

Article

An Improved SVM with Earth Mover's Distance Regularization and Its Application in Pattern Recognition

Rui Feng ¹, Haitao Dong ^{1,2,*}, Xuri Li ³, Zhaochuang Gu ³, Runyang Tian ⁴ and Houde Li ⁵¹ School of Electronic Engineering, Xidian University, Xi'an 710071, China² School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China³ Oil & Gas Technology Institute of Changqing Oilfield Company, Xi'an 710021, China⁴ College of Economics & Management, Shandong University of Science and Technology, Qingdao 266590, China⁵ School of Computer and Information Science, Southwest University, Chongqing 400715, China

* Correspondence: donghaitao@xidian.edu.cn or hai3water@126.com

Abstract: A support vector machine (SVM) aims to achieve an optimal hyperplane with a maximum interclass margin and has been widely utilized in pattern recognition. Traditionally, a SVM mainly considers the separability of boundary points (i.e., support vectors), while the underlying data structure information is commonly ignored. In this paper, an improved support vector machine with earth mover's distance (EMD-SVM) is proposed. It can be regarded as an improved generalization of the standard SVM, and can automatically learn the distribution between the classes. To validate its performance, we discuss the necessity of the structural information of EMD-SVM in the linear and nonlinear cases, respectively. Experimental validation was designed and conducted in different application fields, which have shown its superior and robust performance.

Keywords: support vector machine; structural information; Earth mover's distance; deep convolutional feature; pattern recognition



Citation: Feng, R.; Dong, H.; Li, X.; Gu, Z.; Tian, R.; Li, H. An Improved SVM with Earth Mover's Distance Regularization and Its Application in Pattern Recognition. *Electronics* **2023**, *12*, 645. <https://doi.org/10.3390/electronics12030645>

Academic Editor: Sotirios K. Goudos

Received: 3 December 2022

Revised: 22 January 2023

Accepted: 24 January 2023

Published: 28 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A support vector machine (SVM) is a supervised machine learning model that has been widely utilized in pattern recognition [1], such as text classification [2,3], face recognition [4,5], radar [6], sonar [7], etc. Generally, the basic idea of SVM is to maximize the minimum margin from the samples to the classification hyperplane. Built on the SVM, variants, including some discriminative classifiers featuring large-margin theory, have been proposed to improve the SVM or overcome its limitations [8–13].

For classification tasks, the standard SVM aims to find a hyperplane that allows diverse classes to be separated with a maximal margin. However, traditional SVMs mainly consider the separability of boundary points, while the underlying data structure information is commonly ignored. It is known that in real-world applications, different data sets may have different distributions, and from a statistical perspective, the structural information should be the key factor. Breiman et al. argued this point and showed that maximizing the minimum margin was not the key factor in model generalization [14,15]. Then, Reyzin found that the margin theory was still helpful to model generalization, but the margin distribution seemed more dominant [16]. In this case, a classifier is expected to capture the data structure or distribution information, and a more reasonable discriminant boundary would be available when dealing with the complex structured dataset in certain classification tasks. Gao proved that the margin mean and margin variance do have an essential influence on the generalization performance of the classifier [17]. Subsequently, large margin machine (LDM) and its modified version, i.e., optimal margin distribution learning machine (ODM), are proposed to maximize or minimize the margin mean and margin variance, respectively [18,19]. Considering the sensitivity of the number of samples

and its inclination to generate an imbalanced margin distribution, Cheng considered the statistical characteristics with marginal distribution and constructed a double distribution support vector machine (DDSV) [20]. Additionally, due to the utilization of the sample distribution information, the improved SVMs have shown a superior performance [21–27].

One method is to introduce structural information into the SVM. Belkin et al. [28] proposed a laplacian support vector machine (LapSVM) by constructing a laplacian matrix of the manifold structure of the dataset and embedding a manifold regularization term inside the SVM. This approach is called the semi-supervised learning task. Based on this, a structured large margin machine (SLMM) [29] is proposed to capture the structural information by using clustering techniques, which have proved to be sensitive to data distribution. However, the SLMM is optimized by second-order cone programming (SOCP), which has a large computational complexity. Furthermore, research has improved the SVM from the perspective of the objective function, of which the most representative method is structural regularized support vector machine (SRSVM) [30]. Similar to the SLMM, the SRSVM also obtains structural information by using the clustering method, whereas SRSVM integrates the structural information directly into the objective function of a traditional SVM, rather than into the constraints. That is, SRSVM can also be solved by quadratic programming, hence a SVM with minimum within-class scatter (WCS-SVM) was proposed to combine minimum within-class scatter with SVM [31]. Additionally, it is further extended to a fuzzy version coined FSVM with minimum within-class scatter (WCS-FSVM) [32]. To enhance the discriminative ability, Zhang introduced Fisher regularization into SVM to form a Fisher regularization support vector machine (FisherSVM) [33] that minimizes the within-class samples.

Overall, the structural SVM has matured to the extent that it can utilize structural information from the data and improve the generalization capacities of the model. It is usually expected to construct a classification model by explicitly mining structural information. Therefore, the available model is sensitive to data structure information, thus resulting in a general improvement in the model. Motivated by the aforementioned analysis, a novel pattern recognition classifier, namely a support vector machine based on earth mover's distance (EMD-SVM), is proposed to learn the distribution information between classes automatically. Specifically, we utilized earth mover's distance [34] to capture structural information for data explicitly, and then the structural information will be embedded into the SVM, which acts as a regular term of the objective function optimized by quadratic programming. Additionally, we extended the EMD-SVM formulation from the linear classification to the nonlinear case. Considering the great success and state-of-the-art performance of deep neural networks in machine vision and signal processing fields [35–41], we replaced fully-connected layers in a standard CNN using SVM to cope with classification tasks [42–46], which enabled improvements to be made in the recognition performance and generalization ability of CNN.

In terms of this, an improved support vector machine with earth mover's distance (EMD-SVM) is proposed, which can be regarded as an improved generalization of the standard SVM. The main contributions of this study can be summarized as follows,

- (1) We propose a new strategy to capture the underlying data structural information and thus improve the SVM classifier.
- (2) The principles of the EMD-SVM in the linear and nonlinear cases are discussed in detail, respectively. It is proved to be a convex optimization problem and can be solved by the QP technique.
- (3) We conduct experimental verification on three kinds of classification datasets, including UCI, image recognition, and radar emitter recognition, which have shown that the performance of the proposed EMD-SVM is superior and robust.

The rest of this paper is organized as follows. Section 2 briefly describes SVM and Earth Mover's Distance (EMD). The proposed EMD-SVM is introduced in Section 3, which is followed by numerical results in Section 4. Section 5 presents the conclusions.

2. Related Work

2.1. Support Vector Machine

Taking binary classification for instance, we review the principle of SVM. Usually, we use a testing set D to evaluate the discriminative ability of the classifier for new samples, and then use the “testing error” on it as an approximation of the generalization error. Considering the training samples set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, y_i \in \{-1, +1\}$, the standard SVM aims to find a hyperplane $f = \omega^T x + b$, which can separate the samples of different classes with a margin of $\frac{2}{\|\omega\|}$. Its objective function can be given as follows,

$$\begin{aligned} \min & \frac{1}{2} \omega^T \omega \\ \text{s.t.} & y_i(\omega^T x_i + b) \geq 1, i = 1, 2, \dots, n \end{aligned} \tag{1}$$

For the linear non-separable cases, by using $\zeta_i \geq 0, i = 1, 2, \dots, n$ and penalty factors to penalize the samples that violate inequality constraints, the following soft-margin SVM can be obtained,

$$\begin{aligned} \min & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \zeta_i \\ \text{s.t.} & y_i(\omega^T x_i + b) \geq 1 - \zeta_i \\ & \zeta_i \geq 0, i = 1, 2, \dots, n \end{aligned} \tag{2}$$

where ζ_i is the slack variables, the matrix C denotes to the trade-off between errors of training data and generalization [47].

Then, the standard SVM can be trained by solving a dual quadratic programming problem. The dual problem can be formulated as below,

$$\begin{aligned} \max & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^n \alpha_i \\ \text{s.t.} & 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \tag{3}$$

2.2. The Earth Mover’s Distance

The Earth Mover’s Distance is defined as the minimal cost that transforms one distribution into the other. It is proposed to solve the transportation problem through linear optimization. Next, we explain the algorithm by referring to the cargo transportation problem.

Suppose there are two distributions $P = \{(p_i, u_{p_i})\}_{i=1}^m$ and $Q = \{(q_j, u_{q_j})\}_{j=1}^n$, where p_i is the supplier, u_{p_i} is the quantity of goods it owns, q_j is the warehouse, and u_{q_j} is the quantity of goods it can accommodate. Then, the EMD distance can be expressed as a linear optimization problem as below,

$$\begin{aligned} \text{WORK}(P, Q, F) &= \sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij} \\ \text{s.t.} & f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n \\ & \sum_{j=1}^n f_{ij} \leq \omega_{p_i} \quad 1 \leq i \leq m \\ & \sum_{i=1}^m f_{ij} \leq \omega_{q_j} \quad 1 \leq j \leq n \\ & \sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m \omega_{p_i}, \sum_{j=1}^n \omega_{q_j} \right) \\ & d_{ij} = |p_i - q_j| \end{aligned} \tag{4}$$

where, d_{ij} represents the distance between p_i and q_j , ω_{p_i} and ω_{q_j} denotes the total supply and accommodation capacity, respectively. Here we expect to find a flow $F = [f_{ij}]$ that minimizes the overall transportation costs, where f_{ij} is the flow from p_i to q_j . Then, the EMD distance can be normalized as,

$$EMD(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \tag{5}$$

3. EMD-SVM Model

In this section, the EMD-SVM model is expounded by taking binary classification as an example. Then, the principles of the EMD-SVM in the linear and nonlinear cases are discussed in detail, respectively.

3.1. EMD-SVM for Linear Case

The EMD-SVM model for the linear case can be given as,

$$\begin{aligned} \min & \frac{1}{2} \omega^T \omega + \frac{\lambda}{2} \omega^T E_d \omega + C \sum_{i=1}^n \xi_i \\ \text{s.t.} & y_i (\omega^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \tag{6}$$

where $E_d = \frac{1}{emd} I$, I is the identity matrix, emd represents the EMD distance between the two kinds of distributions, and λ is a parameter used to regulate the relative importance of the distance representation within the distribution of the two classes.

Incorporating the constraints, we can rewrite Equation (5) as a primal Lagrangian equation,

$$L(\omega, b, \alpha) = \frac{1}{2} \omega^T \omega + \frac{\lambda}{2} \omega^T E_d \omega + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i [1 - \xi_i - y_i (\omega^T x_i + b)] - \sum_{i=1}^n \mu_i \xi_i \tag{7}$$

where the KKT conditions for the primal problem can be obtained as follows,

$$\frac{\partial L}{\partial \omega} = (I + \lambda E_d) \omega - \sum_{i=1}^n \alpha_i y_i x_i = 0 \tag{8}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \tag{9}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \tag{10}$$

$$\alpha_i \geq 0, \mu_i \geq 0 \tag{11}$$

$$y_i (\omega^T x_i + b) - 1 + \xi_i \geq 0 \tag{12}$$

$$\alpha_i (y_i (\omega^T x_i + b) - 1 + \xi_i) = 0 \tag{13}$$

$$\xi_i \geq 0, \mu_i \xi_i \geq 0 \tag{14}$$

Substitute Equations (8)–(10) into Equation (7), we can obtain

$$\begin{aligned}
 & \frac{1}{2}\omega^T(I + \lambda E_d)\omega + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i [1 - \xi_i - y_i(\omega^T x_i + b)] - \sum_{i=1}^n \mu_i \xi_i \\
 &= \frac{1}{2}\sum_{i=1}^n \alpha_i y_i x_i^T (I + \lambda E_d)^{-1} \sum_{j=1}^n \alpha_j y_j x_j + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i \sum_{j=1}^n \alpha_j y_j x_j^T (I + \lambda E_d)^{-1} x_i \\
 &= -\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T (I + \lambda E_d)^{-1} x_j + \sum_{i=1}^n \alpha_i
 \end{aligned} \tag{15}$$

Then, we can transform the primal Lagrangian equation into the dual problem,

$$\begin{aligned}
 & \max -\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T (I + \lambda E_d)^{-1} x_j + \sum_{i=1}^n \alpha_i \\
 & \text{s.t. } 0 \leq \alpha_i \leq C, i = 1, \dots, n \\
 & \sum_{i=1}^n \alpha_i y_i = 0
 \end{aligned} \tag{16}$$

Hence, we can obtain the solution α_i by using the QP techniques. In predicting the class labels for testing data x , the classifier function can be derived as below,

$$\begin{aligned}
 \text{Class } x &= \text{sgn}[\omega^T x + b] \\
 &= \text{sgn}\left[\sum_{i=1}^n \alpha_i y_i x_i^T (I + \lambda E_d)^{-1} x + b\right]
 \end{aligned} \tag{17}$$

3.2. EMD-SVM for Nonlinear Case

Like the SVM, we can also construct the kernel functions for EMD-SVM to cope with nonlinear problems. We can construct a mapping function Φ so that to map the training data to a higher-level feature space H , i.e., $\Phi : R^d \mapsto H$. Then, the kernel transposition problem can be described as [47],

$$\text{WORK}(\phi(P), \phi(Q), F) = \sum_{i=1}^m \sum_{j=1}^n d_{\phi(p_i)\phi(q_j)} f_{ij} \tag{18}$$

where the ground distance $d_{\phi(p_i)\phi(q_j)}$ between $\phi(p_i)$ and $\phi(q_j)$ can be calculated by,

$$d_{\phi(p_i)\phi(q_j)} = \sqrt{\|\phi(p_i) - \phi(q_j)\|^2} \tag{19}$$

If there were a “kernel function” K such that $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, we would use K to rewrite the Equation (19) as,

$$d_{\phi(p_i)\phi(q_j)} = \sqrt{K(p_i, p_i) - K(p_i, q_j) - K(q_j, p_i) + K(q_j, q_j)} \tag{20}$$

Then, the kernel of the EMD-SVM can be defined,

$$\begin{aligned}
 & \min \frac{1}{2}\omega^T \omega + \frac{\lambda}{2}\omega^T \phi(X) E_d \phi^T(X) \omega + C\sum_{i=1}^n \xi_i \\
 & \text{s.t. } y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i \\
 & \xi_i \geq 0, i = 1, 2, \dots, n
 \end{aligned} \tag{21}$$

where E_d represents the EMD distance of the two kinds of distributions in the kernel space.

The Lagrangian form for this problem can be formed as below,

$$\begin{aligned}
 L(\omega, b, \alpha) &= \frac{1}{2}\omega^T\omega + \frac{\lambda}{2}\omega^T\phi(X)E_d\phi^T(X)\omega + C\sum_{i=1}^n\zeta_i \\
 &+ \sum_{i=1}^n\alpha_i[1 - \zeta_i - y_i(\omega^T\phi(x_i) + b)] - \sum_{i=1}^n\mu_i\zeta_i
 \end{aligned}
 \tag{22}$$

Setting the partial derivative of $L(\omega, b, \alpha)$ with respect to ω and b equal to zero,

$$\omega = [I + \lambda\phi(X)E_d\phi^T(X)]^{-1}\sum_{i=1}^n\alpha_i y_i \phi(x_i)
 \tag{23}$$

$$\sum_{i=1}^n\alpha_i y_i = 0
 \tag{24}$$

$$C - \alpha_i - \mu_i = 0
 \tag{25}$$

The dual problem can be further formed as,

$$\begin{aligned}
 \max & -\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\alpha_i\alpha_j y_i y_j \phi^T(x_i)(I + \lambda\phi(X)E_d\phi^T(X))^{-1}\phi(x_j) + \sum_{i=1}^n\alpha_i \\
 \text{s.t.} & 0 \leq \alpha_i \leq C, i = 1, \dots, n \\
 & \sum_{i=1}^n\alpha_i y_i = 0
 \end{aligned}
 \tag{26}$$

According to Woodbury’s formula, we could have,

$$(I + UB V)^{-1} = A^{-1} - A^{-1}UB(B + BVA^{-1}UB)^{-1}BVA^{-1}
 \tag{27}$$

Then, we have,

$$\begin{aligned}
 & [I + \lambda\phi(X)E_d\phi^T(X)]^{-1} \\
 &= I - \lambda\phi(X)E_d[E_d + \lambda E_d\phi^T(X)\phi(X)E_d]^{-1}E_d\phi^T(X) \\
 &= I - \lambda\phi(X)E_d[E_d + \lambda E_d K E_d]^{-1}E_d\phi^T(X) \\
 &= I - \lambda\phi(X)P\phi^T(X)
 \end{aligned}
 \tag{28}$$

where

$$P = E_d[E_d + \lambda E_d K E_d]E_d
 \tag{29}$$

Let K_i denote the i -th row of K , K_j denote the j -th column of K , then the dual problem can be cast as

$$\begin{aligned}
 & -\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\alpha_i\alpha_j y_i y_j [K_{ij} - \lambda K_i P K_j] + \sum_{i=1}^n\alpha_i \\
 \text{s.t.} & 0 \leq \alpha_i \leq C, i = 1, \dots, n \\
 & \sum_{i=1}^n\alpha_i y_i = 0
 \end{aligned}
 \tag{30}$$

Once the solution α are obtained from the above convex optimization problem, we can obtain the hyperplane. Then, the class label of a data point $x \in R^n$ can be determined as

$$\begin{aligned}
 \text{Class } x &= \text{sgn}[\omega^T\phi(x) + b] \\
 &= \text{sgn}\left[\sum_{i=1}^n\alpha_i y_i K(x_i, x) - \lambda\sum_{i=1}^n\alpha_i y_i K_i P K(X, x) + b\right]
 \end{aligned}
 \tag{31}$$

4. Experimental Results and Discussion

In this section, the EMD-SVM is evaluated on synthetic and real-world datasets. We compared the performance of the proposed EMD-SVM with standard SVM and some representative large margin methods, including SRSVM, LDM, ODM, and ELM [48]. We first evaluated the effectiveness of the proposed EMD-SVM on a synthetic dataset to illustrate the impact of data distribution information on classification. Then, we evaluated the performance of these methods on UCI datasets and the Caltech 101 dataset. Next, we utilized a deep convolutional neural network to extract convolutional features and discuss the performance of EMD-SVM based on deep convolutional features. Finally, the proposed EMD-SVM was applied to the radar emitter recognition task. All the experiments were carried out on a PC with a 3.50 GHz CPU and 48 GB RAM.

4.1. Recognition Performance of EMD-SVM on Synthetic Dataset

The two-dimensional synthetic dataset consists of three groups of randomly generated Gaussian distributions. The blue plus represents the positive sample and the red star represents the negative samples. Table 1 describes the attributes of the dataset. The hyperplanes of linear SVM, LDM, and EMD-SVM are displayed in Figure 1.

Table 1. The attributes of the synthetic dataset.

Samples	Gaussian Distribution	Num	Mean	Covariance
Positive	Gaussian distribution P	200	[0; 5]	[0.3,0; 0,5]
Negative	Gaussian distribution N_1	180	[6; 8]	[1,0; 0,0.5]
	Gaussian distribution N_2	20	[3; 2]	[1.5,0; 0,1.5]

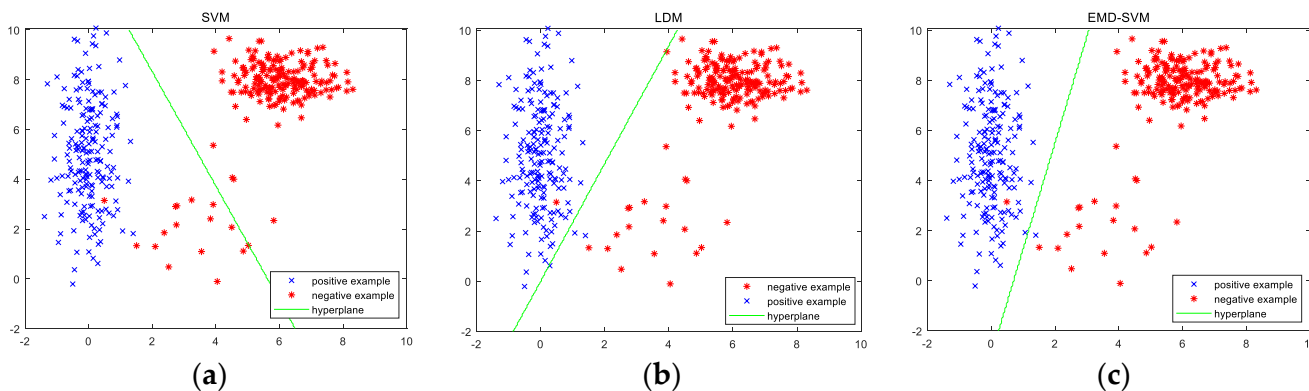


Figure 1. Hyperplanes of SVM and EMD-SVM. (a) SVM; (b) LDM; (c) EMD-SVM.

As can be seen from Figure 1, the positive class has a vertical distribution and the negative class is composed of two horizontal Gaussian distributions. Moreover, the distribution N_1 has more samples than N_2 . It can be seen from Figure 1 that due to the neglect of the structural information, SVM cannot compete with a complex structured dataset; Specifically, SVM ignores the cluster N_2 which has fewer samples than cluster N_1 . The hyperplane only focuses on the separability between cluster P and cluster N_1 . LDM adopts margin mean and variance to characterize the margin distribution and optimizes it to achieve a better generalization performance. On the other hand, considering the structured distance information and the separability between the two distributions, the proposed EMD-SVM can also obtain a more reasonable hyperplane.

4.2. Recognition Performance of EMD-SVM with Hand-Crafted Classifier on UCI Datasets

In this section, we verify the performance of the proposed EMD-SVM on UCI datasets. The attributes of these datasets are presented in Table 2.

Table 2. Attributes of experimental datasets.

Dataset	Feature	Num	Class
Sonar	60	208	2
Breast	9	277	2
Cryotherapy	6	90	2
Fertility	9	100	2
Wdbc	30	569	2
Ionosphere	34	351	2
Hepatitis	19	155	2
Spectf	44	267	2
Pima	8	768	2
Heart	13	303	2
Tae	5	151	3
Iris	4	150	3

We randomly selected half of the samples as the training set and the rest as the testing set. In the linear case, for ODM, the parameter D is selected from the set $[0, 0.1, \dots, 0.5]$, the regularization parameter C_1 and C_2 are selected from the set $[2^{-8}, \dots, 2^8]$, for SVM, SRSVM, EMD-SVM and LDM, the parameter C is selected from the set $[10^{-3}, \dots, 10^3]$, and parameters λ , λ_1 and λ_2 are selected from the set $[2^{-8}, \dots, 2^8]$. For ELM, the number of hidden neurons is defined as 1000, and the activation function is Sigmoid. In the nonlinear case, the RBF kernel $k(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right)$ is used for all algorithms. The width of the RBF kernel is selected from $[2^{-8}, \dots, 2^8]$. Experiments were repeated 10 times with different data partitions.

We compared the average accuracy of all the algorithms. Table 3 shows the accuracy result with linear kernel and Table 4 shows the accuracy result with RBF kernel.

Table 3. Accuracy comparisons with linear kernel.

Dataset	SVM	SRSVM	LDM	ODM	ELM	EMD-SVM
Sonar	77.11 ± 3.84	77.40 ± 1.50	76.83 ± 3.80	76.35 ± 4.30	78.26 ± 3.79	77.69 ± 4.12
Breast	70.86 ± 2.78	71.80 ± 2.20	70.72 ± 2.06	70.58 ± 2.26	57.62 ± 6.64	71.80 ± 2.65
Cryotherapy	83.91 ± 5.10	85.35 ± 5.92	70.39 ± 4.99	68.77 ± 7.98	78.77 ± 7.33	85.64 ± 6.24
Fertility	86.80 ± 3.55	86.80 ± 3.55	86.80 ± 3.55	86.80 ± 3.55	74.2 ± 4.36	87.00 ± 4.02
Wdbc	95.12 ± 1.13	96.63 ± 0.99	94.88 ± 1.11	91.68 ± 1.83	87.75 ± 2.09	96.04 ± 1.14
Ionosphere	87.10 ± 1.69	87.84 ± 1.76	84.03 ± 2.74	84.49 ± 2.40	80.85 ± 2.05	87.73 ± 2.3
Hepatitis	80.75 ± 7.64	83.75 ± 4.28	83.75 ± 3.58	85.50 ± 4.68	76.25 ± 2.71	83.75 ± 4.28
Spectf	78.81 ± 3.75	79.85 ± 3.50	79.18 ± 2.32	79.10 ± 2.25	63.88 ± 5.34	79.78 ± 2.84
Pima	76.07 ± 1.88	76.11 ± 1.84	66.90 ± 1.94	67.14 ± 2.09	60.91 ± 2.48	76.30 ± 1.84
Heart	83.09 ± 2.63	83.28 ± 3.15	83.36 ± 2.15	82.76 ± 2.14	68.94 ± 4.48	83.42 ± 3.05
Tae	48.68 ± 3.59	51.04 ± 4.94	44.73 ± 5.10	44.87 ± 4.51	50.52 ± 4.85	51.45 ± 3.89
Iris	97.60 ± 2.41	97.80 ± 1.56	97.07 ± 1.24	97.07 ± 1.39	77.06 ± 6.85	97.87 ± 1.05

Note: the bold value indicates the best accuracy on each dataset.

From the results, we can draw the following conclusions,

- (1) The EMD-SVM combines the earth mover's distance with standard SVM, which can introduce the data distribution information into the traditional SVM. The outstanding performance of EMD-SVM on most datasets further validates the necessity of distribution information for the classifier's design.
- (2) Although SRSVM can achieve comparable recognition results with EMD-SVM, its recognition performance is highly affected by the clustering method, as SRSVM is based on the clustering structure. In practical applications, different clustering methods must be used for different problems.
- (3) LDM and ODM use the margin mean and variance to describe the margin distribution, while the first- and second-order statistics are often used to characterize Gaussian-distributed data, which has certain limitations. In contrast, EMD-SVM adopts EMD

distance instead of Euclidean distance to describe the data distribution. The distribution information is then incorporated into the SVM object function in the form of regular terms, thus guiding SVM to learn the optimal classification boundary under this distribution metric.

Table 4. Accuracy comparisons with RBF kernel.

Dataset	SVM	SRSVM	LDM	ODM	ELM	EMD-SVM
Sonar	87.88 ± 2.44	87.40 ± 1.84	87.5 ± 2.26	86.54 ± 2.36	86.25 ± 2.31	87.98 ± 2.36
Breast	72.52 ± 3.50	74.39 ± 3.00	74.60 ± 2.81	74.68 ± 3.50	74.31 ± 3.01	74.46 ± 3.31
Cryotherapy	84.00 ± 4.29	85.33 ± 5.36	83.11 ± 8.32	82.22 ± 7.33	82.66 ± 7.96	85.33 ± 6.21
Fertility	87.00 ± 4.02	86.80 ± 3.55	87.00 ± 4.02	86.80 ± 3.55	86.80 ± 3.55	87.00 ± 4.02
Wdbc	94.63 ± 1.44	94.60 ± 1.15	91.61 ± 1.51	91.65 ± 1.37	91.64 ± 1.37	94.35 ± 0.89
Ionosphere	94.71 ± 1.56	94.89 ± 1.20	94.94 ± 1.05	94.94 ± 1.05	94.94 ± 1.05	95.28 ± 1.11
Hepatitis	83.75 ± 4.28	83.75 ± 4.29	78.5 ± 4.74	78.25 ± 4.42	78.25 ± 4.41	84.00 ± 4.74
Spectf	80.22 ± 2.31	79.78 ± 3.25	76.27 ± 2.16	76.34 ± 2.33	76.26 ± 2.16	80.30 ± 2.86
Pima	76.33 ± 1.34	76.74 ± 1.70	76.84 ± 1.88	76.77 ± 1.77	76.85 ± 1.97	76.82 ± 1.47
Heart	83.22 ± 2.65	83.88 ± 3.44	84.14 ± 3.37	84.01 ± 3.22	84.14 ± 2.76	84.41 ± 2.68
Tae	47.63 ± 5.71	46.32 ± 3.33	50.92 ± 7.07	50.92 ± 7.07	41.57 ± 3.73	50.92 ± 7.07
Iris	97.73 ± 1.26	97.87 ± 1.59	97.06 ± 1.64	98.00 ± 1.13	92.00 ± 3.92	98.00 ± 1.13

Note: the bold value indicates the best accuracy on each dataset.

4.3. Experiments on Caltech101 Dataset

In this subsection, we conduct an experiment on the Caltech101 dataset. Caltech101 is a digital image dataset provided by the California Institute of Technology, which contains a total of 9146 images divided into 101 attributes (including face, plane, animal, etc.) and a background category. We chose nine types of images for this experiment: airplanes, bonsai, cars, dolphins, electric guitars, easy faces, helicopters, leopards, and motorbikes. The features of SIFT, LBP, and PHOG are extracted from these images and the attributes are presented in Table 5.

Table 5. Attributes of Caltech101 feature.

Caltech101 Feature	Feature	Num	Class
LBP	37	720	9
SIFT	300	720	9
PHOG	40	720	9

We randomly selected 80 images from each category as datasets, 64 of them as training samples and the remaining 16 as test samples. Ten independent experiments were conducted to evaluate the performance of the proposed EMD-SVM. We used linear kernel in the experiment and the parameters were selected in a similar way as in the UCI dataset experiments. For multi-class problems, the “one-to-one” strategy is adopted. We compared the average accuracy of all the algorithms and the results are shown in Table 6.

Table 6. Accuracy comparisons with linear kernel.

Dataset	SVM	SRSVM	LDM	ODM	ELM	EMD-SVM
LBP	60.34 ± 3.58	60.87 ± 2.18	40.90 ± 4.78	49.09 ± 2.49	60.55 ± 4.93	61.11 ± 1.96
SIFT	82.36 ± 2.14	82.43 ± 2.75	74.24 ± 3.16	78.12 ± 3.37	78.75 ± 1.83	83.26 ± 3.14
PHOG	50.69 ± 2.99	51.59 ± 3.60	41.46 ± 3.43	45.35 ± 3.63	44.03 ± 6.08	51.94 ± 3.87

Note: the bold value indicates the best accuracy on each dataset.

It is clear that the EMD-SVM achieves better accuracy than the SVM, SRSVM, LDM, ODM and ELM methods in the multi-class classification problem. This indicates that distribution information can help to determine a better discriminant boundary. Moreover, the performance of LDM and ODM on the Caltech101 dataset further shows that characterizing the data distribution with first- and second-order statistics still has some limitations.

4.4. Recognition Performance of EMD-SVM Based on Deep Convolutional Features

In this section, we discuss the performance of EMD-SVM and other algorithms on deep convolutional features. We adopted the classical AlexNet as the pretrained CNN model, which contains five convolutional layers and three fully connected layers, and further details of the network can be referred to in [40]. The DSLR and Amazon datasets were used to verify the effectiveness of EMD-SVM in the deep feature. The CNN model was pretrained by the ImageNet dataset and fine-tuned by the DSLR and Amazon datasets. Then, we extracted the fine-tuned deep features Fc6 and Fc7 as the inputs of the above five algorithms for classification, respectively. Table 7 shows the details of the four deep features.

Table 7. Details of the four deep features.

Dataset	CNN Feature	Num	Class
DSLR-Fc6	4096	157	10
DSLR-Fc7	4096	157	10
Amazon-Fc6	4096	958	10
Amazon-Fc7	4096	958	10

In the experiment, we randomly chose 50% of the samples as the training set and the rest as the testing set. Ten independent experiments were conducted to achieve a more stable result. A linear kernel was used in the experiment and the parameters were selected in a similar way as in the UCI dataset experiments. Table 8 compares the accuracy results of EMD-SVM and other methods.

Table 8. Accuracy comparisons with linear kernel.

Dataset	SVM	SRSVM	LDM	ODM	ELM	EMD-SVM
DSLR-Fc6	93.92 ± 2.31	94.05 ± 3.16	95.18 ± 2.58	95.44 ± 2.54	94.55 ± 2.31	96.45 ± 2.04
DSLR-Fc7	94.93 ± 2.05	94.93 ± 2.60	96.20 ± 2.06	96.32 ± 2.26	94.81 ± 3.29	96.96 ± 1.99
AMAZON-Fc6	94.07 ± 0.75	94.17 ± 0.66	94.53 ± 0.83	94.34 ± 0.66	91.19 ± 0.63	94.47 ± 0.69
AMAZON-Fc7	94.55 ± 0.91	94.82 ± 0.80	94.90 ± 0.82	94.92 ± 0.75	92.88 ± 1.04	95.01 ± 0.81

Note: the bold value indicates the best accuracy on each dataset.

As can be seen, the overall performance of EMD-SVM is better than SVM and other methods. In addition, as the large margin algorithms LDM and ODM apply the ideas of maximizing margin mean and minimizing margin variance to the SVM model, compared with SVM, LDM and ODM are also highly competitive with SVM. The results demonstrate that considering the distribution of data can improve the classifier’s classification performance on complex data.

Additionally, we compared EMD-SVM with an MLP with two hidden layers with 1024 and 512 neurons, respectively. The accuracies of EMD-SVM and MLP are shown in Table 9. It can be seen that MLP can achieve recognition results comparable to those of linear EMD-SVM, but still somewhat inferior to nonlinear EMD-SVM. Compared to MLP, EMD-SVM is based on the minimization of structural risk rather than empirical risk, thus avoiding the overfitting problem. By obtaining a structured description of the data distribution, it reduces the requirements for the size and distribution of the data, and has excellent generalization capabilities.

Table 9. Accuracy comparisons with RBF kernel.

Dataset	SVM	MLP	EMD-SVM (Linear Kernel)	EMD-SVM (RBF Kernel)
DSLR-Fc6	93.92 ± 2.31	96.71 ± 1.59	96.45 ± 2.04	96.95 ± 3.16
DSLR-Fc7	94.93 ± 2.05	96.89 ± 2.21	96.96 ± 1.99	96.99 ± 1.34
AMAZON-Fc6	94.07 ± 0.75	94.40 ± 0.88	94.47 ± 0.69	94.58 ± 0.94
AMAZON-Fc7	94.55 ± 0.91	95.21 ± 1.21	95.01 ± 0.81	95.26 ± 1.02

Note: the bold value indicates the best accuracy on each dataset.

4.5. Recognition Performance of EMD-SVM for Radar Emitter Recognition

In order to test the effectiveness of our EMD-SVM in realistic applications, we conducted experiments on radar emitter recognition. The collected data are radar emitter signals with the same type and parameters. We extracted the FFT, Welch power spectrum, ambiguous function slice, and cyclic spectrum slice (denoted as Data1, Data2, Data3, and Data4, respectively). The attributes of these datasets are presented in Table 10. The corresponding waveforms of class1–class4 signals in Data4 are shown in Figure 2.

Table 10. Attributes of the radar datasets.

Dataset	Feature	Num	Class
Data1	5000	833	27
Data2	8192	833	27
Data3	4999	833	27
Data4	2500	833	27

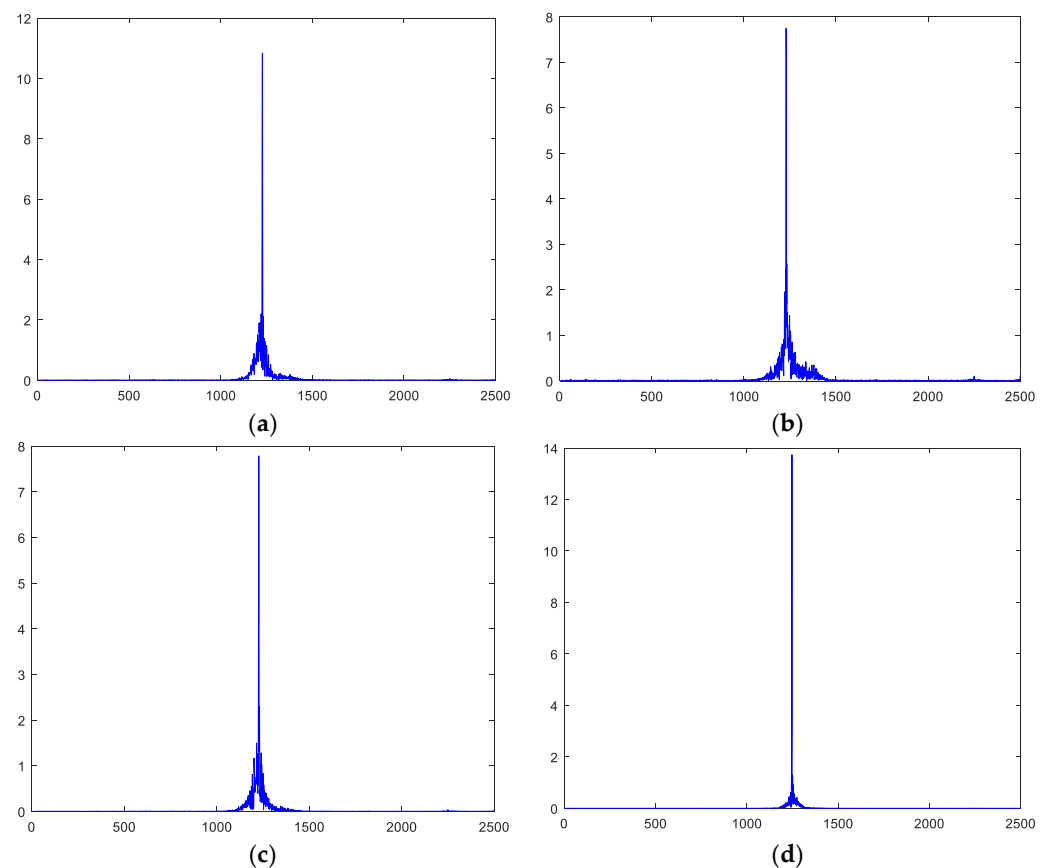


Figure 2. Waveforms of class1–class4 signals in Data4, (a) class1; (b) class2; (c) class3; (d) class4.

To reduce the computation time, the PCA algorithm was utilized to extract 90% of the energy. We randomly chose the 80% percent samples as the training set and the remaining 20% percent samples as the test set. The experiment was repeated 10 times to generate 10 independent results for each dataset, and we compared the average accuracy and the standard deviation of all the algorithms. A linear kernel was used in the experiment and the parameters were selected in a similar way to the UCI dataset experiments. The results of the four radar datasets are shown in Table 11, and it can be seen that our EMD-SVM still achieves superior results.

Table 11. Accuracy comparisons with linear kernel.

Dataset	SVM	SRSVM	LDM	ODM	ELM	EMD-SVM
Data1	54.97 ± 3.65	55.44 ± 3.24	46.23 ± 3.44	50.23 ± 4.05	46.59 ± 2.58	55.51 ± 3.45
Data2	48.02 ± 3.52	48.26 ± 3.30	40.72 ± 3.65	43.29 ± 2.44	45.27 ± 3.89	48.74 ± 3.36
Data3	45.69 ± 3.09	45.69 ± 3.09	43.47 ± 4.28	42.81 ± 1.76	46.29 ± 3.38	46.41 ± 3.27
Data4	50.65 ± 2.43	50.71 ± 2.55	43.89 ± 2.04	45.63 ± 2.59	46.28 ± 3.49	51.14 ± 1.95

Note: the bold value indicates the best accuracy on each dataset.

5. Conclusions

In this paper, we propose a novel SVM classifier with earth mover's distance, which can automatically learn the distribution between the classes. The EMD-SVM can be seen as a generalization of the standard SVM by calculating the EMD distance, and we discuss the principle of the EMD-SVM in linear and nonlinear cases, respectively. The experimental results indicate that the proposed EMD-SVM has a superior and robust performance. In the future, we will pay more attention to overcoming the drawbacks of a long training time for large-scale datasets and sensitivity to hyper-parameters of kernel functions. It would also be interesting to generalize the idea of EMD-SVM to other learning settings.

Author Contributions: Conceptualization, R.F. and Z.G.; methodology, R.F. and H.D.; software, R.F.; validation, H.L.; formal analysis, X.L.; investigation, R.F.; data curation, X.L. and Z.G.; writing—original draft preparation, R.F. and R.T.; writing—review and editing, H.D.; funding acquisition, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (Grant No. 62201439) and China Postdoctoral Science Foundation (Grant No. 2022M722493). The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

Data Availability Statement: The data utilized is open access and can be found by the following websites: http://www.vision.caltech.edu/Image_Datasets/Caltech101/; <https://github.com/jindongwang/transferlearning/blob/master/data/dataset.md#officecaltech>.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

References

1. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
2. Joachims, T. A statistical learning model of text classification for support vector machines. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, LA, USA, 9–13 September 2001; pp. 128–136. [[CrossRef](#)]
3. Lilleberg, J.; Zhu, Y.; Zhang, Y. Support vector machines and Word2vec for text classification with semantic features. In Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), Beijing, China, 6–8 July 2015; pp. 136–140. [[CrossRef](#)]
4. Osuna, E.; Freund, R.; Girosit, F. Training support vector machines: An application to face detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 130–136. [[CrossRef](#)]
5. Ghimire, D.; Jeong, S.; Lee, J.; Park, S.H. Facial expression recognition based on local region specific features and support vector machines. *Multimedia Tools Appl.* **2017**, *76*, 7803–7821. [[CrossRef](#)]
6. Eryildirim, A.; Onaran, I. Pulse Doppler Radar Target Recognition using a Two-Stage SVM Procedure. *IEEE Trans. Aerosp. Electron. Syst.* **2011**, *47*, 1450–1457. [[CrossRef](#)]
7. Dong, H.; Wang, H.; Shen, X.; He, K. Parameter matched stochastic resonance with damping for passive sonar detection. *J. Sound Vib.* **2019**, *458*, 479–496. [[CrossRef](#)]
8. Suykens, J.A.K.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
9. Jayadeva; Khemchandani, R.; Chandra, S. Twin Support Vector Machines for Pattern Classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 905–910. [[CrossRef](#)]
10. Lin, C.-F.; Wang, S.-D. Fuzzy support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 464–471. [[CrossRef](#)]
11. Ding, S.; Zhu, Z.; Zhang, X. An overview on semi-supervised support vector machine. *Neural Comput. Appl.* **2017**, *28*, 969–978. [[CrossRef](#)]

12. Iranmehr, A.; Masnadi-Shirazi, H.; Vasconcelos, N. Cost-sensitive support vector machines. *Neurocomputing* **2019**, *343*, 50–64. [[CrossRef](#)]
13. Huang, S.; Cai, N.; Pacheco, P.P.; Narandes, S.; Wang, Y.; Xu, W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genom. Proteom.* **2018**, *15*, 41–51. [[CrossRef](#)]
14. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1999. [[CrossRef](#)]
15. Breiman, L. Prediction Games and Arcing Algorithms. *Neural Comput.* **1999**, *11*, 1493–1517. [[CrossRef](#)] [[PubMed](#)]
16. Reyzin, L.; Schapire, R.E. How boosting the margin can also boost classifier complexity. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 753–760. [[CrossRef](#)]
17. Gao, W.; Zhou, Z.-H. On the doubt about margin explanation of boosting. *Artif. Intell.* **2013**, *203*, 1–18. [[CrossRef](#)]
18. Zhang, T.; Zhou, Z.-H. Large margin distribution machine. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 313–322. [[CrossRef](#)]
19. Teng, Z.; Zhou, Z.-H. Optimal margin distribution machine. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1143–1156. [[CrossRef](#)]
20. Cheng, F.; Zhang, J.; Li, Z.; Tang, M. Double distribution support vector machine. *Pattern Recognit. Lett.* **2017**, *88*, 20–25. [[CrossRef](#)]
21. Zhou, Y.-H.; Zhou, Z.-H. Large Margin Distribution Learning with Cost Interval and Unlabeled Data. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 1749–1763. [[CrossRef](#)]
22. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5265–5274. [[CrossRef](#)]
23. Elsayed, G.; Krishnan, D.; Mobahi, H.; Regan, K.; Bengio, S. Large margin deep networks for classification. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 850–860.
24. Cheng, F.; Zhang, J.; Wen, C.; Liu, Z.; Li, Z. Large cost-sensitive margin distribution machine for imbalanced data classification. *Neurocomputing* **2017**, *224*, 45–57. [[CrossRef](#)]
25. Zhan, K.; Wang, H.; Huang, H.; Xie, Y. Large margin distribution machine for hyperspectral image classification. *J. Electron. Imaging* **2016**, *25*, 63024. [[CrossRef](#)]
26. Rastogi, R.; Anand, P.; Chandra, S. Large-margin Distribution Machine-based regression. *Neural Comput. Appl.* **2020**, *32*, 3633–3648. [[CrossRef](#)]
27. Abe, S. Unconstrained large margin distribution machines. *Pattern Recognit. Lett.* **2017**, *98*, 96–102. [[CrossRef](#)]
28. Belkin, M.; Niyogi, P.; Sindhvani, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.
29. Yeung, D.S.; Wang, D.; Ng, W.W.Y.; Tsang, E.C.C.; Wang, X. Structured large margin machines: Sensitive to data distributions. *Mach. Learn.* **2007**, *68*, 171–200. [[CrossRef](#)]
30. Xue, H.; Chen, S.; Yang, Q. Structural Regularized Support Vector Machine: A Framework for Structural Large Margin Classifier. *IEEE Trans. Neural Netw.* **2011**, *22*, 573–587. [[CrossRef](#)]
31. An, W.; Liang, M. A new intrusion detection method based on SVM with minimum within-class scatter. *Secur. Commun. Netw.* **2013**, *6*, 1064–1074. [[CrossRef](#)]
32. An, W.; Liang, M. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises. *Neurocomputing* **2013**, *110*, 101–110. [[CrossRef](#)]
33. Zhang, L.; Zhou, W.-D. Fisher-regularized support vector machine. *Inf. Sci.* **2016**, *343–344*, 79–93. [[CrossRef](#)]
34. Rubner, Y.; Tomasi, C.; Guibas, L.J. The Earth Mover’s Distance as a Metric for Image Retrieval. *Int. J. Comput. Vis.* **2000**, *40*, 99–121. [[CrossRef](#)]
35. Kendall, A.; Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5580–5590. [[CrossRef](#)]
36. Zhu, Z.; Yi, Z.; Li, S.; Li, L. Deep Multi-Modal Generic Representation Auxiliary Learning Networks for End-to-End Radar Emitter Classification. *Aerospace* **2022**, *9*, 732. [[CrossRef](#)]
37. Li, L.; Dong, Z.; Zhu, Z.; Jiang, Q. Deep-learning Hopping Capture Model for Automatic Modulation Classification of Wireless Communication Signals. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, 1–12. [[CrossRef](#)]
38. Ribeiro, F.D.S.; Calivá, F.; Swainson, M.; Gudmundsson, K.; Leontidis, G.; Kollias, S. Deep Bayesian Self-Training. *Neural Comput. Appl.* **2020**, *32*, 4275–4291. [[CrossRef](#)]
39. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856. [[CrossRef](#)]
40. Zhu, Z.; Ji, H.; Zhang, W.; Li, L.; Ji, T. Complex Convolutional Neural Network for Signal Representation and Its Application to Radar Emitter Recognition. *IEEE Commun. Lett.* **2023**, *1*. [[CrossRef](#)]
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *60*, 84–90. [[CrossRef](#)]
42. Nagi, J.; Di Caro, G.A.; Giusti, A.; Nagi, F.; Gambardella, L.M. Convolutional Neural Support Vector Machines: Hybrid Visual Pattern Classifiers for Multi-robot Systems. In Proceedings of the 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 12–15 December 2012; pp. 27–32. [[CrossRef](#)]

43. Niu, X.-X.; Suen, C.Y. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recognit.* **2012**, *45*, 1318–1325. [[CrossRef](#)]
44. Elleuch, M.; Maalej, R.; Kherallah, M. A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition. *Procedia Comput. Sci.* **2016**, *80*, 1712–1723. [[CrossRef](#)]
45. Tao, Q.-Q.; Zhan, S.; Li, X.-H.; Kurihara, T. Robust face detection using local CNN and SVM based on kernel combination. *Neurocomputing* **2016**, *211*, 98–105. [[CrossRef](#)]
46. Wu, H.; Huang, Q.; Wang, D.; Gao, L. A CNN-SVM combined model for pattern recognition of knee motion using mechanomyography signals. *J. Electromyogr. Kinesiol.* **2018**, *42*, 136–142. [[CrossRef](#)] [[PubMed](#)]
47. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
48. Huang, G.-B.; Wang, D.H.; Lan, Y. Extreme learning machines: A survey. *Int. J. Mach. Learn. Cybern.* **2011**, *2*, 107–122. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.