)$$

Thus, each bag has a single feature vector or instance representation and the MIL problem can be considered as a regular supervised learning problem. The dissimilarity between two bags B_i and B_j is measured by using different ways which are classified into two main categories on the basis of bag representation. Consider the representation of a bag as a point set of the high-dimensional feature space, and then the dissimilarity between two bags can be measured using a set distance. The following distance metrics have been used to calculate the dissimilarity between bags.

(a) *Hausdorff Distance*. The Hausdorff distance is one of the most popular distance metrics used in object recognition in the field of computer vision. Two bags B_i and B_j are said to be close to each other if every instance of bag B_i is close to an instance in bag B_j . The dissimilarity between two bags B_i and B_j is defined as

$$d(B_i, B_j) = \max \{d_{\text{dir}}(B_i, B_j), d_{\text{dir}}(B_j, B_i)\}. \quad (25)$$

Here, $d_{\text{dir}}(B_i, B_j)$ represents the directed distance between two bags B_i and B_j . In detail, given two bags $B_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ and $B_j = \{x_{j1}, x_{j2}, \dots, x_{jn_j}\}$, the directed distance between B_i and B_j is calculated as

$$d_{\text{dir}}(B_i, B_j) = \max_k \min_l \|x_{ik} - x_{jl}\|_2. \quad (26)$$

Simply, $\|x_{ik} - x_{jl}\|_2$ measures the Euclidean distance between the instances x_{ik} and x_{jl} . The N -dimensional representation of i th bag, v_i , is formed as a vector of such dissimilarities between i th bag and all the other bags in the training set. The final Hausdorff distance between two bags is symmetrized by taking the maximum of the directed distance between them as d_{dir} is not symmetric. The dissimilarity between two bags

can also be defined by taking the minimum and average of squared Euclidean distance as follows:

$$d(B_i, B_j) = \min_k \min_l \|x_{ik} - x_{jl}\|_2, \quad (27)$$

$$d(B_i, B_j) = \frac{1}{n_i} \sum_{k=1}^{n_i} \min_l \|x_{ik} - x_{jl}\|_2. \quad (28)$$

Figure 4 shows the minimum Euclidean distances between instances of two bags. According to Figure 4(a), all the instances in bag B_i have the same closest instance in bag B_j while instances in bag B_j have two different closest instances in bag B_i . Due to which the minimum distance between instances of two bags is asymmetric $d_{\text{min}}(B_i, B_j) \neq d_{\text{min}}(B_j, B_i)$. This distance can be symmetrized by again taking the minimum of these distances; that is, $d_{\text{minmin}}(B_i, B_j) = d_{\text{minmin}}(B_j, B_i)$. For the case of average distance as given by (28), $d_{\text{averagemin}}(B_i, B_j) \neq d_{\text{averagemin}}(B_j, B_i)$.

(b) *City Block Distance*. The dissimilarity between two bags can be defined by using City Block distance metric as follows:

$$d(B_i, B_j) = \frac{1}{n_i} \sum_{k=1}^{n_i} \min_l |x_{ik} - x_{jl}|. \quad (29)$$

(c) *Chi-Squared Distance*. Chi-squared distance is the weighted Euclidean distance which measures the dissimilarity between two bags as follows:

$$d(B_i, B_j) = \frac{1}{n_i} \sum_{k=1}^{n_i} \min_l \frac{(x_{ik} - x_{jl})^2}{2 * (x_{ik} + x_{jl})}. \quad (30)$$

If a bag can be viewed as a probability distribution in the instance space, then the dissimilarity between two bags can be defined through distribution distance. It is not only difficult to determine the probability density function in a higher-dimensional feature space but also very computationally expensive to estimate the true distributions of instances. Therefore, the instance distributions are approximated and the distance is measured between the approximated distributions by using the following two distance metrics.

(a) *Earth Mover's Distance (EMD)*. Earth Mover's distance [71] measures the minimum amount of work to transform one probability distribution B_i into another probability distribution B_j . Consider that each instance has $1/n_i$ of the total probability mass in bag B_i of size n_i . The Earth Mover's distance between two bags is computed as

$$d_{\text{EMD}}(B_i, B_j) = \min_{f_{kl}} \sum_{k,l} f_{kl} d(x_{ik}, x_{jl}), \quad (31)$$

where $d(x_{ik}, x_{jl})$ is the Euclidean distance and f_{kl} is the flow between instances "k" and "l" associated with additional constraints: $f_{kl} > 0$, $\sum_k f_{kl} \leq 1/n_j$, $\sum_l f_{kl} \leq 1/n_i$, and $\sum_{kl} f_{kl} = 1$.

(b) *Mahalanobis Distance*. Each bag B_i is approximated by a single Gaussian distribution with mean μ_i and covariance

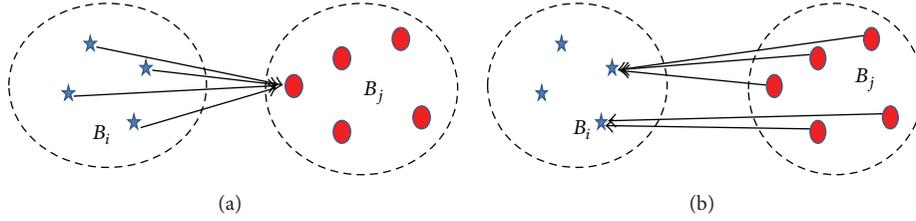


FIGURE 4: Minimum Euclidean distances between instances of two bags.

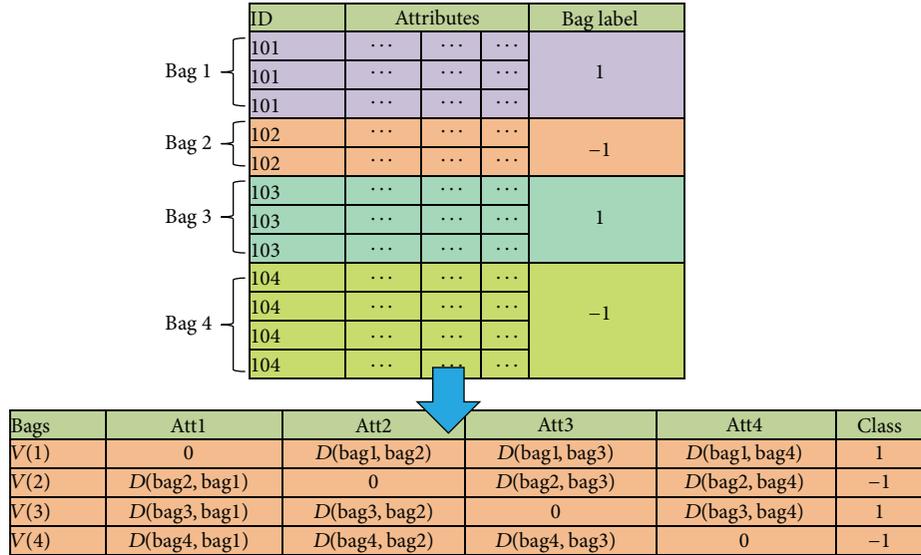


FIGURE 5: Conversion of multiple instance learning data into single instance learning using bag dissimilarity score.

matrix \sum_i parameters. Bag dissimilarity between two bags B_i and B_j through Mahalanobis distance is defined as

$$d_{\text{MAHA}}(B_i, B_j) = (\mu_i - \mu_j)^T \left(\frac{1}{2} \sum_i + \frac{1}{2} \sum_j \right)^{-1} (\mu_i - \mu_j). \quad (32)$$

In this way, we can calculate the dissimilarity score of a bag from the rest of the other bags. The new vector representation of each bag v_i acts as an input to the TWSVM classifier which now works at the bag-level $f(v_i)$. Now the problem has been converted into the single instance binary classification problem in which a bag has either +1 or -1 class label. Figure 5 depicts the example of multiple instance learning data having four bags. Each bag contains different number of instances such that bag 1 and bag 3 have three instances while bag 2 contains two instances and bag 4 consists of four instances. Class label is associated with each bag instead of individual instance. Traditional supervised learning approaches are not designed for such type of problems. Thus a bag-level MIL-TWSVM classifier is trained with this summarized data. During testing phase, the similar representation is obtained for the bag query and the proposed classifier takes a decision for a bag on the basis of minimum distance criteria.

5. Numerical Experiments

This section presents the experimental results of our proposed MIL-TWSVM classifier on ten benchmark MIL datasets. We have analyzed the performance of proposed MIL-TWSVM classifier with different dissimilarity metrics. The results of MIL-TWSVM have been compared with several existing MIL approaches such as Diverse Density (DD), Expectation-Maximization Diverse Density (EMDD), Multi-Instance Logistic Regression (MILR), Citation k -NN, and Multi-Instance Support Vector Machine (MISVM). All these classifiers have been implemented in MATLAB 2012a on Windows 7 operating system with Intel core i-7 processor with 12 GB RAM. This section has been divided into four subsections. The first subsection includes the description of benchmark MIL datasets used in this study. The second subsection analyzes the impact of parameters on the performance of proposed classifier. Experimental results are discussed and analyzed in subsections three and four, respectively.

5.1. Dataset Description. In this study, the experiment has been performed on ten MIL benchmark datasets: Musk 1, Musk 2, Mutagenesis-atoms, Winter Wren, Brown Creeper, Elephant, Fox, Tiger, eastWest, and westEast datasets. These datasets are available online at <http://www.mipproblems.org/>.

TABLE 1: Multiple instance learning datasets.

Datasets	Number of bags			Number of attributes	Number of instances	Average bag size
	+bags	-bags	Total			
Musk 1	47	45	92	167	476	5.17
Musk 2	39	63	102	167	6598	64.69
Mutagenesis-atoms	125	63	188	11	1618	8.61
Winter wren	109	439	548	38	10232	18.67
Brown creeper	197	351	548	38	10232	18.67
Elephant	100	100	200	230	1391	6.96
Fox	100	100	200	230	1320	6.6
Tiger	100	100	200	230	1220	6.2
eastWest	10	10	20	25	213	10.65
westEast	10	10	20	25	213	10.65

The detailed description of these datasets is shown in Table 1. These datasets are widely adopted for the performance evaluation of new MIL approaches. These datasets represent four different categories of MIL example.

Musk 1 and Musk 2 are two standard drug activity prediction benchmark datasets in which a bag is represented by one molecule and different conformations or shapes of these molecules are the instances of a bag. In these drug activity prediction datasets, a bag is assigned with the class label “musk” or “non-musk” by human expert. Musk 1 dataset contains 92 bags while Musk 2 dataset contains 102 bags. Musk 2 dataset contains more number of instances or molecule conformations as compared to the Musk 1 dataset. The objective of MIL is to predict whether a new molecule is “musk” or “non-musk.” Another dataset that belongs to the category of drug activity prediction is Mutagenesis-atoms. This dataset contains 125 positive and 63 negative bags. Brown Creeper and Winter Wren are two audio MIL datasets which contain the audio of bird songs of different species. A bag is represented by an audio fragment. A bag is labeled as positive if particular species is heard in the audio fragment for that category. Since the birds of the same species have similar songs or audio fragment, therefore different bird species have different concepts. It is also possible that some species are heard together more often. In this case, the audio fragments, which are not heard or are negative for one bird species, could be useful to determine whether an audio fragment contains that species or not. Content Base Image Retrieval (CBIR) is another one of the most recognized applications of MIL in which the objective is to determine whether the given image is of interest to user or not. A set of regions or image patches represent an image. An image corresponds to a bag and image regions represent the instances in each bag. The class label of individual instances is unknown. In this study, we have used three image datasets: Elephant, Fox, and Tiger. eastWest and westEast datasets belong to an ILP problem and have been collected from eastWest challenge. The objective of this challenge is to predict whether a train is eastbound or westbound. In eastWest or westEast datasets, a bag represents a train which contains various cars (instances) of different shapes and sizes. Each car having different loads represents

its instance-level attributes. eastWest data challenge has two MI datasets: eastWest and westEast as it is not clear whether an eastbound or westbound train can be considered as positive label example. In eastWest dataset, eastbound trains are regarded as an example of positive class label. Similarly, westbound train is considered as positive example in the westEast dataset.

5.2. Parameters Selection. This study has used Gaussian Kernel function $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ for nonlinear case. MIL-TWSVM classifier has four penalty parameters: $c_1, c_2, c_3,$ and c_4 and an additional kernel parameter sigma (σ). The predictive performance of the classifier gets affected by the choices of these parameters. This study has used Grid Search approach which is one of the widely used approaches for the optimal parameters selection [27, 65, 72–74]. The penalty and kernel parameters are selected from the following range: $c_1, c_2, c_3, c_4 \in \{10^{-6}, \dots, 10^3\}$ and $\sigma \in \{2^{-3}, \dots, 2^6\}$. The experiment has been conducted using 10-fold cross-validation approach. It trains the proposed classifier with each pair (penalty and σ) in the Cartesian product of these two sets and evaluates their performance by internal cross-validation on the training set, in which case multiple MIL-TWSVMs are trained per pair. Finally, it outputs the settings that achieved the highest score in the validation procedure. We have analyzed the influence of these parameters on the performance of MIL-TWSVM on three datasets: Tiger, Fox, and Mutagenesis-atoms as shown in Figures 6, 7, and 8. For linear case, we set $c_1 = c_2$ and $c_3 = c_4$ to reduce the computational complexity and analyze their influence on the predictive performance of linear classifier. However, for nonlinear case, consider $c_1 = c_2 = c_3 = c_4$ to reduce the computational complexity and analyze the influence of these parameters and sigma on the predictive performance of nonlinear MIL-TWSVM classifier. For tiger dataset, the impact of parameters has been analyzed using Max-Hausdorff dissimilarity measure as shown in Figure 6. From the figure, it is observed that the proposed linear MIL-TWSVM classifier has obtained better performance with low value of c_1 and high value

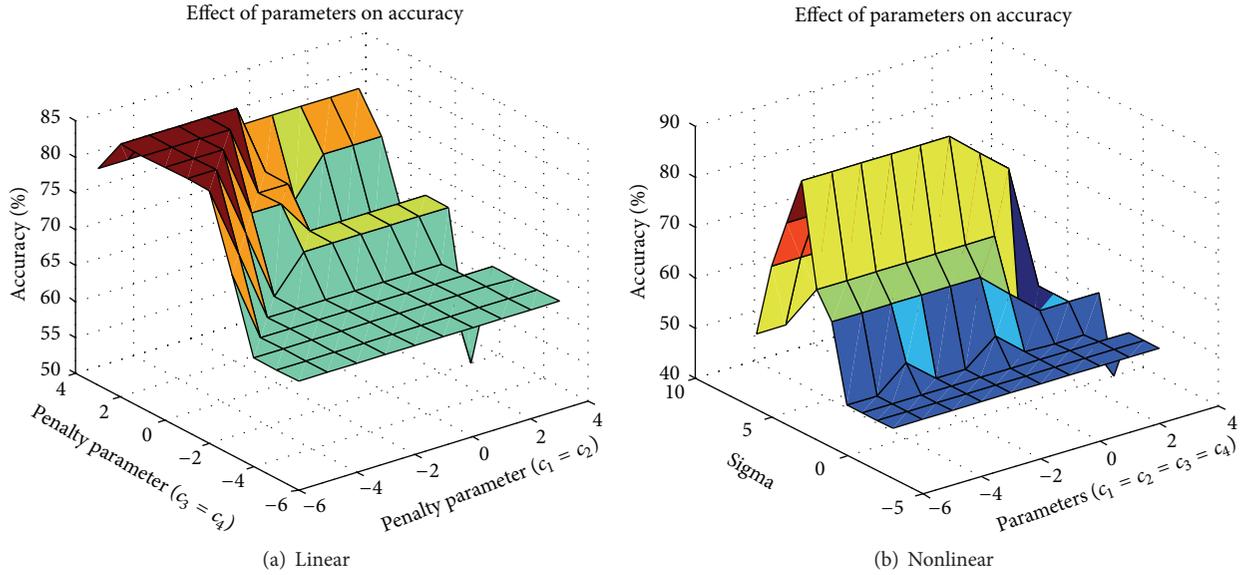


FIGURE 6: Influence of parameters on the predictive accuracy of MIL-TWSVM (Max-Hausdorff) on Tiger dataset.

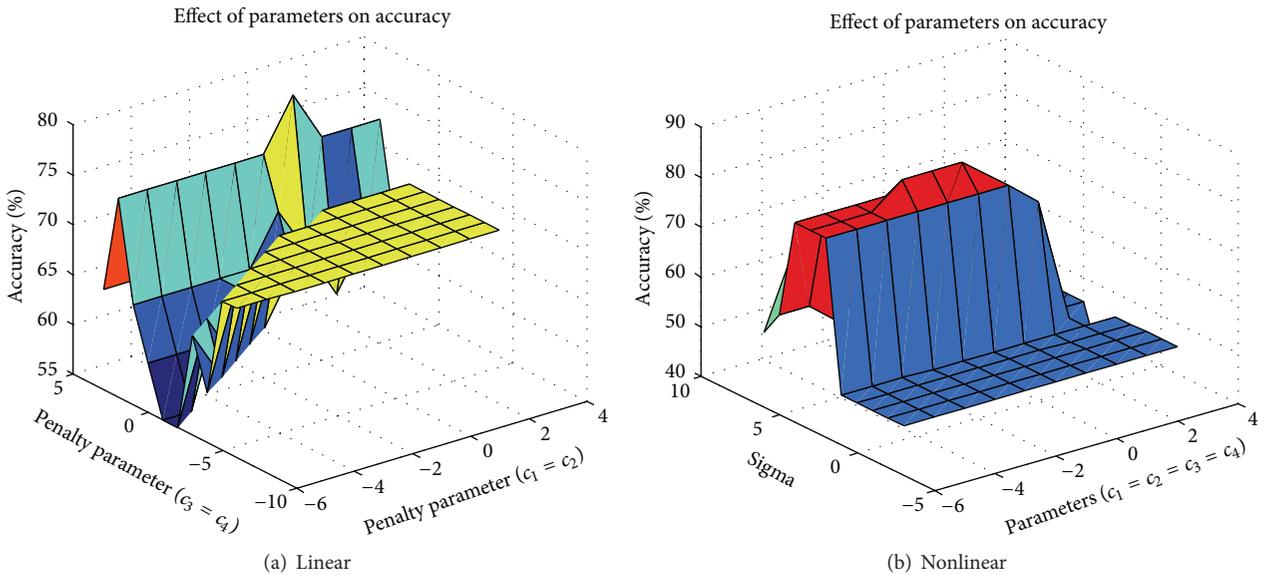


FIGURE 7: Influence of parameters on the predictive accuracy of MIL-TWSVM (Minimum Hausdorff) on Fox dataset.

of c_3 parameters ($c_1 = 10^{-4}$, $c_3 = 10^2$). The performance of MIL-TWSVM suddenly degrades for low value of c_3 parameter. For nonlinear cases, MIL-TWSVM obtains better performance with high value of sigma (σ) and low value of penalty parameter c_1 ($\sigma = 2^3$, $c_1 = 10^{-5}$). Max-Hausdorff based MIL-TWSVM classifier shows better performance for different combinations of penalty and sigma parameters on other datasets.

The impact of these parameters on Fox and Mutagenesis-atoms datasets has been analyzed using Min-Hausdorff and EMD dissimilarity measures, respectively. On fox dataset,

Min-Hausdorff based linear MIL-TWSVM has shown better performance for low value of c_1 and high value of c_3 ($c_1 = 10^{-3}$, $c_3 = 10^2$). Nonlinear MIL-TWSVM classifier has achieved better predictive accuracy with high value of sigma and low value of penalty parameter ($\sigma = 2^3$, $c_1 = 10^{-2}$) on fox dataset as shown in Figure 7. For Mutagenesis-atoms, EMD based linear MIL-TWSVM has obtained highest accuracy for low value of c_1 and c_3 parameters ($c_1 = 10^{-3}$, $c_3 = 10^{-3}$). Nonlinear MIL-TWSVM has gained highest accuracy on Mutagenesis-atoms dataset for low value of sigma and penalty parameter ($\sigma = 2^0$, $c_1 = 10^{-4}$) as shown in Figure 8. For

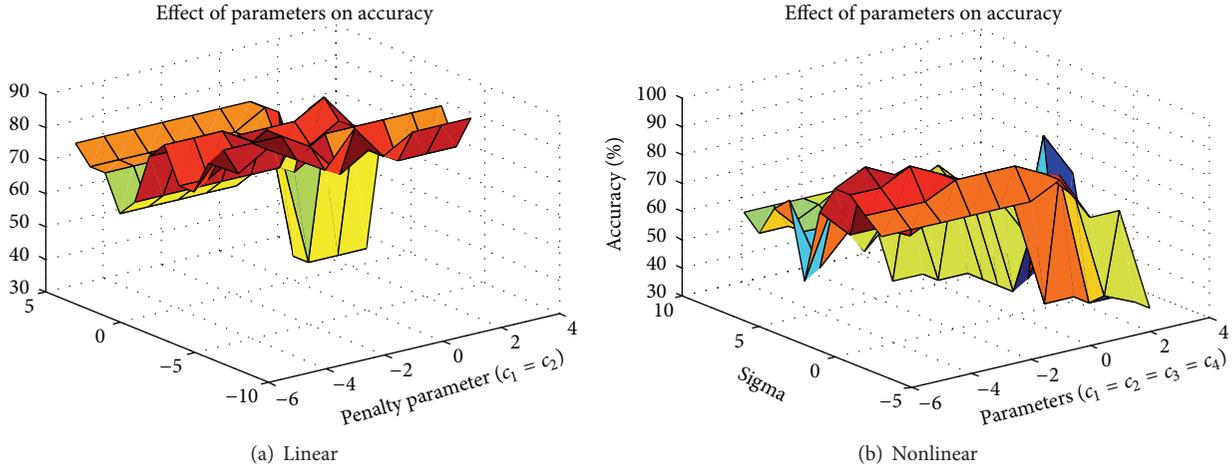


FIGURE 8: Influence of parameters on the predictive accuracy of MIL-TWSVM (EMD) on Mutagenesis-atoms dataset.

every combination of these parameters (penalty and sigma) and dissimilarity measures, the proposed MIL-TWSVM classifier behaves differently on different datasets. Therefore, the appropriate selection of these parameters is essential to obtain better performance of MIL-TWSVM classifier.

5.3. Results and Discussion. The predictive accuracy of the proposed MIL-TWSVM classifier with different dissimilarity measures on ten benchmark datasets is shown in Tables 2 and 3 for linear and nonlinear cases, respectively.

The result includes the average and standard deviation of classification accuracies of the 10-fold cross-validation. Bold values indicate better predictive accuracy of the classifier. Min-Hausdorff dissimilarity score based linear MIL-TWSVM gains highest accuracy on Musk 1, Winter Wren, Elephant, Fox, Tiger, and westEast datasets. Max-Hausdorff based linear MIL-TWSVM classifier obtains highest accuracy on Musk 2 and Brown Creeper datasets. EMD based linear MIL-TWSVM achieves highest accuracy on Mutagenesis-atoms dataset.

Other bag dissimilarity measurements based MIL-TWSVM classifier shows poor performance on all type of datasets. Similarly, for nonlinear case, Min-Hausdorff and Max-Hausdorff dissimilarity score based MIL-TWSVM classifier has shown better performance on Winter Wren, Brown Creeper, Elephant, Tiger, Fox, eastWest, and westEast datasets. EMD based nonlinear MIL-TWSVM classifier achieves highest predictive accuracy on Mutagenesis-atoms dataset. Therefore, we can conclude that the MIL-TWSVM has shown better performance with Min-Hausdorff and Max-Hausdorff dissimilarity scores. Further, we have compared the performance of Min-Hausdorff based MIL-TWSVM classifier with the existing MIL approaches-Expectation-Maximization Diverse Density (EMDD), Diverse Density (DD), Multi-Instance Logistic Regression (MILR), Citation k -NN, and Multi-Instance Support Vector Machine (MISVM) as shown in Table 4.

From Table 4, it is observed that the proposed MIL-TWSVM classifier has achieved highest predictive accuracy

on all ten benchmark datasets and thus performs better than the other existing MIL approaches.

5.4. Statistical Comparison. Friedman test statistic [32, 33] assigns rank to each classifier according to their performance on each dataset independently. For example, the first rank is given to the best performing classifier; second best performing classifier gets second rank. Average rank is given to the classifiers in case they have shown the same performance. Let r_i^j be the rank of j th classifier on i th dataset. Friedman test statistic is calculated as

$$\chi_F^2 = \frac{12D}{K(K+1)} \left[\sum_{j=1}^K AR_j^2 - \frac{K(K+1)^2}{4} \right], \quad (33)$$

$$\text{where } AR_j = \frac{1}{D} \sum_{i=1}^D r_i^j.$$

Here, D represents the number of datasets used in this study for comparison purpose; K denotes the number of classifiers and AR_j is the average rank of j th classifier. Friedman test statistic follows chi-squared distribution with $K - 1$ degrees of freedom. The null hypothesis which states that there is no difference between classifiers can be rejected or accepted according to the value of Friedman test statistic. If the value of Friedman test statistic is large as compared to the critical value corresponds to $K - 1$ degrees of freedom then we can accept or reject the null hypothesis. The Nemenyi post hoc test [75] reports the significant differences between individual classifiers. According to this test, two classifiers are significantly different if their average rank differs by at least the critical difference (CD) which is obtained as

$$CD = q_\alpha \sqrt{\frac{K(K+1)}{6D}}, \quad (34)$$

where q_α is calculated on the basis of studentized range statistic. The results of Friedman test statistic are plotted by using modified Demšar significance diagram [76]. We have

TABLE 2: Performance comparison of linear MIL-TWSVM with different dissimilarity score.

Datasets	Max-Hausdorff		Min-Hausdorff		Average-Hausdorff		City block		Chi-squared		Mahalanobis		EMD	
	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	c_1, c_3 Acc \pm std (%)	
Musk 1	$10^0, 10^0$ 95.56 \pm 2.34	$10^{-2}, 10^{-3}$ 96.67 \pm 5.09	$10^1, 10^1$ 7778 \pm 0.0	$10^0, 10^3$ 61.11 \pm 7.85	$10^{-2}, 10^{-2}$ 64.48 \pm 4.84	$10^{-4}, 10^{-4}$ 70.03 \pm 4.58	$10^{-2}, 10^{-2}$ 75.58 \pm 8.31							
Musk 2	$10^{-3}, 10^{-4}$ 93.02 \pm 4.56	$10^{-2}, 10^{-1}$ 92.85 \pm 4.77	$10^{-3}, 10^{-1}$ 79.84 \pm 2.26	$10^{-4}, 10^{-1}$ 62.46 \pm 4.67	$10^{-2}, 10^{-3}$ 66.67 \pm 4.18	$10^{-2}, 10^{-1}$ 72.34 \pm 4.28	$10^{-3}, 10^{-3}$ 74.47 \pm 7.23							
Mutagenesis-atoms	$10^0, 10^{-1}$ 86.84 \pm 1.52	$10^{-1}, 10^{-3}$ 87.04 \pm 2.63	$10^{-2}, 10^{-1}$ 68.42 \pm 0.0	$10^{-1}, 10^{-1}$ 68.42 \pm 0.0	$10^{-2}, 10^{-1}$ 68.42 \pm 0.0	$10^{-2}, 10^0$ 68.42 \pm 0.0	$10^{-3}, 10^{-3}$ 87.36 \pm 2.57							
Winter wren	$10^{-4}, 10^{-2}$ 96.84 \pm 2.86	$10^{-1}, 10^{-3}$ 97.47 \pm 2.6	$10^{-4}, 10^{-2}$ 89.22 \pm 5.88	$10^{-3}, 10^0$ 79.67 \pm 6.56	$10^{-3}, 10^{-3}$ 84.26 \pm 4.02	$10^{-4}, 10^{-3}$ 76.89 \pm 5.35	$10^{-3}, 10^{-1}$ 94.71 \pm 4.53							
Brown creeper	$10^{-5}, 10^{-3}$ 96.01 \pm 3.66	$10^{-3}, 10^{-4}$ 95.88 \pm 3.24	$10^{-4}, 10^{-1}$ 90.89 \pm 4.22	$10^{-1}, 10^{-3}$ 76.38 \pm 3.63	$10^{-2}, 10^{-4}$ 85.43 \pm 5.83	$10^{-2}, 10^{-3}$ 82.67 \pm 5.56	$10^{-4}, 10^0$ 93.45 \pm 5.15							
Elephant	$10^1, 10^{-3}$ 83.88 \pm 3.01	$10^{-5}, 10^{-1}$ 85.56 \pm 5.49	$10^0, 10^0$ 55 \pm 4.08	$10^0, 10^0$ 50 \pm 0.0	$10^{-1}, 10^{-1}$ 64.16 \pm 13.07	$10^0, 10^0$ 50 \pm 0.0	$10^{-1}, 10^{-1}$ 74 \pm 5.83							
Fox	$10^{-4}, 10^{-5}$ 58.33 \pm 1.74	$10^{-3}, 10^2$ 67.22 \pm 6.7	$10^{-4}, 10^1$ 54.44 \pm 3.5	$10^{-1}, 10^{-1}$ 50 \pm 0.0	$10^{-2}, 10^{-3}$ 55 \pm 0.0	$10^{-1}, 10^0$ 51.5 \pm 1.97	$10^{-7}, 10^2$ 60 \pm 6.32							
Tiger	$10^{-4}, 10^2$ 83.5 \pm 2.45	$10^{-6}, 10^{-2}$ 86.25 \pm 5.89	$10^{-5}, 10^{-2}$ 55 \pm 7.28	$10^0, 10^{-2}$ 50 \pm 0.0	$10^{-3}, 10^{-2}$ 51.5 \pm 4.27	$10^{-1}, 10^{-2}$ 50 \pm 0.0	$10^{-6}, 10^{-5}$ 74.5 \pm 6.5							
eastWest	$10^{-5}, 10^{-1}$ 60 \pm 20	$10^{-3}, 10^0$ 80 \pm 24.49	$10^{-1}, 10^0$ 80 \pm 24.49	$10^{-2}, 10^{-2}$ 50 \pm 0.0	$10^{-1}, 10^0$ 50 \pm 0.0	$10^{-4}, 10^0$ 65 \pm 22.9	$10^{-3}, 10^{-1}$ 80 \pm 24.49							
westEast	$10^0, 10^{-2}$ 75 \pm 25	$10^0, 10^0$ 90 \pm 20	$10^0, 10^{-3}$ 85 \pm 22.9	$10^0, 10^0$ 50 \pm 0.0	$10^{-2}, 10^0$ 50 \pm 0.0	$10^{-1}, 10^{-1}$ 75 \pm 25	$10^{-2}, 10^{-1}$ 75 \pm 25							

TABLE 3: Performance comparison of nonlinear MIL-TWSVM with different dissimilarity score.

Datasets	Max-Hausdorff		Min-Hausdorff		Euclidean (average)		City Block		Chi-squared		Mahalanobis		EMD	
	c_1, σ Acc \pm std (%)													
Musk 1	$10^{-2}, 2^3$ 55.57 \pm 0.0	$10^0, 2^1$ 55.57 \pm 0.0	$10^{-2}, 2^2$ 55.57 \pm 0.0	$10^3, 2^2$ 55.57 \pm 0.0	$10^{-3}, 2^1$ 78.89 \pm 3.68	$10^{-4}, 2^3$ 71.14 \pm 5.22	$10^{-1}, 2^0$ 55.57 \pm 0.0	$10^{-3}, 2^1$ 78.89 \pm 3.68	$10^{-3}, 2^1$ 78.89 \pm 3.68	$10^{-4}, 2^3$ 71.14 \pm 5.22	$10^{-1}, 2^0$ 55.57 \pm 0.0	$10^{-2}, 2^0$ 55.57 \pm 0.0	$10^{-1}, 2^0$ 55.57 \pm 0.0	
Musk 2	$10^{-2}, 2^3$ 58.25 \pm 3.45	$10^{-3}, 2^5$ 55.57 \pm 0.0	$10^{-3}, 2^3$ 57.03 \pm 2.4	$10^{-3}, 2^2$ 55.57 \pm 0.0	$10^0, 2^1$ 77.78 \pm 4.04	$10^{-2}, 2^4$ 77.70 \pm 3.26	$10^{-1}, 2^4$ 58.25 \pm 3.45	$10^{-3}, 2^5$ 55.57 \pm 0.0	$10^{-3}, 2^3$ 57.03 \pm 2.4	$10^{-1}, 2^5$ 77.78 \pm 4.04	$10^{-2}, 2^3$ 77.70 \pm 3.26	$10^{-3}, 2^4$ 55.57 \pm 0.0	$10^{-3}, 2^0$ 55.57 \pm 0.0	
Mutagenesis-atoms	$10^{-2}, 2^1$ 85.52 \pm 1.28	$10^{-1}, 2^4$ 81.58 \pm 5.88	$10^{-3}, 2^3$ 68.42 \pm 0.0	$10^{-3}, 2^3$ 68.42 \pm 0.0	$10^{-1}, 2^5$ 68.42 \pm 0.0	$10^{-4}, 2^3$ 68.42 \pm 0.0	$10^{-1}, 2^4$ 85.52 \pm 1.28	$10^{-4}, 2^4$ 81.58 \pm 5.88	$10^{-4}, 2^3$ 68.42 \pm 0.0	$10^{-1}, 2^5$ 68.42 \pm 0.0	$10^{-3}, 2^3$ 68.42 \pm 0.0	$10^{-3}, 2^4$ 55.57 \pm 0.0	$10^{-3}, 2^0$ 55.57 \pm 0.0	
Winter wren	$10^{-3}, 2^1$ 97.56 \pm 4.67	$10^{-5}, 2^3$ 98.02 \pm 2.2	$10^{-3}, 2^0$ 91.10 \pm 3.95	$10^{-5}, 2^2$ 82.86 \pm 5.69	$10^{-3}, 2^0$ 91.10 \pm 3.95	$10^{-3}, 2^0$ 91.10 \pm 3.95	$10^{-3}, 2^1$ 97.56 \pm 4.67	$10^{-5}, 2^3$ 98.02 \pm 2.2	$10^{-3}, 2^0$ 91.10 \pm 3.95	$10^{-3}, 2^0$ 91.10 \pm 3.95	$10^{-3}, 2^1$ 82.86 \pm 5.69	$10^{-3}, 2^4$ 55.57 \pm 0.0	$10^{-4}, 2^5$ 95.49 \pm 4.66	
Brown creeper	$10^{-5}, 2^3$ 97.27 \pm 3.84	$10^{-7}, 2^5$ 97.89 \pm 3.71	$10^{-1}, 2^0$ 92.30 \pm 4.08	$10^{-5}, 2^2$ 85.8 \pm 4.53	$10^{-1}, 2^0$ 92.30 \pm 4.08	$10^{-1}, 2^0$ 92.30 \pm 4.08	$10^{-7}, 2^5$ 97.27 \pm 3.84	$10^{-7}, 2^5$ 97.89 \pm 3.71	$10^{-1}, 2^0$ 92.30 \pm 4.08	$10^{-1}, 2^0$ 92.30 \pm 4.08	$10^{-5}, 2^2$ 85.8 \pm 4.53	$10^{-3}, 2^3$ 90.03 \pm 3.35	$10^{-4}, 2^5$ 96.72 \pm 3.24	
Elephant	$10^{-5}, 2^3$ 83.62 \pm 3.37	$10^{-2}, 2^3$ 83.33 \pm 5.08	$10^{-5}, 2^4$ 50 \pm 0.0	$10^{-5}, 2^1$ 50 \pm 0.0	$10^{-5}, 2^4$ 50 \pm 0.0	$10^{-5}, 2^4$ 50 \pm 0.0	$10^{-5}, 2^3$ 83.62 \pm 3.37	$10^{-2}, 2^3$ 83.33 \pm 5.08	$10^{-5}, 2^4$ 50 \pm 0.0	$10^{-5}, 2^1$ 50 \pm 0.0	$10^{-5}, 2^1$ 50 \pm 0.0	$10^{-3}, 2^1$ 50 \pm 0.0	$10^{-3}, 2^1$ 50 \pm 0.0	
Fox	$10^{-7}, 2^4$ 63.89 \pm 1.01	$10^{-5}, 2^3$ 75 \pm 4.49	$10^{-5}, 2^1$ 61.11 \pm 6.98	$10^1, 2^1$ 57.78 \pm 7.85	$10^{-5}, 2^3$ 75 \pm 4.49	$10^{-5}, 2^1$ 61.11 \pm 6.98	$10^{-7}, 2^4$ 63.89 \pm 1.01	$10^{-5}, 2^3$ 75 \pm 4.49	$10^{-5}, 2^1$ 57.78 \pm 7.85	$10^{-2}, 2^2$ 55 \pm 0.0	$10^1, 2^1$ 57.78 \pm 7.85	$10^0, 2^1$ 54 \pm 4.76	$10^0, 2^1$ 50 \pm 0.0	
Tiger	$10^{-5}, 2^3$ 85.63 \pm 1.22	$10^{-3}, 2^1$ 83.13 \pm 6.02	$10^{-1}, 2^2$ 50 \pm 0.0	$10^{-2}, 2^5$ 50 \pm 0.0	$10^{-3}, 2^1$ 83.13 \pm 6.02	$10^{-1}, 2^2$ 50 \pm 0.0	$10^{-5}, 2^3$ 85.63 \pm 1.22	$10^{-3}, 2^1$ 83.13 \pm 6.02	$10^{-2}, 2^5$ 50 \pm 0.0	$10^{-2}, 2^2$ 50.5 \pm 2.06	$10^0, 2^1$ 50 \pm 0.0	$10^{-3}, 2^5$ 50 \pm 0.0	$10^0, 2^1$ 50 \pm 0.0	
eastWest	$10^{-3}, 2^4$ 80 \pm 24.49	$10^0, 2^1$ 80 \pm 24.49	$10^{-1}, 2^2$ 80 \pm 24.49	$10^{-2}, 2^4$ 80 \pm 24.49	$10^0, 2^1$ 80 \pm 24.49	$10^{-1}, 2^2$ 80 \pm 24.49	$10^{-3}, 2^4$ 80 \pm 24.49	$10^0, 2^1$ 80 \pm 24.49	$10^{-2}, 2^4$ 80 \pm 24.49	$10^{-2}, 2^2$ 50 \pm 0.0	$10^0, 2^1$ 50 \pm 0.0	$10^{-3}, 2^5$ 80 \pm 24.49	$10^0, 2^1$ 80 \pm 24.49	
westEast	$10^0, 2^5$ 85 \pm 22.9	$10^0, 2^3$ 85 \pm 22.9	$10^0, 2^3$ 80 \pm 24.49	$10^1, 2^4$ 80 \pm 24.49	$10^0, 2^3$ 85 \pm 22.9	$10^0, 2^3$ 80 \pm 24.49	$10^0, 2^5$ 85 \pm 22.9	$10^0, 2^3$ 85 \pm 22.9	$10^1, 2^4$ 80 \pm 24.49	$10^1, 2^4$ 80 \pm 24.49	$10^0, 2^3$ 80 \pm 24.49	$10^0, 2^5$ 80 \pm 24.49	$10^0, 2^1$ 70 \pm 24.49	

TABLE 4: Performance comparison of MIL approaches.

Datasets	EMDD		DD		MILR		Citation k -NN		MIL Boost		BP-MLP		MILES		MISVM		MIL-TWSVM		
	Acc \pm std (%)	Rank																	
Musk 1	83.61 \pm 13.16	8	84.65 \pm 11.83	6	73.51 \pm 13.43	9	90.37 \pm 10.40	3	84.84 \pm 11.55	5	83.7 \pm 9.2	7	92.28 \pm 10.45	2	89.21 \pm 9.84	4	96.67 \pm 5.09	1	
Musk 2	85.57 \pm 10.28	3	80.39 \pm 13.44	7	78.51 \pm 13.06	9	84.61 \pm 11.54	4	79.50 \pm 12.39	8	80.4 \pm 10.34	6	89.66 \pm 12.36	2	83.92 \pm 10.36	5	92.85 \pm 4.77	1	
Mutagenesis-atoms	68.89 \pm 12.29	8	72.89 \pm 9.34	7	74.00 \pm 10.37	5	73.35 \pm 10.13	6	79.26 \pm 9.24	2	78.88 \pm 8.85	3	75.56 \pm 8.88	4	66.54 \pm 1.85	9	87.04 \pm 2.63	1	
Winter wren	96.85 \pm 3.02	3	85.56 \pm 4.16	9	89.45 \pm 2.3	8	94.8 \pm 4.22	6	92.34 \pm 5.16	7	96.67 \pm 7.53	4	96.35 \pm 6.02	5	97.16 \pm 2.85	2	98.02 \pm 2.2	1	
Brown creeper	94.5 \pm 2.26	4	93.78 \pm 3.88	6	93.6 \pm 4.24	7	85.77 \pm 4.35	9	94.02 \pm 5.75	5	95.36 \pm 6.14	3	95.8 \pm 5.84	2	92.80 \pm 3.16	8	97.89 \pm 3.71	1	
Elephant	75.25 \pm 10.57	8	81.45 \pm 10.23	4	78.90 \pm 9.28	7	50 \pm 0.0	9	84.70 \pm 8.46	2	83.72 \pm 8.82	3	80.36 \pm 8.26	2	79.45 \pm 9.32	6	85.56 \pm 5.49	1	
Fox	59.65 \pm 9.16	5	59.40 \pm 9.88	6	57.35 \pm 10.81	7	50 \pm 0.0	8	66.90 \pm 9.58	3	62.05 \pm 7.43	4	68.87 \pm 9.04	5	49.35 \pm 5.71	9	75 \pm 4.49	1	
Tiger	71.35 \pm 9.45	8	72.20 \pm 8.80	7	75.30 \pm 8.98	6	50 \pm 0.0	9	87.62 \pm 8.07	1	79.45 \pm 9.12	5	84.22 \pm 7.65	3	80.70 \pm 7.42	4	86.25 \pm 5.89	2	
eastWest	64.00 \pm 28.50	4	64.50 \pm 32.79	3	67.00 \pm 36.39	2	45 \pm 24.10	9	55.00 \pm 32.18	7.5	60.50 \pm 21.67	5.5	55.00 \pm 32.18	7.5	60.50 \pm 21.67	5.5	80 \pm 24.49	1	
westEast	38.00 \pm 27.63	6.5	36.00 \pm 29.37	8	35.50 \pm 34.30	9	50.5 \pm 27.06	4.5	56.50 \pm 33.83	2	55.00 \pm 32.18	3	50.5 \pm 27.06	4.5	38.00 \pm 25.74	6.5	90 \pm 20.00	1	
Average rank	5.75	6.3	6.9	6.9	6.75	6.75	4.25	4.35	3.7	5.9	3.7	4.35	3.7	5.9	3.7	4.35	3.7	1.1	1.1

Friedman test statistic = 36.83.

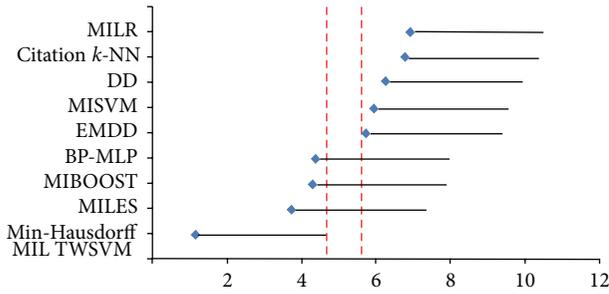


FIGURE 9: Average rank comparison of MIL approaches.

calculated the average rank of each MIL approach on the basis of its performance on each dataset (see Table 4). Then, the Friedman test statistic is calculated according to (31). From Table 4, it is observed that the Min-Hausdorff dissimilarity based MIL-TWSVM classifier achieves highest average rank among all MIL approaches. Maximum Hausdorff based MIL-TWSVM classifier gets second highest average rank. Consider $\alpha = 0.05$; then the critical value for 8-degree of freedom from chi-squared table is 15.507. The obtained Friedman test statistic value is 36.83 which is very higher than the critical value of 8-degree of freedom. Hence, we reject the null hypothesis which states that there is no difference between the classifiers. Critical value $q_{0.05}$ for nine classifiers is 3.102. Critical difference for $\alpha = 0.05$ is determined using (34) as follows:

$$CD = 3.102 \sqrt{\frac{9 \times 10}{6 \times 10}} = 3.799. \quad (35)$$

Figure 9 depicts the Demšar significance diagram in which MIL approaches are arranged in ascending order on the y -axis as per their average rank and their corresponding ranks are mentioned on the x -axis.

Critical difference value has been added to the average rank of each MIL approach in order to analyze whether the proposed approach is significantly better than the other MIL approaches. Two vertical lines in red color depict the difference of the end of the best performing MIL approach's tail and the start of the next significantly different MIL approach. From the figure, it is clear that the other existing MIL approaches such as EMDD, DD, MILR, Citation k -NN, and MISVM perform significantly worse than the best performing approach which is Min-Hausdorff based MIL-TWSVM. Thus, we can conclude that the proposed MIL-TWSVM is a suitable choice in the multiple instance learning problem domains.

6. Conclusion

This study has focused on multiple instance learning in which a classifier learns from a set of feature vectors (bag) instead of single feature vector and has proposed an MIL approach based on TWSVM, termed as MIL-TWSVM. Each bag is denoted by a vector of its dissimilarities to the other bags in the training set and the proposed classifier has been trained with this summarized information. Initially, the performance

of proposed MIL-TWSVM classifier has been compared with different dissimilarity scores on ten benchmark MIL datasets. We have also compared the performance of MIL-TWSVM with six existing MIL approaches. Experimental results demonstrate that the proposed approach has achieved highest predictive accuracy as compared to the other existing MIL approaches on all ten datasets. This further supports the suitability of MIL-TWSVM in multiple instance learning scenarios. The findings of experimental results are also supported by the statistical analysis performed by using Friedman test. The test shows that the MIL-TWSVM is significantly better than the EMDD, DD, MILR, Citation k -NN, and MISVM. In the future, we are interested in extending MIL-TWSVM to multi-instance multilabel scenario.

Competing Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

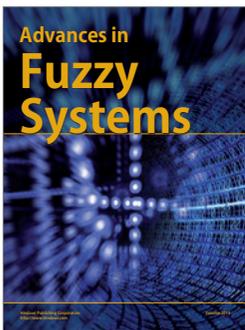
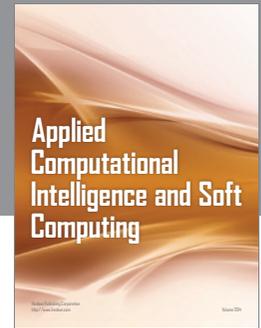
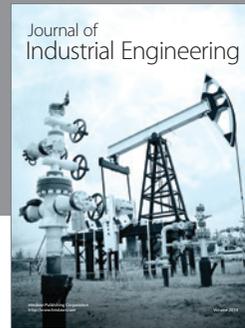
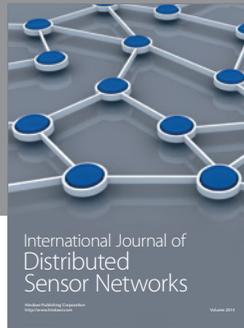
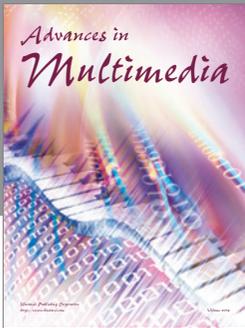
References

- [1] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31-71, 1997.
- [2] O. Maron and T. Lozano-Pérez, *A Framework for Multiple-Instance Learning*. *Advances in Neural Information Processing Systems (NIPS)*, vol. 10, MIT Press, 1998.
- [3] Z. H. Zhou, "Multi-instance learning: a survey," Tech. Rep., Department of Computer Science and Technology, Nanjing University, 2004.
- [4] J. Foulds and E. Frank, "A review of multi-instance learning assumptions," *The Knowledge Engineering Review*, vol. 25, no. 1, pp. 1-25, 2010.
- [5] J. Amores, "Multiple instance classification: review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81-105, 2013.
- [6] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," *Advances in Neural Information Processing Systems*, vol. 15, MIT Press, no. Cambridge, Mass, USA, pp. 561-568, 2003.
- [7] Y. Chen, J. Bi, and J. Z. Wang, "MILES: multiple-instance learning via embedded instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1931-1947, 2006.
- [8] I. Gondra and T. Xu, "A multiple instance learning based framework for semantic image segmentation," *Multimedia Tools and Applications*, vol. 48, no. 2, pp. 339-365, 2010.
- [9] Q. Zhang, S. A. Goldman, W. Yu, and J. E. Fritts, "Content-based image retrieval using multiple-instance learning," in *Proceedings of the International Conference on Machine Learning (ICML '02)*, vol. 2, pp. 682-689, 2002.
- [10] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-I.I.D. samples," in *Proceedings of the 26th International Conference on Machine Learning (ICML '09)*, pp. 1249-1256, Montreal, Canada, June 2009.
- [11] H. Wang, F. Nie, and H. Huang, "Learning instance specific distance for multi-instance classification," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, pp. 507-512, San Francisco, Calif, USA, August 2011.

- [12] P. Viola, J. C. Platt, and C. Zhang, "Multiple Instance boosting for object detection," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS '05)*, vol. 18, pp. 1417–1424, December 2005.
- [13] B. Babenko, N. Verma, P. Dollár, and S. J. Belongie, "Multiple instance learning with manifold bags," in *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*, pp. 81–88, 2011.
- [14] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 288–303, 2010.
- [15] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 983–990, IEEE, Miami Beach, Fla, USA, June 2009.
- [16] C. Leistner, A. Saffari, and H. Bischof, "Miforests: multiple-instance learning with randomized trees," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, pp. 29–42, Springer, Berlin, Germany, 2010.
- [17] Y. Xie, Y. Qu, C. Li, and W. Zhang, "Online multiple instance gradient feature selection for robust visual tracking," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1075–1082, 2012.
- [18] B. Zeisl, C. Leistner, A. Saffari, and H. Bischof, "On-line semi-supervised multiple-instance boosting," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, p. 1879, IEEE, San Francisco, Calif, USA, June 2010.
- [19] Z. Jorgensen, Y. Zhou, and M. Inge, "A multiple instance learning strategy for combating good word attacks on spam filters," *The Journal of Machine Learning Research*, vol. 9, pp. 1115–1146, 2008.
- [20] L. Sørensen, M. Loog, D. M. J. Tax, W. J. Lee, M. De Bruijne, and R. P. W. Duin, "Dissimilarity-based multiple instance learning," in *Structural, Syntactic, and Statistical Pattern Recognition*, vol. 6218, pp. 129–138, Springer, 2010.
- [21] V. Cheplygina, D. M. J. Tax, and M. Loog, "Multiple instance learning with bag dissimilarities," *Pattern Recognition*, vol. 48, no. 1, pp. 264–275, 2015.
- [22] J. Wang and J. D. Zucker, "Solving multiple-instance problem: a lazy learning approach," in *Proceedings of the 17th International Conference on Machine Learning*, pp. 1119–1125, San Francisco, Calif, USA, 2000.
- [23] T. Gärtner, P. A. Flach, A. Kowalczyk, and A. J. Smola, "Multi-instance kernels," in *Proceedings of the International Conference on Machine Learning (ICML '02)*, vol. 2, pp. 179–186, 2002.
- [24] H.-Y. Wang, Q. Yang, and H. Zha, "Adaptive p-posterior mixture-model kernels for multiple instance learning," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1136–1143, Helsinki, Finland, July 2008.
- [25] D. Zhang, Y. Liu, L. Si, J. Zhang, and R. D. Lawrence, "Multiple instance learning on structured data," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 24, pp. 145–153, 2011.
- [26] Q. Zhang and S. A. Goldman, "EM-DD: an improved multiple-instance learning technique," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 14, pp. 1073–1080, MIT Press, 2001.
- [27] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, 2007.
- [28] Y. Tian, Z. Qi, X. Ju, Y. Shi, and X. Liu, "Nonparallel support vector machines for pattern classification," *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 1067–1079, 2014.
- [29] D. Tomar and S. Agarwal, "Twin support vector machine: a review from 2007 to 2014," *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 55–69, 2015.
- [30] O. L. Mangasarian and E. W. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 69–74, 2006.
- [31] X. Xu and E. Frank, "Logistic regression and boosting for labeled bags of instances," in *Advances in Knowledge Discovery and Data Mining*, H. Dai, R. Srikant, and C. Zhang, Eds., vol. 3056 of *Lecture Notes in Computer Science*, pp. 272–281, Springer, 2004.
- [32] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.
- [33] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
- [34] Z. Zhao, G. Fu, S. Liu et al., "Drug activity prediction using multiple-instance learning via joint instance and feature selection," *BMC Bioinformatics*, vol. 14, supplement 14, article S16, 2013.
- [35] Z.-H. Zhou, K. Jiang, and M. Li, "Multi-instance learning based web mining," *Applied Intelligence*, vol. 22, no. 2, pp. 135–147, 2005.
- [36] Y. Chevaleyre and J. D. Zucker, "Solving multiple-instance and multiple-part learning problems with decision trees and rule sets. Application to the mutagenesis problem," in *Advances in Artificial Intelligence: 14th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI 2001 Ottawa, Canada, June 7–9, 2001 Proceedings*, vol. 2056 of *Lecture Notes in Computer Science*, pp. 204–214, Springer, Berlin, Germany, 2001.
- [37] Z. Fu and A. Robles-Kelly, "Fast multiple instance learning via L1,2 logistic regression," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, IEEE, Tampa, Fla, USA, December 2008.
- [38] J. Ramon and L. De Raedt, "Multi instance neural networks," in *Proceedings of the ICML-2000 Workshop on Attribute-Value and Relational Learning*, pp. 53–60, 2000.
- [39] Z. H. Zhou and M. L. Zhang, "Neural networks for multi-instance learning," in *Proceedings of the International Conference on Intelligent Information Technology*, pp. 455–459, Beijing, China, 2002.
- [40] W. Shen, X. Bai, Z. Hu, and Z. Zhang, "Multiple instance subspace learning via partial random projection tree for local reflection symmetry in natural images," *Pattern Recognition*, vol. 52, pp. 306–316, 2016.
- [41] M. Birisan and P. A. Beling, "A multi-instance learning approach to filtering images for presentation to analysts," *Environment Systems and Decisions*, vol. 34, no. 3, pp. 406–416, 2014.
- [42] I. Gondra and T. Xu, "Image region re-weighting via multiple instance learning," *Signal, Image and Video Processing*, vol. 4, no. 4, pp. 409–417, 2010.
- [43] Y.-Y. Xu, "Multiple-instance learning based decision neural networks for image retrieval and classification," *Neurocomputing*, vol. 171, pp. 826–836, 2016.

- [44] O. Maron and A. L. Ratan, "Multiple-instance learning for natural scene classification," in *Proceedings of the International Conference on Machine Learning (ICML '98)*, pp. 341–349, Madison, Wis, USA, 1998.
- [45] H. Cheng, K. A. Hua, and N. Yu, "An automatic feature generation approach to multiple instance learning and its applications to image databases," *Multimedia Tools and Applications*, vol. 47, no. 3, pp. 507–524, 2010.
- [46] H. T. Pao, S. C. Chuang, Y. Y. Xu, and H.-C. Fu, "An EM based multiple instance learning method for image classification," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1468–1472, 2008.
- [47] F. Sener and N. Ikizler-Cinbis, "Ensemble of multiple instance classifiers for image re-ranking," *Image and Vision Computing*, vol. 32, no. 5, pp. 348–362, 2014.
- [48] F. Li and R. Liu, "Graph-based multiple-instance learning with instance weighting for image retrieval," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 2453–2456, IEEE, Brussels, Belgium, September 2011.
- [49] S. Feng, W. Xiong, B. Li, C. Lang, and X. Huang, "Hierarchical sparse representation based multi-instance semi-supervised learning with application to image categorization," *Signal Processing*, vol. 94, no. 1, pp. 595–607, 2014.
- [50] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I.-C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '14)*, pp. 1626–1630, Florence, France, May 2014.
- [51] J. Wu, Y. Yinan, C. Huang, and Y. Kai, "Deep multiple instance learning for image classification and auto-annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 3460–3469, IEEE, Boston, Mass, USA, June 2015.
- [52] D. Kotzias, M. Denil, P. Blunsom, and N. de Freitas, "Deep multi-instance transfer learning," <https://arxiv.org/abs/1411.3128>.
- [53] J. Ding, H. D. Cheng, J. Huang, J. Liu, and Y. Zhang, "Breast ultrasound image classification based on multiple-instance learning," *Journal of Digital Imaging*, vol. 25, no. 5, pp. 620–627, 2012.
- [54] C. Li, S. Zhang, H. Zhang et al., "Using the K-nearest neighbor algorithm for the classification of lymph node metastasis in gastric cancer," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 876545, 11 pages, 2012.
- [55] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," *Medical Image Analysis*, vol. 18, no. 5, pp. 808–818, 2014.
- [56] G. Quellec, M. Lamard, M. D. Abràmoff et al., "A multiple-instance learning framework for diabetic retinopathy screening," *Medical Image Analysis*, vol. 16, no. 6, pp. 1228–1240, 2012.
- [57] W. He and Y. Wang, "Text representation and classification based on multi-instance learning," in *Proceedings of the 16th International Conference on Management Science and Engineering (ICMSE '09)*, pp. 34–39, Moscow, Russia, September 2009.
- [58] C. Xu, W. Tao, Z. Meng, and Z. Feng, "Robust visual tracking via online multiple instance learning with Fisher information," *Pattern Recognition*, vol. 48, no. 12, pp. 3917–3926, 2015.
- [59] J. Xu, S. Denman, V. Reddy, C. Fookes, and S. Sridharan, "Real-time video event detection in crowded scenes using MPEG derived features: a multiple instance learning approach," *Pattern Recognition Letters*, vol. 44, pp. 113–125, 2014.
- [60] X. Ding, G. Zhang, Y. Ke, B. Ma, and Z. Li, "High efficient intrusion detection methodology with twin support vector machines," in *Proceedings of the International Symposium on Information Science and Engineering (ISISE '08)*, vol. 1, pp. 560–564, Shanghai, China, December 2008.
- [61] J. He and S.-H. Zheng, "Intrusion detection model with twin support vector machines," *Journal of Shanghai Jiaotong University (Science)*, vol. 19, no. 4, pp. 448–454, 2014.
- [62] J. A. Nasiri, N. M. Charkari, and K. Mozafari, "Energy-based model of least squares twin Support Vector Machines for human action recognition," *Signal Processing*, vol. 104, pp. 248–257, 2014.
- [63] H.-Y. Yang, X.-Y. Wang, P.-P. Niu, and Y.-C. Liu, "Image denoising using nonsubsampling shearlet transform and twin support vector machines," *Neural Networks*, vol. 57, pp. 152–165, 2014.
- [64] D. Tomar, D. Ojha, and S. Agarwal, "An emotion detection system based on multi least squares twin support vector machine," *Advances in Artificial Intelligence*, vol. 2014, Article ID 282659, 11 pages, 2014.
- [65] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7535–7543, 2009.
- [66] S. Agarwal and D. Tomar, "A feature selection based model for software defect prediction," *International Journal of Advanced Science and Technology*, vol. 65, pp. 39–58, 2014.
- [67] D. Tomar and S. Agarwal, "Prediction of defective software modules using class imbalance learning," *Applied Computational Intelligence and Soft Computing*, vol. 2016, Article ID 7658207, 12 pages, 2016.
- [68] D. Tomar and S. Agarwal, "Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes," *Advances in Artificial Neural Systems*, vol. 2015, Article ID 265637, 10 pages, 2015.
- [69] D. Tomar, B. R. Prasad, and S. Agarwal, "An efficient Parkinson disease diagnosis system based on least squares twin support vector machine and particle swarm optimization," in *Proceedings of the 9th IEEE International Conference on Industrial and Information Systems (ICIIS '14)*, pp. 1–6, IEEE, Gwalior, India, December 2014.
- [70] Z. Wu and C. Yang, "Study to multi-twin support vector machines and its applications in speaker recognition," in *Proceedings of the International Conference on Computational Intelligence and Software Engineering (CiSE '09)*, pp. 1–4, Wuhan, China, December 2009.
- [71] Y. Rubner, C. Tomasi, and L. J. Guibas, "Earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [72] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Tech. Rep., Department of Computer Science, National Taiwan University, Taipei, Taiwan, 2013.
- [73] R. Khemchandani and S. Sharma, "Robust Least Squares Twin Support vector machine for human activity recognition," *Applied Soft Computing*, vol. 47, pp. 33–46, 2016.
- [74] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

- [75] P. Nemenyi, *Distribution-free multiple comparisons [Ph.D. thesis]*, Princeton University, 1963.
- [76] I. Brown and C. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 3446–3453, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

