informatics

MDPI

Article

Robust Multi-View Ensemble Broad Learning for Semi-Supervised Classification

Ziyang Dong [1], Mianfen Lin [1,*] and Zhiwen Yu [1,2]

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China; csdzy@mail.scut.edu.cn (Z.D.); zhwyu@scut.edu.cn (Z.Y.)
[2] Pengcheng Laboratory, Shenzhen 518066, China
* Correspondence: cslmfen@mail.scut.edu.cn

Abstract

In semi-supervised learning scenarios, the presence of limited labeled data and abundant unlabeled samples poses significant challenges to model robustness and generalization. Although the semi-supervised broad learning system (SSBLS) effectively exploits manifold structure through graph Laplacian regularization, its optimization is typically formulated under the mean square error (MSE) criterion, which is sensitive to noise and outliers. To address this limitation, this paper introduces the maximum mixture correntropy criterion (MMC) into the SSBLS framework and proposes a model termed M2C-SSBLS. By replacing the conventional MSE loss with a mixture correntropy-based objective, the proposed method enhances robustness against non-Gaussian noise and abnormal samples while preserving the computational efficiency and analytical solution property of the BLS. Furthermore, to improve representation diversity and reduce model variance, a multi-view ensemble extension, named EC-SSBLS, is proposed. This method constructs multiple feature views through a random feature subspace strategy, and independently trains an M2C-SSBLS base learner on each subspace. Finally, the predicted results of each view are fused through a voting mechanism. Experiments on benchmark UCI datasets under noise-free, 10% and 20% label noise settings demonstrate that the proposed M2C-SSBLS consistently outperforms conventional SSBLS and other advanced semi-supervised learning approaches. The ensemble extension EC-SSBLS further enhances performance, particularly in noisy environments, validating the effectiveness of combining MMC-based optimization with multi-view ensemble learning.

Keywords: semi-supervised learning; broad learning system; ensemble learning; maximum mixture correntropy criterion

Check for updates

Academic Editor: Olga Kurasova

Received: 27 February 2026
Revised: 15 May 2026
Accepted: 19 May 2026
Published: 21 May 2026

Copyright: © 2026 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.

1. Introduction

In many real-world machine learning applications, acquiring sufficient labeled data is often expensive, time-consuming, or even infeasible, whereas large amounts of unlabeled data can be obtained easily [1]. Semi-supervised learning (SSL) [2], which aims to leverage both labeled and unlabeled samples during model training, has therefore attracted increasing attention over the past decades [3]. Representative SSL methods, such as Laplacian regularized least squares [4] and Laplacian support vector machine [5], exploit the intrinsic geometric structure of data by enforcing smoothness constraints on prediction functions.

Recently, random-feature-based neural models have emerged as efficient alternatives for semi-supervised learning due to the fast training speed and competitive generalization performance. Among them, the semi-supervised broad learning system (SSBLS) [6],

built upon the broad learning system (BLS) [7], inherits the advantages of flat network architecture, analytical solution of output weights, and incremental learning capability. By incorporating graph Laplacian regularization, the SSBLS effectively exploits the manifold structure shared by labeled and unlabeled data, making it suitable for large-scale and weakly labeled learning scenarios.

Despite these advantages, the optimization of the SSBLS is typically formulated under the mean square error (MSE) criterion. Although MSE is optimal under Gaussian noise assumptions, it is well known to be sensitive to non-Gaussian noise, outliers, and label contamination [8]. As the MSE loss amplifies residual errors through quadratic penalization, mislabeled samples with large residuals tend to dominate the optimization objective. As a result, the model may excessively fit abnormal labels during training. In particular, in semi-supervised learning scenarios, mislabeled samples may propagate their errors to nearby unlabeled samples through manifold regularization, leading to performance degradation and unstable decision boundaries [9]. In addition, the graph regularization term in the SSBLS relies on the similarity relationships among samples to construct the graph Laplacian matrix. When noisy samples or incorrect neighborhood relationships exist, the graph structure itself may become corrupted, thereby weakening the effectiveness of manifold smoothness constraints. In such cases, inaccurate local structural information may propagate through the graph regularization term and negatively affect the entire learning process [10].

To address these limitations, correntropy-based learning [11] criteria have been widely investigated as robust alternatives to MSE. Correntropy is a local similarity measure defined in kernel space, which assigns smaller weights to samples with large errors and thus naturally suppresses the influence of outliers [12]. Furthermore, the maximum mixture correntropy criterion (MMC) [13], which combines multiple kernel functions with different bandwidths, provides enhanced flexibility and robustness when handling complex noise distributions. MMC has been successfully applied to semi-supervised extreme learning machines (MMC-SSELM) [14], demonstrating superior robustness against non-Gaussian noise and large outliers compared with single-kernel correntropy-based methods.

Motivated by these advances, this paper extends the maximum mixture correntropy criterion to the semi-supervised broad learning system framework and proposes a robust MMC-based SSBLS, termed M2C-SSBLS. By replacing the conventional MSE loss with the MMC-based objective, the proposed model significantly enhances robustness to noisy labels and abnormal samples while preserving the computational efficiency of the SSBLS. The output weights are optimized via a fixed-point iterative scheme, yielding a stable and effective semi-supervised learning algorithm. Furthermore, inspired by the success of multi-view learning [15,16] and ensemble methods [17], we further develop a multi-view ensemble extension of M2C-SSBLS. Specifically, multiple diverse feature subspaces are generated from the original dataset through random feature partitioning strategies. Each subspace is regarded as an independent feature view and used to train an individual M2C-SSBLS. The final prediction is obtained through a voting mechanism. This multi-view ensemble design not only increases model diversity but also further improves generalization performance and robustness. Extensive experiments conducted on benchmark UCI datasets demonstrate that the proposed M2C-SSBLS consistently outperforms the conventional SSBLS and several advanced semi-supervised learning methods in terms of classification accuracy and robustness under weakly labeled conditions.

The main contributions of this paper can be summarized as follows:

- We develop a robust M2C-SSBLS by introducing the maximum mixture correntropy criterion into the semi-supervised broad learning system.

- We further propose a multi-view ensemble learning framework, EC-SSBLS, to enhance the performance of M2C-SSBLS.
- Experimental results on benchmark datasets validate the effectiveness and robustness of the proposed method.

2. The Proposed Method

2.1. Mixture Correntropy

Correntropy has emerged as an effective information-theoretic similarity measure for robust learning, especially in the presence of non-Gaussian noise and outliers [11]. Unlike the MSE, correntropy emphasizes local similarity in kernel space and naturally suppresses the influence of large deviations. This property makes correntropy particularly attractive for learning tasks involving corrupted labels or abnormal observations.

Given two random variables U and V, correntropy is commonly defined as a kernel-based similarity measure, formulated as the expectation of a kernel function:

$$\mathcal{C}(U,V) = \mathbb{E}\big[\kappa(U,V)\big]. \tag{1}$$

where $\kappa(U,V)$ denotes a kernel function. When using a single kernel function, correntropy can provide robustness through local similarity mechanisms. However, its performance is sensitive to the kernel scale, which controls the locality of the similarity measure. A single scale often struggles to adapt to different levels of noise distribution simultaneously.

To alleviate the limitation of a single kernel correntropy, researchers proposed the mixture correntropy [18], which constructs a more flexible similarity measure by weighting and combining multiple kernel functions. Compared with a single kernel function, mixture correntropy can more comprehensively characterize the error characteristics, thereby significantly improving the model's adaptability to complex noise environments [19]. Formally, the mixture correntropy can be expressed as

$$\mathcal{M}(U,V) = \frac{1}{N}\sum_{k=1}^{K} \omega_k \kappa_k(U,V) \tag{2}$$

where $\{\kappa_k(\cdot)\}_{k=1}^{K}$ denotes a set of kernel-induced similarity functions with different shape parameters, ω_k are the mixture weights satisfying $\sum_{k=1}^{K}\omega_k = 1$, and K is the number of kernels. By maximizing $\mathcal{M}(r)$, the learning process encourages small residuals while effectively down-weighting samples associated with large or irregular errors.

Compared with single-kernel correntropy, mixture correntropy exhibits enhanced flexibility and robustness, as it can simultaneously accommodate fine-grained local errors and broader deviation patterns. This advantage has motivated its application in various robust learning models [19,20], where MMC-based optimization consistently outperforms methods based on MSE or single correntropy in noise and outlier interference scenarios.

2.2. SSBLS

The Broad Learning System (BLS) [7] is a learning model that emphasizes network width expansion rather than depth. The basic idea is to construct feature mapping nodes and enhancement nodes, and to solve the output weights in an analytical form, thereby avoiding the complex parameter tuning and iterative training procedures commonly required in deep neural networks. Due to its simple structure, fast training speed, and good generalization performance, the BLS has been widely used in small and medium-sized data modeling tasks [21–23].

In the classical BLS framework, assuming the input data is $\mathbf{X}$, the input data are first transformed into feature nodes. The i-th group of feature maps can be formulated as

$$\mathbf{Z}_i = \phi\left(\mathbf{X}\mathbf{W}_i^{(f)} + \mathbf{b}_i^{(f)}\right) \tag{3}$$

where $\mathbf{W}_i^{(f)}$ and $\mathbf{b}_i^{(f)}$ denote the weight matrix and bias of the feature mapping layer, respectively, and $\phi(\cdot)$ represents the activation function. By concatenating all feature mapping nodes, the overall output of the feature layer is obtained as

$$\mathbf{Z} = \left[\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_{N_f}\right] \tag{4}$$

In order to further enhance the non-linear expression ability of the model, the BLS further generated enhancement nodes. The j-th enhanced node can be represented as

$$\mathbf{H}_j = \xi\left(\mathbf{Z}\mathbf{W}_j^{(e)} + \mathbf{b}_j^{(e)}\right) \tag{5}$$

where $\mathbf{W}_j^{(e)}$ and $\mathbf{b}_j^{(e)}$ denote the randomly generated weight matrix and bias vector of the enhancement layer, respectively, and $\xi(\cdot)$ represents the activation function. By concatenating all enhancement nodes, the overall output of the enhancement layer is obtained as

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_{N_e}] \tag{6}$$

Finally, the outputs of the feature layer and the enhancement layer are concatenated to form the intermediate representation of the BLS, given by

$$\mathbf{A} = [\mathbf{Z} \mid \mathbf{H}] \tag{7}$$

The model output is then produced through a linear mapping:

$$\mathbf{Y} = \mathbf{A}\mathbf{W} \tag{8}$$

where $\mathbf{W}$ denotes the output weight matrix to be learned, which is typically obtained by solving a regularized least squares problem.

However, the conventional BLS relies heavily on sufficiently labeled data for supervised training, and its performance is often significantly degraded in scenarios where labeled samples are scarce or partially missing. To address this limitation, semi-supervised learning principles have been incorporated into the BLS framework, leading to the development of the Semi-Supervised Broad Learning System (SSBLS), which aims to exploit both a small number of labeled samples and a large amount of unlabeled data to enhance model generalization capability.

The fundamental idea of the SSBLS is based on the smoothness assumption commonly adopted in semi-supervised learning, which states that samples that are close to each other in the feature space should yield similar predictive outputs. This assumption is typically modeled through graph-based regularization. Specifically, consider a dataset consisting of l labeled samples and u unlabeled samples, with the total number of samples given by $n = l + u$. Manifold regularization is typically represented as

$$\mathcal{R}_{\text{manifold}}(f) = \frac{1}{2}\sum_{i,j} S_{ij}\left\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\right\|^2, \tag{9}$$

where $f(\mathbf{x}_j)$ is the model output. $\mathbf{S}$ is a similarity matrix used to measure the pairwise similarity between samples. For samples $\mathbf{x}_i$ and $\mathbf{x}_j$, the similarity weight is defined as

$$S_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\tau^2}\right), & \text{if } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \text{ or } \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where τ is the bandwidth parameter and $\mathcal{N}_k(\mathbf{x}_i)$ denotes the set of k nearest neighbors of $\mathbf{x}_i$.

Equation (9) is commonly rewritten as

$$\mathcal{R}_{\text{manifold}}(f) = \text{Tr}\left(\mathbf{F}^\top \mathbf{L}\mathbf{F}\right), \mathbf{F} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)]^\top. \tag{11}$$

$\text{Tr}(\cdot)$ represents the trace of a matrix. $\mathbf{L}$ is the graph Laplacian matrix constructed from sample similarity relationships. The graph Laplacian matrix is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{S}. \tag{12}$$

where $\mathbf{D}$ is constructed as a diagonal matrix with entries $D_{ii} = \sum_j^{u+l} S_{ij}$.

On this basis, the objective function of the SSBLS generally consists of three components: the structural regularization term of output weights, the empirical loss term defined over the labeled samples, and the smoothness regularization term that characterizes the manifold structure of both labeled and unlabeled data. The objective function of the SSBLS can be formulated as follows:

$$\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W}\|_F^2 + \frac{\alpha}{2}\|\mathbf{Y}_l - \mathbf{A}_l\mathbf{W}\|_F^2 + \frac{\beta}{2}\text{Tr}\left(\mathbf{W}^\top \mathbf{A}^\top \mathbf{L}\mathbf{A}\mathbf{W}\right) \tag{13}$$

where $\mathbf{A}_l$ and $\mathbf{Y}_l$ denote the intermediate-layer output matrix and the label matrix, respectively, and α and β are regularization parameters that control the trade-off among different terms.

By solving the above objective function, a closed-form solution for the output weights can be obtained. This allows the structural information carried by unlabeled samples to be incorporated into the training process while preserving the computational efficiency of the BLS. As a result, the SSBLS exhibits superior generalization performance compared with the conventional BLS in scenarios with insufficient labeled data or missing labels [6,24].

Nevertheless, the SSBLS based on the MSE criterion still suffers from limited robustness when confronted with noisy labels and outlier samples, and it typically relies on a single feature view for modeling. These limitations may hinder its effectiveness in complex noise environments or in scenarios with restricted feature representations. Consequently, further investigation is required to enhance the robustness and representational capacity of the SSBLS framework.

2.3. M2C-SSBLS

Although the SSBLS is capable of exploiting unlabeled samples through manifold regularization, its learning objective is typically constructed based on the mean square error (MSE) criterion. While MSE is optimal under Gaussian noise assumptions, its performance degrades significantly in the presence of non-Gaussian noise, noisy labels, or outlier samples, which are commonly encountered in real-world datasets. As a result, the robustness of the conventional SSBLS remains limited in complex learning environments.

To overcome this limitation, we introduce the maximum mixture correntropy criterion into the SSBLS framework, called M2C-SSBLS. MMC is a similarity measure that extends conventional correntropy by combining multiple kernel functions, thereby providing enhanced robustness against noise and outliers. Compared with single-kernel correntropy or MSE-based loss functions, MMC can effectively reduce the impact of outliers or noise

on the model. Specifically, based on Equation (2), given a training dataset with N samples $\mathcal{D}_N = \{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, where $x_i \in R^d$ and $y_i \in R^c$ represent the input data and the associated target output, we use Gaussian kernel as the kernel function. It should be noted that variables U and V in Equation (1) are generic random variables used for the theoretical definition of correntropy, whereas $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{x}_i$, and $\mathbf{y}_i$ are adopted to represent the practical dataset samples and prediction outputs in the proposed learning model. Equation (2) can be rewritten as

$$\mathcal{M}(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \sum_{k=1}^{K} \omega_k G_{\sigma_k}(\mathbf{X}, \mathbf{Y}) \tag{14}$$

where $\{G_{\sigma_k}(\cdot)\}_{k=1}^K$ denotes K Gaussian kernels with different kernel bandwidths σ_k, and $\{\omega_k\}_{k=1}^K$ are the corresponding kernel weights satisfying $\sum_{k=1}^K \omega_k = 1$. The Gaussian kernel function is defined as

$$G_{\sigma_k}(x, y) = \exp\left(-\frac{\|x - y\|_2^2}{2\sigma_k^2}\right) \tag{15}$$

The maximum mixture correntropy criterion-based loss [25] can be formalized as

$$\begin{aligned} \widehat{\mathcal{O}}_{\text{mix}} &= 1 - \mathcal{M}(\mathbf{X}, \mathbf{Y}) \\ &= 1 - \frac{1}{N} \sum_{k=1}^{K} \left[\sum_{i=1}^{N} \omega_k \, G_{\sigma_k}(\mathbf{X}, \mathbf{Y}) \right] \end{aligned} \tag{16}$$

When the prediction error is small, the Gaussian kernel output approaches 1, resulting in a small loss; when the error becomes large, the kernel value decays rapidly, thereby automatically suppressing the influence of outliers. Based on Equation (15), minimizing the prediction error is equivalent to maximizing $\mathcal{M}(\mathbf{X}, \mathbf{Y})$, which further corresponds to minimizing the objective function $\widehat{\mathcal{O}}_{\text{mix}}$. Based on this, we introduce MMC into the SSBLS, and the Equation (13) can be rewritten as

$$\min_{\mathbf{W}} \ \frac{1}{2}\|\mathbf{W}\|_F^2 - \frac{C}{2N} \sum_{k=1}^{K} \left[\sum_{i=1}^{N} \omega_k \, G_{\sigma_k}(y_i, a_i \mathbf{W}) \right] + \frac{\beta}{2} \operatorname{Tr}\left(\mathbf{W}^\top \mathbf{A}^\top \mathbf{L} \mathbf{A} \mathbf{W} \right) \tag{17}$$

The first term $\frac{1}{2}\|\mathbf{W}\|_F^2$ is a structural regularization term. By penalizing excessively large weight values, this regularization term limits the model capacity and suppresses overly complex decision boundaries, thereby controlling the model complexity and preventing overfitting. N denotes the number of labeled samples, and K is the number of Gaussian kernels used in the mixture correntropy formulation. α and β are parameters that balance the influence of the correntropy-based loss term and the manifold regularization term, respectively. ω_k denotes the weight of the k-th Gaussian kernel. $G_{\sigma_k}(y_i, a_i \mathbf{W})$ measures the similarity between the true label y_i and the predicted output $a_i \mathbf{W}$ for the i-th labeled sample with bandwidths σ_k.

We optimized Equation (17) by calculating the gradient with respect to $\mathbf{W}$, so the optimized objective is as follows:

$$\mathbf{W} = \left(\mathbf{A}^\top \mathbf{P} \mathbf{A} + \frac{2N}{C} + \frac{2N\beta}{C} \mathbf{A}^\top \mathbf{L} \mathbf{A} \right)^{-1} \mathbf{A}^\top \mathbf{P} \mathbf{Y}. \tag{18}$$

where $\mathbf{P}$ is a diagonal matrix, with the first l diagonal elements being $[\mathbf{P}]_{ii} = \bar{\phi}(\epsilon_i)$. And

$$\phi(\epsilon_i) = \sum_{k=1}^{K} \omega_k e^{\left(-\frac{\epsilon_i^2}{2\sigma_k^2}\right)} \tag{19}$$

while the remaining diagonal elements are set to 0. Here, ϵ_i denotes the residual of the i-th labeled sample. The structural diagram of M2C-SSBLS is shown in Figure 1. The pseudocode of the M2C-SSBLS algorithm is shown in Algorithm 1.

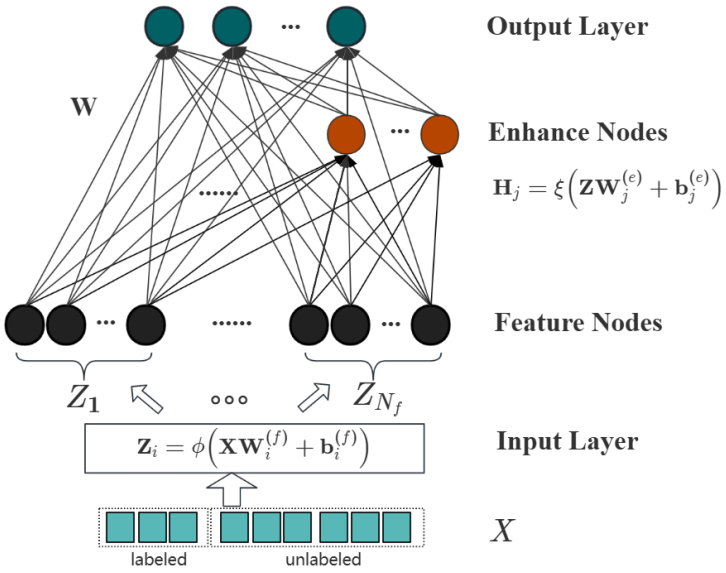


Figure 1. The framework of M2C-SSBLS.

Algorithm 1 M2C-SSBLS: Maximum Mixture Correntropy based Semi-Supervised BLS.

Input:

1. Labeled set $\mathcal{D}_l = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$
2. Unlabeled set $\mathcal{D}_u = \{\mathbf{x}_j\}_{j=1}^{u}$
3. Numbers of feature nodes N_f and enhancement nodes N_e
4. Activation functions $\phi(\cdot)$ and $\xi(\cdot)$
5. Mixture kernel number K, bandwidths $\{\sigma_k\}_{k=1}^{K}$ and weights $\{\omega_k\}_{k=1}^{K}$
6. Regularization parameters C and β
7. Maximum iterations T and tolerance ϵ

Training:

1. Calculate the graph Laplacian $\mathbf{L}$ through Equations (10) and (12).
2. Randomly generate the mapping parameters of the feature nodes and enhancement nodes.
3. Construct the feature layer $\mathbf{Z}$ and the enhancement layer $\mathbf{H}$ using Equations (4) and (6), then form the feature matrix $\mathbf{A} = [\mathbf{Z} \,|\, \mathbf{H}]$.
4. **For** $t = 1, \ldots, T$:

   Compute residuals for labeled samples: $\epsilon_i^{(t)} = \mathbf{y}_i - \mathbf{a}_i \mathbf{W}^{(t)}$, $i = 1, \ldots, N$,
   Form the diagonal matrix $\mathbf{P}^{(t)} = \mathrm{diag}(\phi(\epsilon_1)^{(t)}, \ldots, \phi(\epsilon_N)^{(t)})$.
   Update the output weights $\mathbf{W}^{(t)}$ by Equation (18)
   break if $|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}| < \varepsilon$.

2.4. EC-SSBLS

Although M2C-SSBLS significantly improves robustness under noisy environments by incorporating the maximum mixture correntropy criterion, relying on a single feature space is often insufficient to fully capture the underlying distribution characteristics of complex data [26]. To further enhance generalization performance and reduce model instability, we extend M2C-SSBLS by introducing a random feature subspace strategy and construct a multi-view ensemble framework termed EC-SSBLS.

Let the original input data matrix be

$$\mathbf{X} \in \mathbb{R}^{n \times d} \tag{20}$$

To increase the diversity of the ensemble classifiers, distinct feature spaces are generated for different base learners. Specifically, multiple feature subspaces are constructed via random feature sampling without replacement. For the m-th feature view, a subset of features is randomly selected from the original feature set to form a subspace representation.

$$\mathbf{X}^{(m)} \in \mathbb{R}^{n \times d_M} \tag{21}$$

where $d_M < d$. This random subspace strategy preserves the statistical properties of the original features while introducing structural diversity across different views. Compared with random projection or feature transformation approaches, random feature sampling does not alter the physical meaning or geometric structure of the features [3]. This property is suitable for semi-supervised learning scenarios involving graph Laplacian regularization, where preserving the intrinsic data geometry is essential.

An independent M2C-SSBLS model is trained on each feature subspace. Let the intermediate-layer output corresponding to the m-th view be denoted as

$$\mathbf{A}^{(m)} = \left[\mathbf{Z}^{(m)} \mid \mathbf{H}^{(m)}\right] \tag{22}$$

where $\mathbf{Z}^{(m)}$ represents the output of the feature mapping layer and $\mathbf{H}^{(m)}$ denotes the output of the enhancement layer. The corresponding output weight matrix is obtained in closed form as

$$\mathbf{W}^{(m)} = \left(\mathbf{A}^{(m)\top}\mathbf{P}^{(m)}\mathbf{A}^{(m)} + \frac{2N}{C}\mathbf{I} + \frac{2N\beta}{C}\mathbf{A}^{(m)\top}\mathbf{L}\mathbf{A}^{(m)}\right)^{-1}\mathbf{A}^{(m)\top}\mathbf{P}^{(m)}\mathbf{Y}. \tag{23}$$

In this manner, each base learner simultaneously benefits from the robustness induced by the MMC criterion and the manifold smoothness constraint introduced by graph regularization when learning different feature representations.

Let the prediction result of the m-th base learner be

$$\hat{\mathbf{Y}}^{(m)} = \mathbf{A}^{(m)}\mathbf{W}^{(m)} \tag{24}$$

EC-SSBLS fuses the outputs of all base learners through a voting mechanism:

$$\hat{y} = \mathrm{mode}\left(\hat{y}^{(1)}, \ldots, \hat{y}^{(m)}\right) \tag{25}$$

This ensemble mechanism effectively reduces model variance and improves overall stability by integrating complementary decision boundaries learned from different feature views.

By integrating complementary decision boundaries learned from different feature views, the ensemble mechanism effectively reduces model variance and enhances overall stability. Through the combination of correntropy-based robustness, graph-structured semi-supervised regularization, and random subspace diversity, EC-SSBLS exhibits stronger generalization ability and stability in environments with noisy labels, scarce annotations. The structural diagram of EC-SSBLS is shown in Figure 2. The pseudocode of the EC-SSBLS algorithm is shown in Algorithm 2.

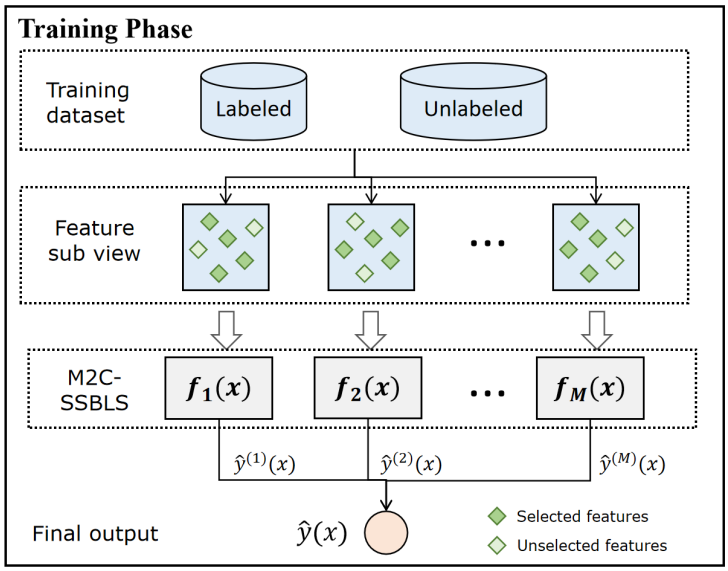


Figure 2. The framework of EC-SSBLS.

Algorithm 2 EC-SSBLS.

Multi-view ensemble of M2C-SSBLS. **Input**:

1. Training labeled set $\mathcal{D}_l = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$
2. Unlabeled set $\mathcal{D}_u = \{\mathbf{x}_j\}_{j=1}^{u}$
3. M: Number of views (base learners)
4. d_M: Number of selected features dimensions
5. Kernel bandwidth parameters $\{\omega_k, \sigma_k\}_{k=1}^{K}$
6. Regularization parameters C and β
7. Maximum iterations T and tolerance ε.

Training:
For $m = 1$ **to** M **do**

1. Randomly sample a feature index set $\mathcal{I}_m \subset \{1, \ldots, d\}$ with $|\mathcal{I}_m| = d_M$, and form the m-th view data $\mathbf{X}^{(m)}$.
2. Calculate the intermediate representation $\mathbf{A}^{(m)} = [\mathbf{Z}^{(m)} \mid \mathbf{H}^{(m)}]$ using Equation (22).
3. Train a base learner f_m through Algorithm 1.
4. Save the m-th base learner as $(\mathcal{I}_m, \mathbf{W}^{(m)})$.

End For
Testing:
Utilize Equations (24) and (25) to determine the class of the test sample.

3. Experiment

3.1. Experimental Setting

To evaluate the effectiveness and stability of the proposed method, extensive experiments were conducted on several publicly available benchmark datasets. All datasets were obtained from the UCI machine learning repository, and their detailed descriptions are summarized in Table 1. In the training set, the labeled and unlabeled samples were partitioned with a ratio of 3:7 to assess the learning capability of the model under limited annotation conditions. The experiment is implemented based on Scikit-learn on an Intel Core i7-8550U CPU at 1.80 GHz.

To simulate the influence of label noise and data perturbations encountered in real-world scenarios, experiments were carried out under different noise settings, including (1) the original data without noise; (2) 10% proportion label noise scene; (3) 20% proportion

label noise scene. Due to the high computational complexity on large-scale datasets, only the 10% noise scenario was evaluated for the last four datasets.

Table 1. Summary of datasets used in the experiments.

Dataset	Training Data	Testing Data	Features	Class
Breast Cancer	300	269	30	2
Heart disease	100	170	13	2
wine	100	78	13	3
COIL20	860	580	1024	20
Diabetic	806	345	19	2
g50c	440	110	50	2
Protien	949	534	56	10
abalone	2088	2089	8	2
BASEHOCK	1195	798	4862	2
Cardiotocography	1701	425	21	3
PCMAC	1166	777	3289	2

To evaluate the performance of the model, this article uses classification accuracy as the main evaluation metric and reports the average results of 5 repeated experiments. To comprehensively validate the superiority of the proposed method, several representative semi-supervised learning approaches were selected as baseline methods for comparison, including LapSVM [5], SSBLS [6], RC-SSELM [27], and CC-SSBLS [28]. The parameters of all competing methods were configured according to the settings recommended in their original research.

3.2. Results

Table 2 presents the classification accuracy of all methods under different datasets and noise settings. The experimental results indicate that, under the original (noise-free) condition, M2C-SSBLS achieves superior classification performance compared with existing approaches on most datasets. The advantage becomes more pronounced in the presence of label noise. These results demonstrate the effectiveness of incorporating the MMC criterion into the SSBLS framework. Furthermore, EC-SSBLS attains higher classification accuracy than the single M2C-SSBLS model on the majority of datasets. This suggests that constructing multi-view representations through random feature subspaces effectively enhances the expressive capability of the model. The complementary information learned from different subspaces contributes to refining the overall decision boundary, thereby reducing model variance and improving classification stability. Overall, under the noise-free condition, both M2C-SSBLS and EC-SSBLS exhibit strong learning capability. Among them, EC-SSBLS further improves performance through the ensemble mechanism, validating the effectiveness of the proposed multi-view ensemble strategy.

Under the 10% and 20% label noise condition, MSE-based semi-supervised learning methods generally suffer noticeable performance degradation, indicating that traditional loss functions are sensitive to outliers. In contrast, M2C-SSBLS maintains relatively high classification accuracy across most datasets, with significantly smaller performance degradation compared to other methods. This highlights the advantage of the maximum mixture correntropy criterion in suppressing abnormal errors. By modeling residuals through a multi-kernel weighted formulation, MMC assigns lower weights to large residual samples, thereby reducing the adverse impact of noisy labels on model training and enhancing robustness. On this basis, EC-SSBLS further outperforms the single M2C-SSBLS model under noisy conditions. By integrating the learning results from multiple feature subspaces, EC-SSBLS can reduce the instability caused by noisy samples in a single view and leverages

complementary information across views to improve overall classification performance. Consequently, EC-SSBLS demonstrates stronger stability and generalization capability in noisy environments.

Table 2. Accuracy results (%) under different noise settings.

Dataset	Ratio	LapSVM	CC-SSBLS	SSBLS	RC-SSELM	M2C-SSBLS	EC-SSBLS
Breast Cancer	0%	90.09 ± 1.83	91.95 ± 0.53	91.15 ± 1.19	91.97 ± 1.22	92.25 ± 0.89	**92.56 ± 0.71**
	10%	88.48 ± 0.98	90.86 ± 0.76	89.59 ± 0.53	89.89 ± 1.47	91.14 ± 0.56	**91.42 ± 0.64**
	20%	85.11 ± 0.37	89.77 ± 0.82	89.14 ± 0.51	89.10 ± 0.82	90.88 ± 1.23	**91.16 ± 0.70**
Heart disease	0%	81.18 ± 3.28	83.65 ± 0.49	82.47 ± 0.26	82.85 ± 0.87	86.48 ± 1.23	**87.23 ± 0.36**
	10%	80.78 ± 2.78	83.53 ± 0.72	79.88 ± 0.77	80.98 ± 0.95	84.76 ± 1.01	**86.45 ± 0.49**
	20%	79.92 ± 3.09	82.69 ± 0.58	78.14 ± 1.81	81.17 ± 1.42	83.66 ± 0.83	**85.19 ± 0.57**
wine	0%	94.87 ± 3.39	98.93 ± 1.22	98.72 ± 0.00	98.72 ± 0.00	98.48 ± 0.74	**99.12 ± 0.62**
	10%	95.30 ± 3.23	97.88 ± 1.78	97.44 ± 1.28	97.69 ± 1.07	98.12 ± 0.51	**98.44 ± 0.78**
	20%	91.15 ± 2.09	94.37 ± 1.08	92.44 ± 1.51	93.88 ± 2.05	96.57 ± 0.84	**97.66 ± 1.20**
COIL20	0%	96.15 ± 0.98	95.20 ± 0.31	96.17 ± 0.56	96.31 ± 0.34	96.73 ± 0.40	**97.45 ± 0.08**
	10%	93.45 ± 0.20	93.72 ± 0.76	92.07 ± 0.40	93.72 ± 0.26	95.21 ± 0.33	**95.24 ± 0.23**
	20%	86.15 ± 0.87	88.37 ± 1.08	87.44 ± 1.51	87.71 ± 0.33	89.15 ± 0.40	**90.14 ± 0.50**
Diabetic	0%	70.72 ± 1.61	72.53 ± 1.30	72.35 ± 1.02	72.58 ± 0.60	73.18 ± 0.56	**76.83 ± 1.23**
	10%	67.25 ± 1.09	69.76 ± 0.70	69.04 ± 0.43	69.04 ± 0.56	70.22 ± 0.68	**72.12 ± 1.16**
	20%	62.26 ± 1.44	64.76 ± 0.87	63.56 ± 0.99	63.41 ± 0.58	67.71 ± 0.44	**69.54 ± 0.87**
g50c	0%	96.97 ± 0.52	96.85 ± 1.29	96.91 ± 0.50	96.73 ± 1.52	**97.64 ± 0.73**	96.99 ± 1.50
	10%	96.67 ± 0.52	96.66 ± 0.84	96.36 ± 1.57	97.09 ± 0.41	**97.31 ± 0.87**	95.74 ± 0.82
	20%	94.31 ± 2.13	95.39 ± 0.97	94.87 ± 0.69	94.09 ± 1.35	**96.88 ± 0.74**	95.51 ± 1.01
Protein	0%	89.95 ± 1.20	89.61 ± 1.83	89.59 ± 1.51	89.89 ± 0.70	90.58 ± 1.34	**91.08 ± 0.47**
	10%	86.14 ± 0.65	88.15 ± 1.69	86.70 ± 0.30	86.85 ± 0.48	**89.15 ± 0.97**	88.98 ± 0.51
	20%	82.19 ± 1.44	85.62 ± 0.88	85.50 ± 1.15	84.99 ± 0.79	86.69 ± 0.74	**87.23 ± 0.46**
abalone	10%	83.25 ± 0.67	85.77 ± 1.28	82.46 ± 0.88	84.77 ± 0.11	85.99 ± 0.49	**87.12 ± 0.63**
BASEHOCK	10%	59.87 ± 1.67	61.64 ± 0.52	57.83 ± 1.25	58.95 ± 1.48	65.80 ± 1.55	**68.47 ± 0.33**
Cardiotocography	10%	80.19 ± 1.05	84.75 ± 0.58	81.54 ± 0.36	82.04 ± 0.81	**85.68 ± 0.51**	85.53 ± 0.29
PCMAC	10%	54.12 ± 0.94	57.33 ± 0.43	55.89 ± 1.75	56.26 ± 0.93	57.91 ± 1.76	**57.98 ± 0.78**

3.3. Parameter Analysis

To further analyze the impact of key hyperparameters on performance of the model, this section conducts sensitivity experiments on key parameters in M2C-SSBLS and EC-SSBLS through grid search. The search scope is shown in Table 3.

Table 3. Parameter selection range of M2C-SSBLS and EC-SSBLS.

Parameter	Range
N_f and N_e	[50, 100, 300, 500]
C and β	$[10^{-5}, 10^{-3}, 10^{-1}, 10^{1}, 10^{2}, 10^{3}]$
σ_1	$[10^{-3}, 10^{-1}, 10^{1}, 10^{2}]$
σ_2	[1, 3, 5, 7, 9]
M	[1, 3, 5, 7, 9]
$r(d_m/d)$	[0.95, 0.9, 0.85, 0.8, 0.7, 0.6, 0.5]

Specifically, for the number of base learners M, as shown in Figure 3. As the number of base learners M increases, the model performance gradually improves. However, when M exceeds a certain value, the performance improvement tends to stabilize, while the

computational cost significantly increases. Considering comprehensive performance and efficiency, this article sets M to 5.

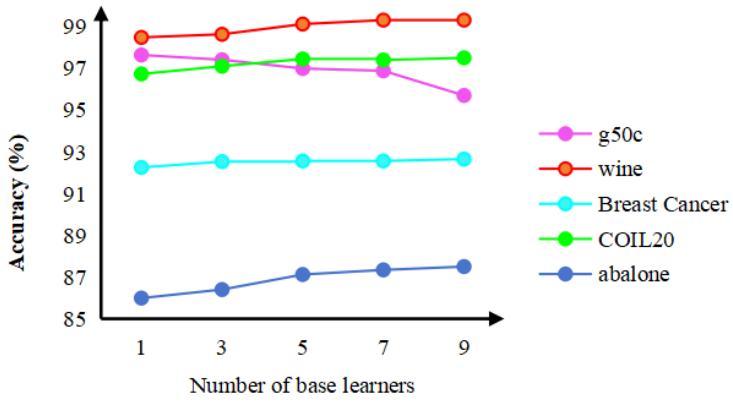


Figure 3. The impact of the number of classifiers.

In our experiments, for datasets with dimensions exceeding 1000, PCA was first applied to extract the most informative components before performing the random feature selection procedure. For the subspace size d_m, different settings are adopted depending on the feature dimensionality. For low-dimensional datasets, where the number of features is limited and each feature typically carries substantial information, setting a small subspace size may lead to insufficient representation and potential underfitting. In contrast, high-dimensional datasets often contain considerable feature redundancy. Appropriately reducing the subspace size can enhance the diversity among base learners and improve the effectiveness of the ensemble model. Accordingly, for datasets with fewer than 100 dimensions, the feature retention ratio r ($r = d_m/d$) is set to 0.85 of the original feature dimension. For datasets with more than 100 features, r is set to 0.70 of the original feature dimension.

4. Discussion

The proposed M2C-SSBLS and EC-SSBLS are mainly designed for scenarios involving limited labeled samples, noisy labels, and complex feature distributions. Experimental results demonstrate that the proposed methods achieve superior robustness and generalization performance compared with conventional methods based on the MSE criterion. First, the proposed methods are particularly advantageous in noisy environments containing corrupted labels or outliers. Traditional SSL methods based on the MSE criterion are generally sensitive to large residual samples and can be significantly affected by noisy labels or abnormal observations, leading to performance degradation. In contrast, M2C-SSBLS introduces the MCC criterion, which models residual errors through a multi-kernel weighting mechanism and automatically suppresses the influence of large-error samples during optimization. Therefore, in scenarios where label quality is unreliable or data contains random noise and outliers, M2C-SSBLS usually performs better than conventional methods. Second, EC-SSBLS is more suitable for datasets with complex feature distributions or strong feature redundancy. Conventional single-model approaches rely on a single feature representation and may suffer from local feature bias or redundant information. In contrast, EC-SSBLS constructs multiple feature views using a random subspace strategy and integrates the predictions from different subspaces through ensemble learning. This mechanism enhances classifier diversity and reduces model variance. In addition, as the BLS obtains output weights in an analytical manner without relying on complicated backpropagation and iterative optimization, the proposed framework requires less computational cost. Nevertheless, several limitations still exist. First, the performance of MMC depends on the selection of kernel bandwidth parameters, which may require dataset-specific tuning. Second, although

EC-SSBLS improves performance through ensemble learning, integrating multiple base learners inevitably increases computational overhead. Finally, the current framework is mainly designed for static datasets, and its extension to online learning or streaming data scenarios remains a potential direction for future work.

5. Conclusions

This paper proposed a semi-supervised learning framework by integrating the MMC criterion into the BLS. The proposed M2C-SSBLS effectively alleviates the sensitivity of conventional SSBLS to noise and outliers by adopting a mixture correntropy-based optimization objective. To further improve representation capacity and stability, a multi-view ensemble extension, EC-SSBLS, was developed by incorporating a random feature subspace strategy. By training multiple M2C-SSBLS models on diverse feature views and aggregating their predictions via voting, the ensemble framework reduces model variance and leverages complementary information across subspaces. The experimental results show that M2C-SSBLS outperforms the compared methods in both noise-free and noisy scenarios, while EC-SSBLS further improves classification accuracy. Future work will investigate adaptive kernel weighting strategies, dynamic view generation mechanisms, and incremental extensions for streaming data scenarios to further enhance adaptability and scalability.

Author Contributions: Conceptualization, Z.D. and M.L.; methodology, Z.D. and M.L.; software, Z.D.; validation, Z.D.; writing—original draft preparation, Z.D.; writing—review and editing, Z.Y. and M.L.; visualization, Z.D.; supervision, Z.Y.; project administration, Z.Y.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was sponsored by National Natural Science Foundation of China No. 62572199, 92467109, U21A20478, and in part by the Major Key Project of PCL (Grant No. PCL2025A11 and No. PCL2025A13), in part by the National Key R&D Program of China 2023YFA1011601.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, G.; Fan, Z.; Yu, Z.; Yang, K.; Chen, C.P. Semi-Supervised Ensemble Classifier Based on Distance Constraint for High-Dimensional Data. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *56*, 724–736. [CrossRef]
2. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
3. Li, G.; Yu, Z.; Yang, K.; Lin, M.; Chen, C.P. Exploring feature selection with limited labels: A comprehensive survey of semi-supervised and unsupervised approaches. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6124–6144. [CrossRef]
4. Chen, J.; Wang, C.; Sun, Y.; Shen, X.S. Semi-supervised Laplacian regularized least squares algorithm for localization in wireless sensor networks. *Comput. Netw.* **2011**, *55*, 2481–2491. [CrossRef]
5. Melacci, S.; Belkin, M. Laplacian Support Vector Machines Trained in the Primal. *J. Mach. Learn. Res.* **2011**, *12*, 1149–1184.
6. Zhao, H.; Zheng, J.; Deng, W.; Song, Y. Semi-supervised broad learning system based on manifold regularization and broad network. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2020**, *67*, 983–994. [CrossRef]
7. Chen, C.P.; Liu, Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 10–24. [CrossRef] [PubMed]
8. Lin, M.; Yang, K.; Yu, Z.; Shi, Y.; Chen, C.P. Hybrid ensemble broad learning system for network intrusion detection. *IEEE Trans. Ind. Inform.* **2023**, *20*, 5622–5633. [CrossRef]
9. Yang, X.; Song, Z.; King, I.; Xu, Z. A survey on deep semi-supervised learning. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8934–8954. [CrossRef]
10. Jin, W.; Ma, Y.; Liu, X.; Tang, X.; Wang, S.; Tang, J. Graph structure learning for robust graph neural networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 66–74.

11. Liangjun, C.; Honeine, P.; Hua, Q.; Jihong, Z.; Xia, S. Correntropy-based robust multilayer extreme learning machines. *Pattern Recognit.* **2018**, *84*, 357–370. [CrossRef]
12. Ren, Z.; Yang, L. Correntropy-based robust extreme learning machine for classification. *Neurocomputing* **2018**, *313*, 74–84. [CrossRef]
13. Wang, T.; Cao, J.; Dai, H.; Lei, B.; Zeng, H. Robust maximum mixture correntropy criterion based one-class classification algorithm. *IEEE Intell. Syst.* **2021**, *37*, 69–78. [CrossRef]
14. Yang, J.; Cao, J.; Xue, A. Robust maximum mixture correntropy criterion-based semi-supervised ELM with variable center. *IEEE Trans. Circuits Syst. II Express Briefs* **2020**, *67*, 3572–3576. [CrossRef]
15. Yu, Z.; Dong, Z.; Yu, C.; Yang, K.; Fan, Z.; Chen, C.P. A review on multi-view learning. *Front. Comput. Sci.* **2025**, *19*, 197334. [CrossRef]
16. Zhao, J.; Xie, X.; Xu, X.; Sun, S. Multi-view learning overview: Recent progress and new challenges. *Inf. Fusion* **2017**, *38*, 43–54. [CrossRef]
17. Lin, M.; Yu, Z.; Yang, K.; Chen, C.P. Hybrid ensemble framework for imbalanced data streams with concept drift. *IEEE Trans. Big Data* **2025**, *11*, 3430–3442. [CrossRef]
18. Chen, B.; Wang, X.; Lu, N.; Wang, S.; Cao, J.; Qin, J. Mixture correntropy for robust learning. *Pattern Recognit.* **2018**, *79*, 318–327. [CrossRef]
19. Lu, M.; Xing, L.; Zheng, N.; Chen, B. Robust sparse channel estimation based on maximum mixture correntropy criterion. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–6.
20. Zheng, Y.; Chen, B.; Wang, S.; Wang, W.; Qin, W. Mixture correntropy-based kernel extreme learning machines. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *33*, 811–825. [CrossRef]
21. Li, G.; Yu, Z.; Yang, K.; Fan, Z.; Chen, C.P. Incremental Semisupervised Learning with Adaptive Locality Preservation for High-Dimensional Data. *IEEE Trans. Artif. Intell.* **2025**, *6*, 2990–3004. [CrossRef]
22. Lin, M.; Yu, Z.; Yang, K.; Chen, C.P. Dynamic Chunk-Based Active Learning Based on Enhanced Broad Learning System for Imbalanced Drifting Data Streams. *IEEE Trans. Knowl. Data Eng.* **2025**, *38*, 997–1010. [CrossRef]
23. Chen, W.; Yang, K.; Yu, Z.; Nie, F.; Chen, C.P. Adaptive broad network with graph-fuzzy embedding for imbalanced noise data. *IEEE Trans. Fuzzy Syst.* **2025**, *33*, 1949–1962. [CrossRef]
24. Pu, X.; Li, C. Online semisupervised broad learning system for industrial fault diagnosis. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6644–6654. [CrossRef]
25. Jia, W.; Li, X.; Bi, D.; Xie, Y. Maximum mixture correntropy based Student-t kernel adaptive filtering for indoor positioning of internet of Things. *Inf. Sci.* **2025**, *696*, 121729. [CrossRef]
26. Wang, J.; Luo, S.-w.; Zeng, X.-h. A random subspace method for co-training. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 195–200.
27. Yang, J.; Cao, J.; Wang, T.; Xue, A.; Chen, B. Regularized correntropy criterion based semi-supervised ELM. *Neural Netw.* **2020**, *122*, 117–129. [CrossRef] [PubMed]
28. Wang, L.; Wang, Y.; Chen, J.; Zhang, S.; Zhang, L. Research on CC-SSBLS model-based air quality index prediction. *Atmosphere* **2024**, *15*, 613. [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.