

Learnable Unified Structure under Global-Local Consistency for Unsupervised Feature Selection

Guoqiu Wen, Xinjie Han, Peng Liu*, Jiayang Su, Yuanpeng Li, Jiale Liu,
Shuyu Li, Cong Lei, and Qinghong Jiang*

Abstract: Previous unsupervised feature selection methods typically model global data distributions and local geometric structures separately or via parallel constraints. Such decoupled modeling often leads to structural inconsistencies due to the lack of a unified representation of sample relationships. Moreover, existing local structure-based methods generally rely on predefined and fixed adjacency graphs, which often contain noise and redundancy, further degrading the learning performance. To address these issues, this paper proposes a novel Unified Structural Matrix Framework (USMF) for unsupervised feature selection. Specifically, a dynamically learnable, non-negative, symmetric structure matrix is introduced to consistently capture both global and local structural relationships within a single variable. This matrix preserves the global data structure via a trace-maximizing term, similar to weighted principal component analysis, while maintaining local geometric relationships via weighted distance constraints in the embedding space. Furthermore, a Frobenius norm (F-norm) consistency constraint is introduced to preserve neighborhood information and stabilize structure learning. Concurrently, an $l_{2,1}$ -norm regularization is imposed on the projection matrix to achieve effective feature selection. An alternating optimization algorithm is designed to solve the proposed model. Extensive experiments on multiple public datasets demonstrate that USMF outperforms existing state-of-the-art unsupervised feature selection methods in terms of Clustering Accuracy (ACC) and Normalized Mutual Information (NMI).

Key words: unsupervised feature selection; Unified Structural Matrix Framework (USMF); consistency

1 Introduction

With the widespread application of high-dimensional data in fields such as image recognition, bioinformatics, and text analysis^[1–4], unsupervised

feature selection has emerged as a crucial technique for mitigating the curse of dimensionality and enhancing model discriminability power^[5, 6]. In the absence of labeled information, effectively uncovering the

- Guoqiu Wen, Xinjie Han, Yuanpeng Li, Jiale Liu, Shuyu Li, and Qinghong Jiang are with the Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin, 541004, China, and also with the Guangxi Key Laboratory of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin, 541004, China. E-mail: wenguoqiu2008@163.com; hanxinjie@stu.gxnu.edu.cn; liyuanpeng@stu.gxnu.edu.cn; lishuyu@stu.gxnu.edu.cn; qhjiang@mailbox.gxnu.edu.cn.
- Peng Liu is with School of Computer Science and Engineering, Guilin University of Aerospace. E-mail: guaiguazhuhai@126.com.
- Jiayang Su is with Hainan University, Hainan, China. E-mail: sujiayang2000@gmail.com.
- Cong Lei is with University of Sydney, Sydney, NSW 2006, Australia. E-mail: clei9249@uni.sydney.edu.au.

* To whom correspondence should be addressed.

Manuscript received: 2026-03-07; revised: 2026-06-09; accepted: 2026-06-19

intrinsic structure of data and selecting representative features remains a central challenge in current research^[7, 8].

Structural information is widely recognized as critical for unsupervised feature selection, encompassing both local geometric structures that characterize sample neighborhood relationships and global structural features reflecting overall distribution patterns^[9]. Recent graph similarity learning methods have further demonstrated that modeling structural interactions across different representation levels can notably improve the consistency and discriminability of the resulting relationships^[10]. Existing methods typically model structural information from either a local or a global perspective. Local structure-based approaches often rely on pre-constructed adjacency graphs to preserve geometric relationships between samples^[11, 12], while global structure-based methods characterize overall data distributions through covariance or low-rank representations^[13]. Some recent studies have attempted to incorporate both local and global constraints simultaneously^[14, 15], or improve graph construction efficiency via anchor graphs^[16]. In addition, adaptive structural fusion strategies have been explored to integrate structural information from different scales, and their the effectiveness of adaptive structural weighting in representation learning^[17]. However, these methods often combine different types of structural information in a parallel manner, without a unified structural representation mechanism. Consequently, the sample relationships emphasized by local and global constraints may not align, misaligned, which limits the model's ability to capture complex data structures.

To address the aforementioned issues, this paper proposes a Unified Structural Modeling Framework (USMF). By introducing a learnable structural matrix, it collaboratively captures local geometric relationships and global structural information through a single shared variable. Specifically, the structural matrix participates in both local distance weighting and global covariance weighting, ensuring the consistency of sample relationship importance across different structural metrics. To enhance the stability of structure learning, a Frobenius-norm (F-norm)-based structural consistency constraint is that aligns to align the learned structural matrix with a reference structure, ensuring reliable structure estimation. Concurrently, row

sparsity regularization is applied to the projection matrix, enabling joint optimization of feature selection and structural modeling within a unified framework. The main contributions of this paper are as follows.

A unified structural-matrix-based unsupervised feature selection model is introduced, which jointly captures local and global structural relationships within a single framework.

A structural consistency regularization mechanism is developed to enhance model stability while retaining structural expressiveness.

Extensive experimental results on multiple public datasets are conducted, and the results confirm the effectiveness of the proposed method.

2 Related Work

Subspace learning is a key paradigm in high-dimensional data analysis and feature extraction, with its core objective being to preserve critical structural information during dimensionality reduction, thereby obtaining more discriminative low-dimensional representations^[18, 19]. Existing subspace learning methods can be broadly categorized into two types: local structure preservation methods and global information preservation methods.

Local structure preservation methods are based on the manifold hypothesis, which posits that high-dimensional data typically reside on a low-dimensional manifold space^[20]. Therefore, these methods emphasize maintaining local neighborhood relationships and manifold geometric features during dimensionality reduction. Belkin et al.^[11] proposed Laplacian Embedding (LE), a method that preserves local neighborhood relationships using the graph Laplacian operator. Subsequently, He et al.^[9] developed Locality-Preserving Projection (LPP) and proposed Neighborhood-Preserving Embedding (NPE), extending the preservation of local structure to the framework of linear subspace learning. Such methods typically rely on pre-constructed similarity graphs in the high-dimensional observation space, making them sensitive to noise and outliers while struggling to correct initial graph construction errors.

Global information preservation methods focus on extracting the primary variation patterns and overall distribution of data. Abdi et al.^[21] summarized Principal Component Analysis (PCA), a method that captures the primary information in data by identifying

the projection direction with the greatest variance. To address non-linear data structures, Ebied et al.^[22] extended PCA to kernel space and proposed Kernel Principal Component Analysis (KPCA), thereby enabling non-linear feature extraction via kernel mapping. While such methods effectively capture global features, they often neglect local manifold structures, leading to insufficient discriminative power on complex data. Consequently, Wen et al.^[23] proposed a variant of spectral-analysis-based clustering that employs geometric transformations and rotation mechanisms to enhance the stability of the representation.

Additionally, to jointly exploit both local and global structural information, recent studies have focused on unified feature selection frameworks that integrate multiple structural cues within a single optimization model. For example, Zhang et al.^[24] learned adaptive similarity structures across multiple views and performs feature selection in a unified inference framework. Wang et al.^[25] introduced group-wise sparsity into spatial pattern learning to jointly capture discriminative structures and feature relevance. Jiang et al.^[26] impose constraints to enhance feature selection performance under semi-supervised settings. Lu et al.^[27] proposed a dual-enhancement framework based on graph attention to simultaneously strengthen various structural interactions in multi-view clustering. Wu et al.^[28] combined a pseudo-label strategy with knowledge distillation to improve learning performance in graph-based few-shot scenarios.

However, most existing methods either rely on predefined structural assumptions, or model local and global information separately, or merely perform a simple weighted fusion of the two, making it difficult to fully exploit their complementary nature. In contrast, this paper proposes a unified structural learning framework that simultaneously models local manifold structure and global statistical consistency within a single representation space, thereby enabling the collaborative optimization of these two types of information to achieve unsupervised feature selection.

3 Method

This paper proposes an unsupervised feature selection method based on a unified and learnable structural matrix, which directly models the importance of relationships between samples. USMF uses a shared

structural matrix to jointly guide the learning of local and global structures. As shown in Fig. 1, the structural matrix encodes the importance of pairwise sample relationships. Sample pairs with stronger structural associations are pulled closer in the learned representation space, while sample pairs with weaker associations are pushed farther apart. In this way, local relationships are no longer determined solely by predefined neighborhoods but are adaptively adjusted according to global structural information.

3.1 Problem definition and symbols

In this method, we simultaneously characterize both the local geometric relationships and the global structural information of samples within the same representation space via a unified structural modeling mechanism. To this end, USMF introduces a learnable structural matrix to capture the structural associations among samples. Unlike approaches that introduce independent local and global constraints, our method enables the collaborative optimization of representation learning and structural modeling by leveraging the same structural matrix across different structural metrics. Table 1 lists the main general symbols used in this paper and their meanings.

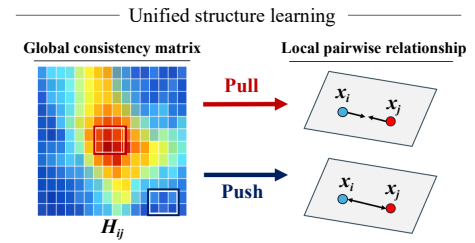


Fig. 1 Structure consistency learning where x_i and x_j denote two samples, and H_{ij} represents the weight of the (i, j) -th entry in the structure matrix, which reflects the importance of the relationship between samples.

Table 1 Notations and descriptions.

Notation	Description
$X \in \mathbb{R}^{d \times n}$	Original data matrix
$x_i \in \mathbb{R}^d$	Feature vector of the i -th sample
k	Dimension of the target low-dimensional subspace
d	Dimension of the original feature space
n	Number of samples
I	Identity matrix with compatible dimension
$\mathbf{1}$	All-one vector
K	Number of nearest neighbors in the KNN graph

3.2 Data reconstruction

A fundamental premise of dimensionality reduction or representation learning is that the low-dimensional representation should adequately preserve the primary information structure of the original data. To this end, this paper first learns a low-dimensional representation of the data through a linear reconstruction model. Given the centered data matrix $X \in \mathbb{R}^{d \times n}$, we introduce a projection matrix $W \in \mathbb{R}^{d \times k}$ and a low-dimensional representation $Z \in \mathbb{R}^{k \times n}$ (z_i and z_j denote the low-dimensional representation of x_i and x_j). By minimizing the reconstruction error (Eq. (1)), we ensure that the learned low-dimensional representation Z and projection matrix W effectively capture the primary information patterns within the data.

$$\min_{W, Z} \|X - WZ\|_F^2 = \min_{W, Z} \sum_{i=1}^n \|x_i - Wz_i\|_2^2 \quad (1)$$

To further enable feature selection, an $\ell_{2,1}$ -norm regularization is imposed on W , encouraging uninformative features to be automatically suppressed.

3.3 Unified structure matrix learning

To effectively exploit structural information in the learned representation space, we introduce a learnable unified structure matrix H to characterize the relationships among samples. Unlike conventional methods that separately model local and global structures, the proposed structure matrix serves as a shared structural representation and simultaneously participates in both local and global structure learning. Based on this unified representation, the following sections describe its learning strategy and structural modeling process.

3.3.1 Adaptive learning with structure prior

Traditional unsupervised feature selection methods typically construct local and global structural constraints separately; however, there is a lack of a unified structural representation mechanism between the two, which may lead to inconsistencies in the sample relationships emphasized by different structural objectives. To address this issue, we introduce a unified structure matrix $H \in \mathbb{R}^{n \times n}$ to model the structural relationships between samples, where the element H_{ij} represents the importance of the structural association between samples x_i and x_j .

Unlike traditional fixed graph structures, the structural matrix H described in this paper is learned adaptively through an optimization process and is

simultaneously involved in subsequent local structure preservation and global structure modeling processes, thereby enabling consistent collaborative learning among different structural objectives within a unified structural space.

Although allowing H to optimize freely maximizes the flexibility of structure learning, an unconstrained dynamic matrix is highly susceptible to numerical degradation or overfitting during the alternating optimization process. For instance, without proper boundary constraints, H might easily drift towards trivial solutions, causing severe structural drift. To enhance the stability of structural learning, this paper further introduces an initial structure matrix $S \in \mathbb{R}^{n \times n}$ as a structural prior.

The prior structure matrix S is constructed based on a symmetric K -nearest-neighbor (KNN) graph in the original feature space. Specifically, we first define an adjacency matrix $A \in \mathbb{R}^{n \times n}$ as

$$A_{ij} = \begin{cases} 1, & x_j \in \mathcal{N}_k(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\mathcal{N}_k(x_i)$ denotes the set of the KNNs of the sample x_i . The neighborhood size is fixed at $K = 5$ for all datasets in line with common practice. The prior structure matrix S is then obtained by symmetrizing.

$$S = \frac{A + A^T}{2} \quad (3)$$

Based on this, this paper introduces the following structural consistency constraints:

$$\min_H \|H - S\|_F^2 \quad (4)$$

This regularization term serves as a key prior consensus mechanism. It ensures that, while H has the freedom to adaptively optimize and refine its alignment in a low-dimensional space, it is prevented H from degenerating during the joint optimization process. In particular, the F-norm is used to measure the overall difference between the learned structure and the initial structure. By limiting the degree of deviation between the learned structure and the initial structure, this constraint ensures that the structure matrix H maintains consistency with the original neighborhood structure while retaining its adaptive learning capability, thereby enhancing the stability of structure learning.

3.3.2 H -induced local structure preservation

Local structure is commonly used to describe the geometric relationships between samples within a neighborhood. The underlying assumption is that if two

samples, $\mathbf{x}_i$ and $\mathbf{x}_j$, are intrinsically close, their low-dimensional representations, $\mathbf{z}_i$ and $\mathbf{z}_j$ (where $\mathbf{Z} = \mathbf{W}^T \mathbf{X}$), should remain adjacent. Unlike traditional methods that utilize a static heat-kernel weight, we employ the dynamically learned H_{ij} to weight the pairwise Euclidean distances in the embedding space. Formally, this local structure is formulated as the following minimization objective:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} \sum_{i,j=1}^n H_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 &= \min_{\mathbf{W}, \mathbf{H}} \text{Tr}(\mathbf{Z} \mathbf{L}_H \mathbf{Z}^T) \\ &= \min_{\mathbf{W}, \mathbf{H}} \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_H \mathbf{X}^T \mathbf{W}) \end{aligned} \quad (5)$$

where $\mathbf{L}_H = \mathbf{D}_H - \mathbf{H}$ denotes the Laplacian matrix derived from the learnable structure matrix $\mathbf{H}$, and $\mathbf{D}_H$ is the diagonal degree matrix with its i -th diagonal element defined as $(\mathbf{D}_H)_{ii} = \sum_{j=1}^n H_{ij}$. By minimizing this term, H_{ij} is adaptively measure the strength of the local structural correlation between samples $\mathbf{x}_i$ and $\mathbf{x}_j$. When H_{ij} is large, the distance between the corresponding samples in the low-dimensional space is subject to stronger constraints; conversely, it weakens the constraints on their local association.

3.3.3 H -induced global structure preservation

In the classical PCA framework, by minimizing the linear reconstruction error $\|\mathbf{X} - \mathbf{W} \mathbf{W}^T \mathbf{X}\|_F^2$, we ensure that the low-dimensional representation preserves as much global information from the original data as possible.

Let $\mathbf{P} = \mathbf{W} \mathbf{W}^T$ with $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, then we get:

$$\begin{aligned} \|\mathbf{X} - \mathbf{W} \mathbf{W}^T \mathbf{X}\|_F^2 &= \|(\mathbf{I} - \mathbf{P}) \mathbf{X}\|_F^2 \\ &= \text{tr}(\mathbf{X}^T (\mathbf{I} - \mathbf{P})^T (\mathbf{I} - \mathbf{P}) \mathbf{X}) \\ &= \text{tr}(\mathbf{X}^T (\mathbf{I} - \mathbf{P}) \mathbf{X}) \\ &= \text{tr}(\mathbf{X}^T \mathbf{X}) - \text{tr}(\mathbf{X}^T \mathbf{P} \mathbf{X}) \\ &= \text{tr}(\mathbf{X}^T \mathbf{X}) - \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) \end{aligned} \quad (6)$$

Since $\text{tr}(\mathbf{X}^T \mathbf{X})$ is constant with respect to $\mathbf{W}$, minimizing $\|\mathbf{X} - \mathbf{W} \mathbf{W}^T \mathbf{X}\|_F^2$ is equivalent to maximizing $\text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W})$, where the projection matrix $\mathbf{W}$ defines the learned low-dimensional space.

Expanding the trace term reveals that this objective essentially measures the preservation of global structure via the inner products of the resulting low-dimensional projection vectors:

$$\begin{aligned} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) &= \sum_{i=1}^n \|\mathbf{W}^T \mathbf{x}_i\|_2^2 \\ &= \sum_{i=1}^n (\mathbf{W}^T \mathbf{x}_i)^T (\mathbf{W}^T \mathbf{x}_i) \end{aligned} \quad (7)$$

This formulation implicitly assumes that all samples contribute uniformly to the global structure, meaning the global structure is dominated solely by the overall variance of the data without distinguishing the importance of different sample relationships. To explicitly preserve the weighted global structural relationships among samples in the low-dimensional space, we introduce a non-negative symmetric structural matrix $\mathbf{H} \in \mathbb{R}^{n \times n}$, where H_{ij} denotes the structural correlation weight between samples $\mathbf{x}_i$ and $\mathbf{x}_j$. Based on this structural matrix, the trace maximization term in classical PCA can be naturally extended to the following weighted form:

$$\sum_{i,j=1}^n H_{ij} (\mathbf{W}^T \mathbf{x}_i)^T (\mathbf{W}^T \mathbf{x}_j) = \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) \quad (8)$$

This weighted trace term characterizes the ability of low-dimensional representations to preserve the original weighted global structure in the inner product sense. Unlike traditional PCA, the global structure is no longer solely determined by data variance but is instead regulated by the learnable structure matrix $\mathbf{H}$. This ensures that the subspace learning process maintains structural consistency with the global structure.

It should be noted that although the reconstruction term and the weighted trace term formally involve high-to-low and low-to-high dimensional mappings, respectively, they are unified by sharing the same projection matrix $\mathbf{W}$: specifically, $\mathbf{W}^T$ maps data to a low-dimensional space for global structural weighting, while $\mathbf{W}$ performs linear mapping and representation reconstruction within the same subspace. This weight-sharing mechanism ensures that global structural modeling and representation learning proceed collaboratively within the same subspace.

Through the aforementioned modeling approach, the structural matrix $\mathbf{H}$ is naturally incorporated into the global subspace learning process. This ensures that the learned low-dimensional representations not only preserve holistic information but also reflect global associative relationships among samples at the structural level. Since maximizing this term is equivalent to minimizing its negative counterpart. Consequently, this global structure is formulated as the following minimization objective:

$$\min_{\mathbf{W}, \mathbf{H}} -\text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) \quad (9)$$

3.3.4 The unified objective function

Within the unified structural learning framework described above, $\mathbf{H}$ captures and coordinates both local and global structures in a unified manner. Taking into account reconstruction error, local structure preservation, global structure preservation, and structural consistency, the final optimization model in this paper can be expressed as

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{Z}, \mathbf{H}} \quad & \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{W}\mathbf{z}_i\|_2^2 + \alpha \sum_{i,j=1}^n H_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \\ & - \gamma \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) + \frac{\delta}{2} \|\mathbf{H} - \mathbf{S}\|_F^2 \\ & + \beta \|\mathbf{W}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{W}^T \mathbf{W} = \mathbf{I}, \quad \mathbf{H} = \mathbf{H}^T, \quad H_{ij} \geq 0 \end{aligned} \quad (10)$$

3.4 Convergence analysis

Since the objective function constitutes a joint non-convex optimization problem with respect to the variables $\mathbf{W}$, $\mathbf{Z}$, and $\mathbf{H}$, it is difficult to obtain a global optimum for the overall problem directly. To address this, this paper employs a Block Coordinate Descent (BCD) strategy to perform alternating iterative optimization of the variables $\mathbf{W}$, $\mathbf{Z}$, and $\mathbf{H}$. In each iteration, two of the variables are fixed while the remaining one is optimized, until the objective function converges.

Let the objective function be denoted as

$$J(\mathbf{W}, \mathbf{Z}, \mathbf{H}) \quad (11)$$

the variables in the t -th iteration are set and are set in the t -th iteration as

$$\begin{aligned} J(\mathbf{W}^T, \mathbf{Z}^{t+1}, \mathbf{H}^T) &\leq J(\mathbf{W}^T, \mathbf{Z}^T, \mathbf{H}^T) \\ J(\mathbf{W}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{H}^T) &\leq J(\mathbf{W}^T, \mathbf{Z}^{t+1}, \mathbf{H}^T) \\ J(\mathbf{W}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{H}^{t+1}) &\leq J(\mathbf{W}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{H}^T) \end{aligned} \quad (12)$$

The final solution is:

$$J(\mathbf{W}^T, \mathbf{Z}^T, \mathbf{H}^T) \quad (13)$$

The proofs for each sub-problem are as

(1) Update $\mathbf{Z}$:

With $\mathbf{W}$ and $\mathbf{H}$ being fixed, the sub-problem of the objective function with respect to $\mathbf{Z}$ is:

$$\min_{\mathbf{Z}} J(\mathbf{Z}) = \min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{W}\mathbf{Z}\|_F^2 + \alpha \text{tr}(\mathbf{Z} \mathbf{L}_H \mathbf{Z}^T) \quad (14)$$

Expanding the F-norm and focusing solely on the terms containing $\mathbf{Z}$, we obtain:

$$J(\mathbf{Z}) = \text{tr}(\mathbf{Z}^T \mathbf{Z}) - 2\text{tr}(\mathbf{Z}^T \mathbf{W}^T \mathbf{X}) + \alpha \text{tr}(\mathbf{Z} \mathbf{L}_H \mathbf{Z}^T) \quad (15)$$

Differentiating $J(\mathbf{Z})$ with respect to $\mathbf{Z}$ and setting the gradient to zero leads to

$$\frac{\partial J}{\partial \mathbf{Z}} = 2\mathbf{Z} - 2\mathbf{W}^T \mathbf{X} + 2\alpha \mathbf{Z} \mathbf{L}_H = 0 \quad (16)$$

Solving the above system of linear equations yields:

$$\mathbf{Z} = \mathbf{W}^T \mathbf{X} (\mathbf{I} + \alpha \mathbf{L}_H)^{-1} \quad (17)$$

Since matrix $(\mathbf{I} + \alpha \mathbf{L}_H)$ is always invertible, the sub-problem regarding $\mathbf{Z}$ has a unique closed-form solution; furthermore, as this sub-problem is strictly convex quadratic optimization problem with respect to $\mathbf{Z}$, the solution obtained is the global optimum.

(2) Update $\mathbf{W}$:

If $\mathbf{Z}$ and $\mathbf{H}$ are fixed, the sub-problem of the objective function with respect to $\mathbf{W}$ is:

$$\begin{aligned} \min_{\mathbf{W}} J(\mathbf{W}) &= \min_{\mathbf{W}} \|\mathbf{X} - \mathbf{W}\mathbf{Z}\|_F^2 - \gamma \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) \\ &+ \beta \|\mathbf{W}\|_{2,1} \end{aligned} \quad (18)$$

Since the $\ell_{2,1}$ -norm is non-smooth at the origin, we introduce a diagonal auxiliary matrix $\mathbf{D} \in \mathbb{R}^{d \times d}$, where the i -th diagonal element is defined as $D_{ii} = \frac{1}{2\|\mathbf{W}^i\|_2}$. Thus, Eq. (18) becomes:

$$\begin{aligned} \min_{\mathbf{W}} \text{tr}(\mathbf{Z}^T \mathbf{W}^T \mathbf{W} \mathbf{Z}) - 2\text{tr}(\mathbf{Z}^T \mathbf{W}^T \mathbf{X}) \\ - \gamma \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) + \beta \text{tr}(\mathbf{W}^T \mathbf{D} \mathbf{W}) \end{aligned} \quad (19)$$

Given that $\mathbf{Z} = \mathbf{W}^T \mathbf{X}$, and $\sum_{i,j=1}^n H_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 = 2\text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_H \mathbf{X}^T \mathbf{W})$, Eq. (19) can be transferred to:

$$\max_{\mathbf{W}} \text{tr}(\mathbf{W}^T (\mathbf{X} \mathbf{X}^T - \alpha \mathbf{X} \mathbf{L}_H \mathbf{X}^T + \gamma \mathbf{X} \mathbf{H} \mathbf{X}^T - \beta \mathbf{D}) \mathbf{W}) \quad (20)$$

Let $\mathbf{Q} = \mathbf{X} \mathbf{X}^T - \alpha \mathbf{X} \mathbf{L}_H \mathbf{X}^T + \gamma \mathbf{X} \mathbf{H} \mathbf{X}^T - \beta \mathbf{D}$, the problem then becomes $\max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{Q} \mathbf{W})$, and satisfies $\mathbf{W}^T \mathbf{W} = \mathbf{I}$. According to the Rayleigh-Ritz theorem, the optimal solution to this problem consists of the eigenvectors corresponding to the k largest eigenvalues of matrix $\mathbf{Q}$. Consequently, the $\mathbf{W}$ sub-problem has an optimal closed-form solution at every iteration.

(3) Update $\mathbf{H}$:

When $\mathbf{Z}$ and $\mathbf{W}$ are fixed, the sub-problem of the objective function with respect to $\mathbf{H}$ is:

$$\begin{aligned} \min_{\mathbf{H}} J(\mathbf{H}) &= \min_{\mathbf{H}} \alpha \sum_{i,j=1}^n H_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \\ &- \gamma \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{W}) + \frac{\delta}{2} \|\mathbf{H} - \mathbf{S}\|_F^2 \end{aligned} \quad (21)$$

Let $\mathbf{D}_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|_2^2$ and $\mathbf{G} = (\mathbf{W}^T \mathbf{X})^T (\mathbf{W}^T \mathbf{X})$, Eq. (21) can be transferred to:

$$\alpha \text{tr}(\mathbf{H}^T \mathbf{D}) - \gamma \text{tr}(\mathbf{H}^T \mathbf{G}) + \frac{\delta}{2} \|\mathbf{H} - \mathbf{S}\|_F^2 \quad (22)$$

Let $\mathbf{M} = \alpha\mathbf{D} - \gamma\mathbf{G}$, Eq. (22) can be formulated as

$$\langle \mathbf{H}, \mathbf{M} \rangle + \frac{\delta}{2} \|\mathbf{H} - \mathbf{S}\|_F^2 \quad (23)$$

Expanding the F-norm gives

$$\|\mathbf{H} - \mathbf{S}\|_F^2 = \text{tr}((\mathbf{H} - \mathbf{S})^T(\mathbf{H} - \mathbf{S})) \quad (24)$$

Therefore, the objective function is a standard convex quadratic function on $\mathbf{H}$, we differentiate it with respect to $\mathbf{H}$ and set it to zero:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{H}} = \mathbf{M} + \delta(\mathbf{H} - \mathbf{S}) = 0 \quad (25)$$

Solving the above equation gives:

$$\mathbf{H} = \mathbf{S} - \frac{1}{\delta}(\alpha\mathbf{D} - \gamma\mathbf{G}) \quad (26)$$

Since $\mathbf{H}$ is non-negative and symmetric, and $\delta > 0$, the objective function is strictly convex with respect to $\mathbf{H}$; consequently, the $\mathbf{H}$ sub-problem has a unique global optimum.

In summary, as each sub-problem has a solution during the optimization process, the optimization algorithm proposed in this paper ensures that the objective function value remains non-increasing throughout the iterative process and ultimately converges. For the specific optimization algorithm, see Algorithm 1.

4 Experiment

We carry out all experiments on a server with 8 NVIDIA GeForce RTX 4090 GPUs. A grid search is applied to explore key hyperparameters over the range from 10^{-4} to 10^1 . During training, each parameter combination runs independently on a separate GPU, guaranteeing stable and reproducible experimental results. All model training and evaluation procedures are automated using Python scripts.

4.1 Datasets and comparison methods

To comprehensively evaluate USMF's performance, we select 9 public datasets from different domains, including face image datasets (Yale, ORL, JAFFE,

Algorithm 1 Optimization of USMF

Require: $\mathbf{X} \in \mathbb{R}^{d \times n}$, $\mathbf{S} \in \mathbb{R}^{n \times n}$, parameters $\alpha, \beta, \gamma, \delta$, target dimension k , neighborhood size K

Ensure: $\mathbf{W} \in \mathbb{R}^{d \times k}$, $\mathbf{Z} \in \mathbb{R}^{k \times n}$, $\mathbf{H} \in \mathbb{R}^{n \times n}$

- 1: **repeat**
- 2: Construct Laplacian matrix $\mathbf{L}_H$
- 3: **update Z:**
- 4: $\mathbf{Z} = \mathbf{W}^T \mathbf{X} (\mathbf{I} + \alpha \mathbf{L}_H)^{-1}$
- 5: **update W:**
- 6: $\mathbf{D} = \text{diag}\left(\frac{1}{2\|\mathbf{w}^i\|_2}\right)$
- 7: $\mathbf{Q} = \mathbf{X}\mathbf{X}^T - \alpha\mathbf{X}\mathbf{L}_H\mathbf{X}^T + \gamma\mathbf{X}\mathbf{H}\mathbf{X}^T - \beta\mathbf{D}$
- 8: Compute the top- K eigenvectors of $\frac{1}{2}(\mathbf{Q} + \mathbf{Q}^T)$ to obtain $\mathbf{W}$
- 9: Orthogonalize $\mathbf{W}$ via QR decomposition
- 10: **update H:**
- 11: Compute $D_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|_2^2$ and $\mathbf{G} = \mathbf{Z}^T \mathbf{Z}$
- 12: $\mathbf{M} = \alpha\mathbf{D} - \gamma\mathbf{G}$
- 13: $\mathbf{H} = \mathbf{S} - \frac{1}{\delta} \mathbf{M}$
- 14: $\mathbf{H} = \frac{1}{2}(\mathbf{H} + \mathbf{H}^T)$
- 15: $\mathbf{H} = \max(\mathbf{H}, 0)$
- 16: **until** convergence

ATT40), biological datasets (lung_small, lung), and object image datasets (COIL20, binalpha, USPS). Their detail information is shown in Table 2.

To systematically evaluate the effectiveness of the proposed method in unsupervised representation learning and feature selection tasks, this paper adopts a variety of representative existing methods as baselines for comparison.

Fast Sparse Discriminative K-Means (FSDK)^[29] incorporates a weighted discrete pseudo-label matrix into least-squares regression to eliminate trivial solutions, and utilizes $\ell_{2,p}$ -norm regularization to achieve row sparsity, thereby unifying discriminant K-means and sparse regression into an unsupervised

Table 2 Dataset statistics and sources.

Dataset	Sample	Feature	Classe	Source
Yale	165	1024	15	http://www.cad.zju.edu.cn/home/dengcai/Data/data.html
COIL20	1440	1024	20	http://www.cad.zju.edu.cn/home/dengcai/Data/data.html
ORL	400	1024	40	https://jundongli.github.io/scikit-feature/datasets.html
JAFFE	213	256	10	https://arxiv.org/abs/2403.10910
ATT40	400	1024	40	https://doi.org/10.1109/TIP.2015.2479560
binalpha	1404	320	36	https://doi.org/10.1016/j.knosys.2021.106847
lung_small	73	325	36	https://jundongli.github.io/scikit-feature/datasets.html
lung	203	3312	5	https://jundongli.github.io/scikit-feature/datasets.html
USPS	9298	256	10	https://postalpro.usps.com

feature selection framework.

Sparse Feature Selection via Fast Embedding Spectral Analysis (SFESA)^[30] generates anchors using the Balanced Hierarchical K-means (BKHK) algorithm and constructs an adaptive anchor neighborhood graph; it utilizes spectral analysis to learn the data manifold structure, combines this with an $\ell_{2,0}$ -norm constrained projection matrix to achieve strong sparsity, and designs a fast iterative algorithm to solve non-convex optimization problems.

Graph-Regularized Consensus Learning and Diversity Representation for Unsupervised multi-view Feature Selection (GCDUFS)^[31] employs graph regularization to constrain the consensus matrix in order to capture consistent manifold structures across multiple views, utilizes kernel functions and the HSIC metric to learn independent diversity information from each view, and combines this with symmetric non-negative matrix factorization for clustering analysis, thereby unifying consistency learning, diversity representation, and sparse feature selection within an optimization framework.

Sparse Projection and Discriminant Feature Selection (SPDFS)^[32] combines sparse projection with an $\ell_{2,0}$ -norm constraint and fuzzy k-means clustering to perform both fuzzy membership learning and discriminative projection learning simultaneously. It employs both non-iterative and iterative optimization strategies to solve non-convex problems, thereby directly selecting the most discriminative subset of features.

Adaptive and Flexible ℓ_1 -norm Graph embedding (RAFG)^[33] incorporates the $\ell_{2,1}$ -norm into the elastic regression terms and employs adaptive ℓ_1 -norm graph learning to characterize cluster distributions, thereby achieving unsupervised feature selection.

Deterministic Column Subset Selection (DCSS)^[34] employs an autoencoder and an RBF-based clustering network to generate probabilistic cluster assignments, and optimizes a differentiable soft silhouette objective to learn compact and well-separated latent representations for clustering.

To verify the clustering performance of the algorithm, we adopt Clustering Accuracy (ACC) and Normalized Mutual Information (NMI) as evaluation metrics. ACC reflects the consistency between the clustering results and the ground-truth labels, while NMI measures the mutual information between the clustering and the true distribution. Both values lie within the range [0, 1]. Higher ACC and NMI values indicate better clustering performance and a more representative selected feature subset.

4.2 Experimental results and analysis

During the training process, we set the hyper parameters using grid search and select the parameter combination that yields the best performance for testing. We use the K-Means algorithm to cluster the data points based on the selected features. Considering the sensitivity of this algorithm to initialization, each experiment is repeated five times, and the final results are presented as the average of all trials.

Tables 3 and 4 show the mean and standard deviation of the best ACC and NMI achieved by USMF and other compared algorithms on different datasets, respectively. The results demonstrate that our proposed method outperforms all compared methods on the vast majority of datasets.

Specifically, in terms of the ACC metric, USMF consistently achieves the highest accuracy across all datasets except JAFFE. On the JAFFE dataset, USMF is only 0.14% lower. For the NMI metric, USMF

Table 3 ACC of different comparison methods on various datasets (mean $\pm$ std).

(Unit: %)

Dataset	FSDK	SFESA	GCDUFS	SPDFS	NFRFS	DCSS	USMF
Yale	37.06 $\pm$ 2.51	41.30 $\pm$ 2.78	39.64 $\pm$ 5.91	42.79 $\pm$ 2.82	50.39 $\pm$ 3.18	41.70 $\pm$ 3.41	51.64 $\pm$ 0.97
COIL20	63.61 $\pm$ 1.04	64.65 $\pm$ 2.76	66.97 $\pm$ 0.61	62.01 $\pm$ 2.91	69.94 $\pm$ 2.25	69.90 $\pm$ 0.97	84.79 $\pm$ 0.12
ORL	55.60 $\pm$ 1.97	59.49 $\pm$ 2.40	54.55 $\pm$ 1.29	57.50 $\pm$ 2.30	58.66 $\pm$ 2.37	59.75 $\pm$ 7.26	62.90 $\pm$ 1.16
JAFFE	75.31 $\pm$ 1.99	75.05 $\pm$ 4.67	82.16 $\pm$ 0.59	82.72 $\pm$ 4.48	83.43 $\pm$ 1.90	82.15 $\pm$ 2.61	83.29 $\pm$ 1.05
ATT40	58.43 $\pm$ 1.42	60.90 $\pm$ 1.74	56.20 $\pm$ 1.48	58.45 $\pm$ 2.21	59.31 $\pm$ 2.37	54.70 $\pm$ 4.92	63.15 $\pm$ 1.47
binalpha	41.62 $\pm$ 0.83	44.82 $\pm$ 1.82	44.53 $\pm$ 1.04	43.52 $\pm$ 1.59	45.08 $\pm$ 1.36	37.27 $\pm$ 2.32	49.13 $\pm$ 1.71
lung_small	70.40 $\pm$ 2.51	74.66 $\pm$ 2.06	87.12 $\pm$ 0.67	72.59 $\pm$ 5.89	83.97 $\pm$ 5.55	77.26 $\pm$ 7.46	89.32 $\pm$ 1.03
lung	62.44 $\pm$ 2.24	78.19 $\pm$ 0.91	76.06 $\pm$ 5.91	79.31 $\pm$ 8.73	84.04 $\pm$ 4.14	81.94 $\pm$ 1.67	86.31 $\pm$ 1.18
USPS	53.57 $\pm$ 3.55	73.19 $\pm$ 4.14	70.26 $\pm$ 1.29	66.63 $\pm$ 2.84	66.12 $\pm$ 0.21	62.21 $\pm$ 1.16	76.67 $\pm$ 0.04

Note: NFRFS: Unsupervised Feature Selection Algorithm based on $\ell_{2,p}$ -norm Feature Reconstruction.

Table 4 NMI of different comparison methods on various datasets (mean $\pm$ std).

(Unit: %)

Dataset	FSDK	SFESA	GCDUFS	SPDFS	NFRFS	DCSS	USMF
Yale	43.52 $\pm$ 2.71	48.41 $\pm$ 1.88	45.43 $\pm$ 1.82	47.64 $\pm$ 2.53	59.10 $\pm$ 2.15	50.05 $\pm$ 3.63	56.53 $\pm$ 1.17
COIL20	74.57 $\pm$ 0.64	77.76 $\pm$ 0.75	76.79 $\pm$ 0.89	73.01 $\pm$ 1.46	80.04 $\pm$ 1.04	79.32 $\pm$ 1.13	90.29 $\pm$ 0.14
ORL	73.53 $\pm$ 1.56	79.61 $\pm$ 1.13	72.41 $\pm$ 0.84	74.01 $\pm$ 1.33	78.19 $\pm$ 1.05	65.19 $\pm$ 2.61	79.97 $\pm$ 0.84
JAFPE	76.30 $\pm$ 1.91	77.78 $\pm$ 3.33	82.56 $\pm$ 0.87	82.75 $\pm$ 3.42	84.67 $\pm$ 1.50	81.80 $\pm$ 2.51	84.87 $\pm$ 0.70
ATT40	75.67 $\pm$ 0.77	79.22 $\pm$ 0.88	73.71 $\pm$ 0.82	74.79 $\pm$ 1.44	78.53 $\pm$ 1.31	71.42 $\pm$ 3.95	79.93 $\pm$ 0.56
binalpha	57.43 $\pm$ 0.58	60.52 $\pm$ 0.69	58.89 $\pm$ 0.66	59.05 $\pm$ 0.73	59.68 $\pm$ 0.93	52.43 $\pm$ 1.82	63.37 $\pm$ 0.69
lung_small	66.61 $\pm$ 1.91	70.90 $\pm$ 4.10	80.01 $\pm$ 0.31	67.50 $\pm$ 4.93	78.36 $\pm$ 4.33	72.58 $\pm$ 1.10	83.14 $\pm$ 1.39
lung	45.72 $\pm$ 1.70	67.28 $\pm$ 3.17	55.36 $\pm$ 2.89	58.22 $\pm$ 6.51	66.78 $\pm$ 3.16	64.13 $\pm$ 0.81	68.18 $\pm$ 1.82
USPS	53.40 $\pm$ 1.87	76.73 $\pm$ 1.56	63.62 $\pm$ 3.01	61.15 $\pm$ 0.79	64.69 $\pm$ 1.13	65.23 $\pm$ 1.82	72.11 $\pm$ 0.01

obtains the best results on most datasets and achieves the second-best performance on the Yale and USPS datasets, demonstrating its strong capability in preserving the intrinsic cluster structure of the data. These results indicate that the proposed framework can effectively identify informative features and enhance clustering quality from multiple perspectives.

Another notable observation is that the performance gains become more significant on relatively complex datasets. For instance, on the COIL20 dataset, USMF achieves an ACC of 84.79%, outperforming the second-best method by 14.85%, while the NMI reaches 90.29%, yielding an improvement of 9.49 %. Similarly, on the lung-small dataset, USMF obtains an ACC of 89.32%, which is 2.20% higher than that of the best competing method (87.12%), and achieves an NMI of 83.14%, surpassing the second-best result by 3.13%. These substantial improvements indicate that the proposed unified structural modeling mechanism is particularly effective in capturing complex intrinsic data structures and identifying discriminative features for clustering.

4.3 Ablation study

To further analyze the impact of each component module on the overall model performance, we conduct a systematic ablation study of the proposed method. Specifically, we remove the local structure preservation term, the global structure preservation term, and the structure consistency constraint term, respectively. Furthermore, we investigate the scenario in which the reference structure matrix S is not introduced under the F-norm constraint. This allows us to evaluate the actual contribution of each component within the unified structure modeling framework. It should be noted that the row-sparsity regularization term, which is employed for unsupervised feature selection,

constitutes a fundamental component of the model's task definition and is therefore not included as a removable module for individual ablation. The experimental results for each configuration are presented in Table 5.

The experimental results demonstrate that removing the local structure preservation term leads to varying degrees of performance degradation across multiple datasets. This indicates that relying solely on global structure is insufficient to fully capture the fine-grained neighborhood relationships between samples. Conversely, when the global structure term is removed, although the local geometric relationships are maintained, the overall discriminative ability is significantly weakened. This suggests that structural information at the global covariance level is equally crucial for stabilizing subspace learning.

Furthermore, when the consistency constraint on the structure matrix H is not enforced, the model's performance fluctuations increase. This suggests that excessive structural freedom may lead to overfitting to noise or anomalous relationships. Moreover, when the reference structure S is omitted from the F-norm regularization—even though structural complexity is partially controlled—the lack of prior structural guidance weakens the consistency between the learned structure and the original neighborhood patterns, thereby affecting the final results.

In summary, the complete model exhibits more stable and superior performance across all comparisons. This indicates that synergistically characterizing both local and global relationships via a unified structure matrix H , combined with stabilization through structure consistency constraints, is a key to enhancing the effectiveness of unsupervised representation learning.

Table 5 ACC and NMI of the ablation study (mean $\pm$ std).**(Unit: %)**

Dataset	Metric	Full Model	-Local	-Global	-Consistency	-Prior Structure
Yale	ACC	51.64 $\pm$ 0.97	42.79 $\pm$ 3.57	45.21 $\pm$ 2.05	45.09 $\pm$ 2.19	47.15 $\pm$ 1.69
	NMI	56.53 $\pm$ 1.17	49.60 $\pm$ 4.00	51.73 $\pm$ 2.31	51.88 $\pm$ 2.11	53.54 $\pm$ 2.43
COIL20	ACC	84.79 $\pm$ 0.12	65.53 $\pm$ 1.32	69.51 $\pm$ 1.65	67.72 $\pm$ 1.64	78.14 $\pm$ 0.73
	NMI	90.29 $\pm$ 0.14	78.43 $\pm$ 0.67	79.87 $\pm$ 0.78	79.18 $\pm$ 0.61	88.28 $\pm$ 0.51
ORL	ACC	62.90 $\pm$ 1.16	59.80 $\pm$ 0.81	60.35 $\pm$ 2.36	61.35 $\pm$ 0.46	61.75 $\pm$ 1.19
	NMI	79.97 $\pm$ 0.84	78.40 $\pm$ 0.46	78.45 $\pm$ 1.25	79.59 $\pm$ 0.22	79.46 $\pm$ 0.74
JAFPE	ACC	83.29 $\pm$ 1.05	82.07 $\pm$ 0.86	81.22 $\pm$ 1.85	82.79 $\pm$ 0.23	81.50 $\pm$ 2.79
	NMI	84.87 $\pm$ 0.70	82.92 $\pm$ 1.37	81.76 $\pm$ 1.95	83.91 $\pm$ 0.34	83.41 $\pm$ 2.28
ATT40	ACC	63.15 $\pm$ 1.47	56.90 $\pm$ 2.26	58.60 $\pm$ 1.40	59.70 $\pm$ 1.96	59.55 $\pm$ 2.13
	NMI	79.93 $\pm$ 0.56	77.12 $\pm$ 1.10	78.03 $\pm$ 0.62	78.65 $\pm$ 0.75	78.03 $\pm$ 0.64
binalpha	ACC	49.13 $\pm$ 1.71	47.42 $\pm$ 1.84	46.27 $\pm$ 2.11	48.18 $\pm$ 1.61	49.06 $\pm$ 1.32
	NMI	63.37 $\pm$ 0.69	60.73 $\pm$ 0.72	60.61 $\pm$ 0.99	61.48 $\pm$ 0.59	62.42 $\pm$ 0.46
lung_small	ACC	89.32 $\pm$ 1.03	80.55 $\pm$ 7.22	80.82 $\pm$ 7.50	83.56 $\pm$ 6.43	87.67 $\pm$ 2.87
	NMI	83.14 $\pm$ 1.39	76.53 $\pm$ 3.54	76.99 $\pm$ 4.11	78.05 $\pm$ 3.84	81.37 $\pm$ 2.93
lung	ACC	86.31 $\pm$ 1.18	72.22 $\pm$ 6.49	65.02 $\pm$ 1.25	80.89 $\pm$ 0.20	86.08 $\pm$ 0.98
	NMI	68.18 $\pm$ 1.82	64.16 $\pm$ 1.63	62.35 $\pm$ 0.40	63.63 $\pm$ 0.11	67.84 $\pm$ 1.50
USPS	ACC	76.67 $\pm$ 0.04	67.12 $\pm$ 0.06	67.20 $\pm$ 0.06	68.05 $\pm$ 0.02	76.55 $\pm$ 0.08
	NMI	72.11 $\pm$ 0.01	61.12 $\pm$ 0.05	61.22 $\pm$ 0.10	62.98 $\pm$ 0.05	71.88 $\pm$ 0.02

4.4 Parameter sensitivity experiment

To analyze the model's sensitivity to changes in key hyperparameters and the stability of its internal structural mechanisms, this paper conducts systematic joint experiments on structure-related parameters. Considering that the proposed model centers around a unified structure matrix H , whose role is simultaneously reflected in local structure preservation, global structure modeling, and structure consistency constraints, different parameters may exhibit mutual influence and synergistic regulatory relationships. Therefore, it is necessary to examine the interactions between parameters from the perspectives of structural synergy and structural stability.

On the one hand, the local structure weight α and the global structure weight γ jointly determine the intensity of the structure matrix's influence under different geometric measures. Their relative proportion directly affects the balance between preserving local geometric relationships and maintaining the global structure. On the other hand, the global structure weight γ and the consistency regularization parameter δ respectively regulate the structural representation capability and the strength of the structural complexity constraint. Their combined relationship reflects the trade-off mechanism between structure modeling and structure stabilization.

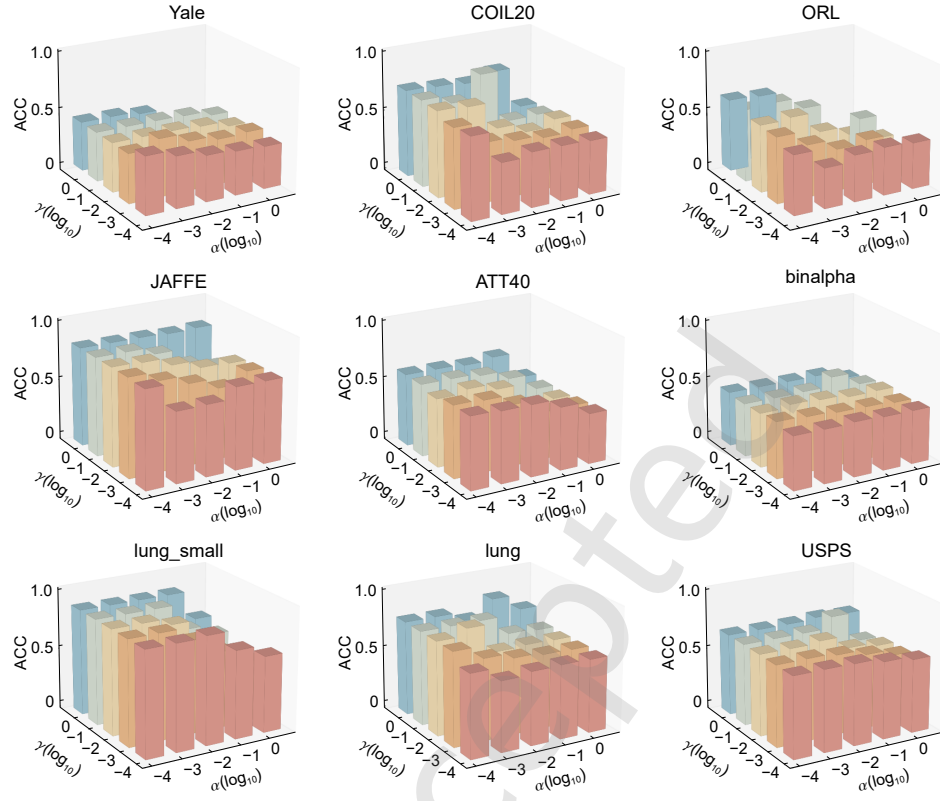
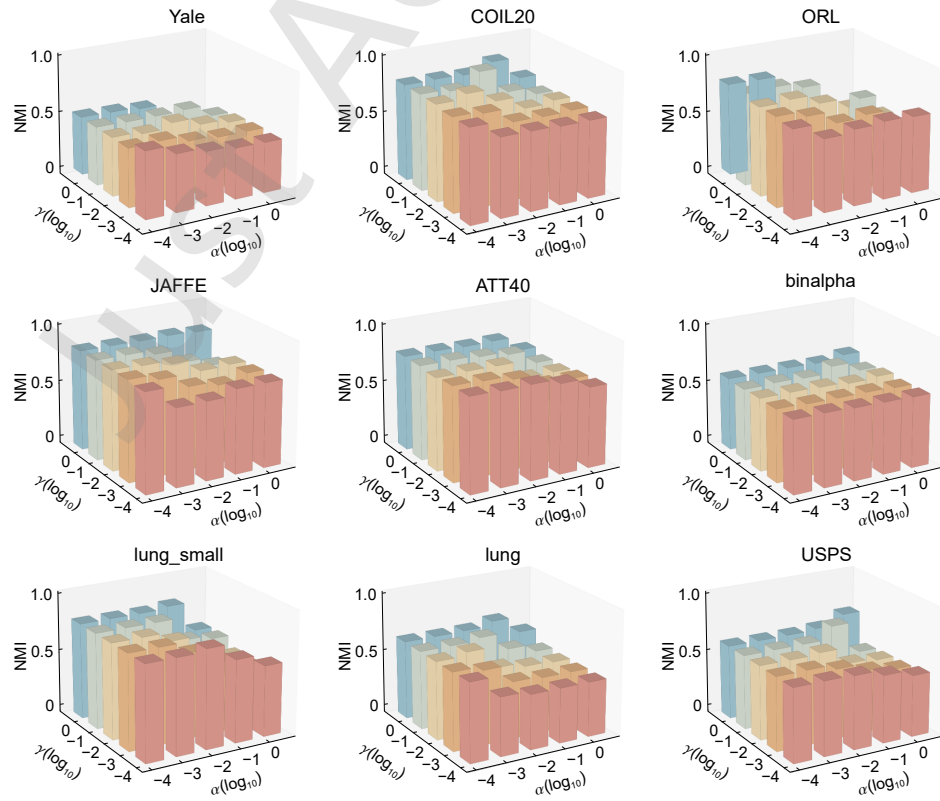
Based on the above considerations, this paper conducts a joint sensitivity analysis on the parameter pairs (α, γ) and (γ, δ) , respectively, to systematically evaluate the stability and robustness of the unified structural modeling framework under different weight configurations. The experimental results are shown in Figs. 2, 3, 4, and 5.

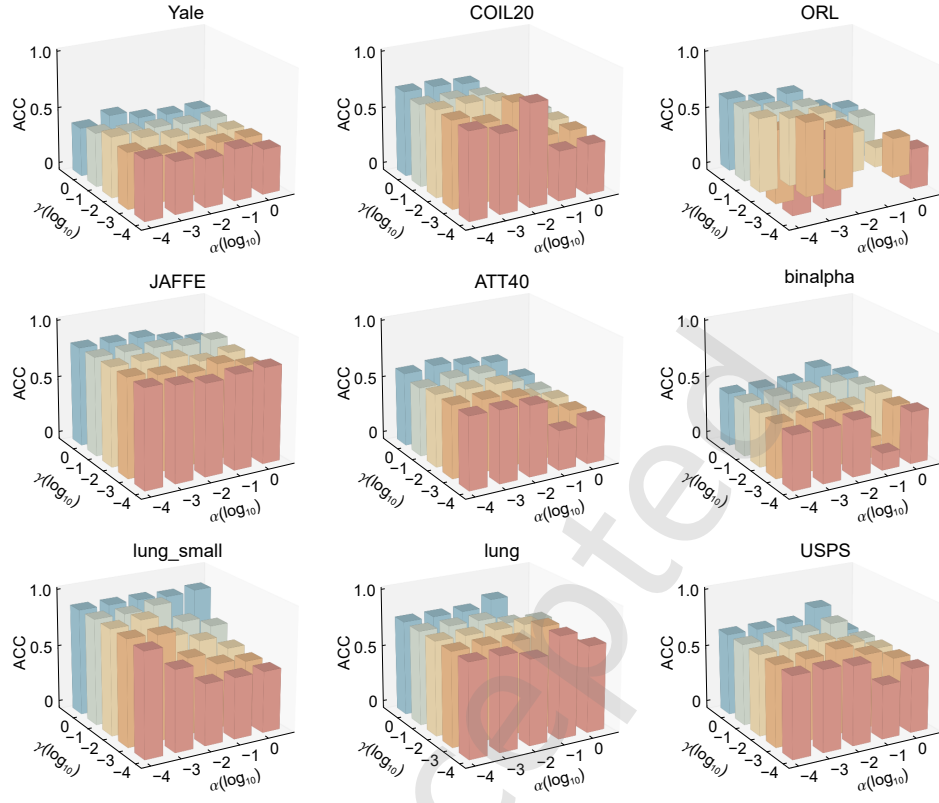
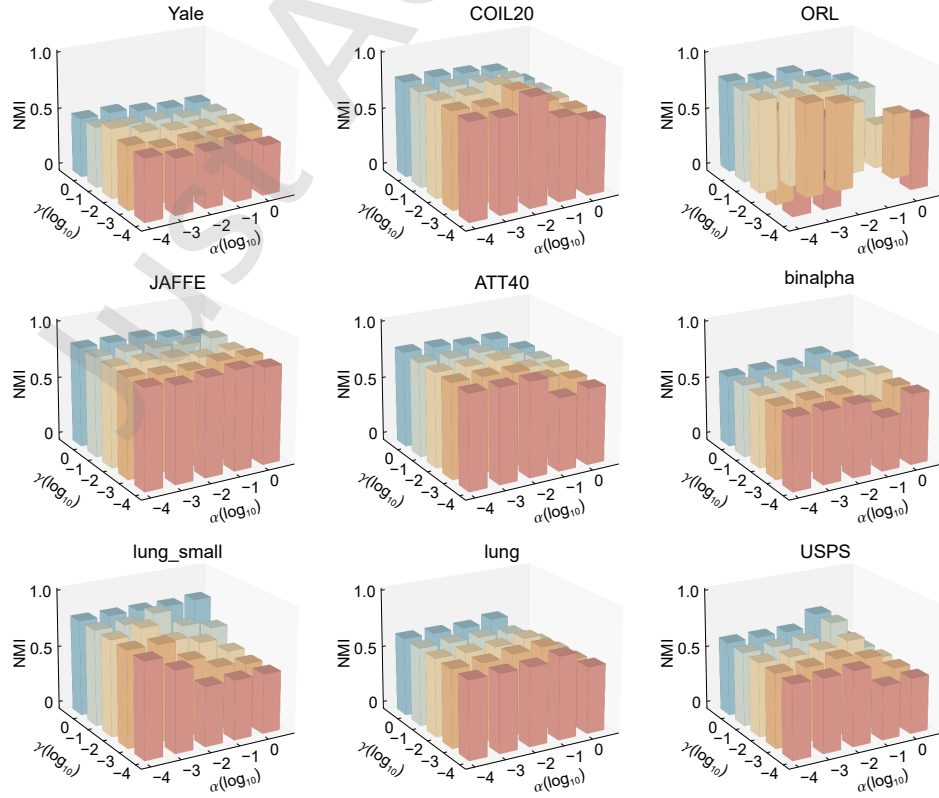
4.5 Convergence test

Figure 6 shows the convergence curves of the objective function values for USMF across all datasets. It can be observed that, as the number of iterations increases, the objective function values for each dataset exhibit a monotonically decreasing trend and gradually stabilize within a finite number of steps. This indicates that the USMF optimization process achieves good convergence. This is because the model employs a block-wise alternating optimization strategy: with other variables held fixed, each sub-problem can obtain an exact or optimal solution, ensuring that no update increases the objective function value. Furthermore, since the objective function is bounded below under the imposed constraints, the algorithm is guaranteed to converge to a stable solution.

4.6 Application on the ORL dataset

Feature selection plays a significant role in image

Fig. 2 ACC on different α and γ .Fig. 3 NMI on different α and γ .

Fig. 4 ACC on different γ and δ .Fig. 5 NMI on different γ and δ .

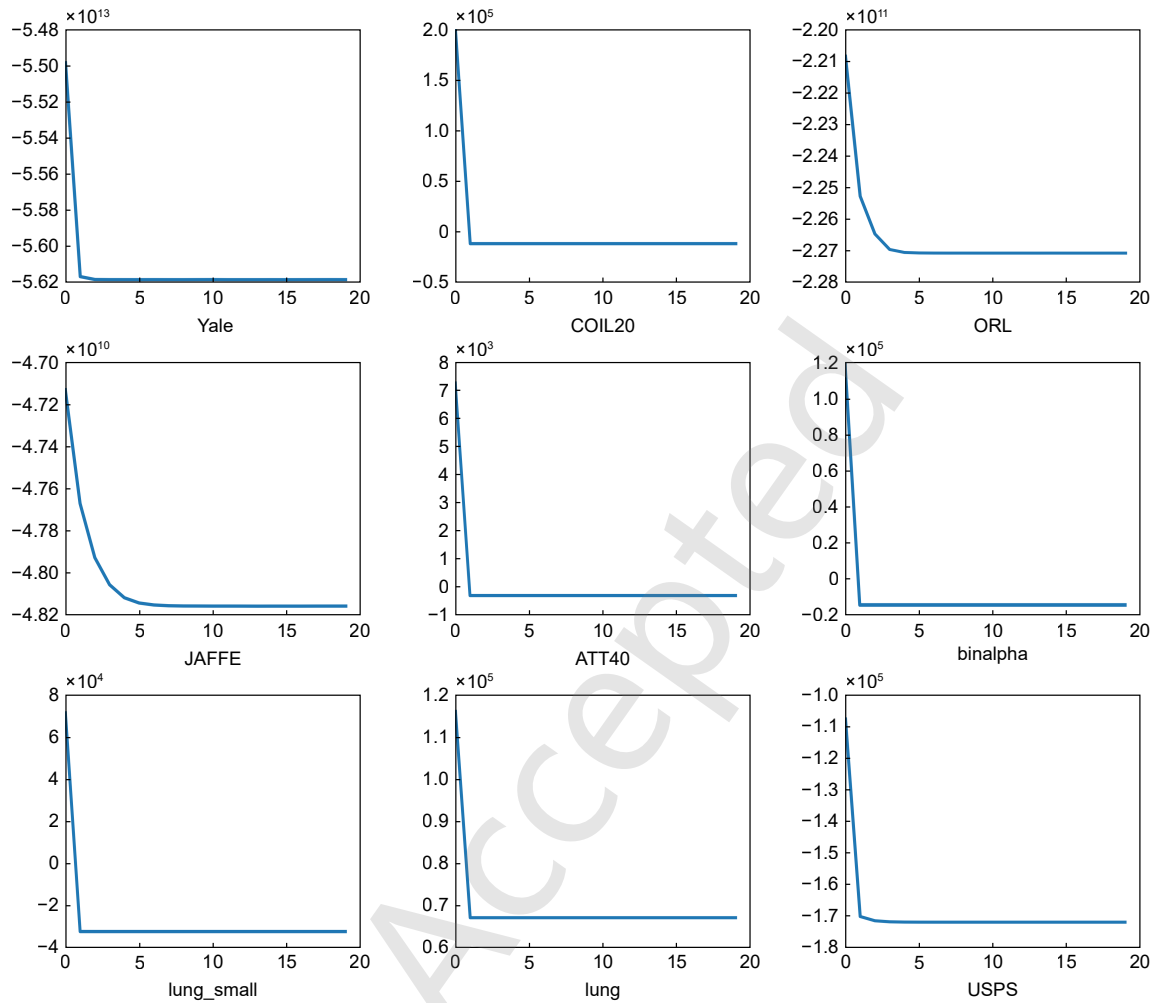


Fig. 6 Convergence plot of the objective function.

recognition tasks, particularly in the context of face recognition, where the objective is to identify the most discriminative key features from high-dimensional pixel spaces. To verify the effectiveness and interpretability of the proposed method in feature selection, we apply it to the ORL face dataset and randomly selected 10 facial samples for visual analysis.

Figure 7 presents the results when selecting the top 10%–40% of features. In the visualization, selected pixels are marked in red, while unselected ones retain their original grayscale values. As the proportion of selected features gradually increases, it can be observed that the model prioritizes representative regions within the images, including key facial structures such as the face contour, brow bones, eyes, nose, and mouth. These areas typically carry the primary discriminative information for face recognition.

These results demonstrates that the proposed method can effectively identify and preserve critical features

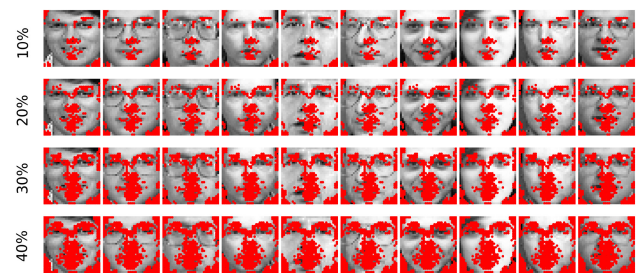


Fig. 7 The application on ORL.

related to facial structures, while suppressing redundant or background information. This indicates reflects its strong structural awareness and interpretability.

5 Conclusion

This paper proposes a unified structural modeling framework for unsupervised feature selection. By introducing a learnable structural matrix H , the method simultaneously preserves global structural information

and local geometric relationships, enabling structural consistency across different representation objectives. In addition, a F-norm-based structural consistency constraint and a row-sparsity regularization term are incorporated to enhance stability and enable effective feature selection. Experimental results on multiple benchmark datasets demonstrate that the proposed method achieves superior clustering performance compared to several representative approaches.

In the future, this framework will incorporate more efficient computational methods, such as sparse representations and low-rank approximation, to enhance the model's scalability. At the same time, in practical applications, data often exhibits multi-modal or multi-view characteristics; therefore, extending the unified-structure approach to multi-view or multi-source data scenarios is a direction worthy of further research. Furthermore, with the advancement of deep learning, combining unified structure modeling with deep clustering methods is another avenue that merits in-depth exploration. On the one hand, deep neural networks can be utilized to learn more discriminative non-linear representations, while introducing a unified structure matrix as a constraint in the latent space enables collaborative optimization from representation learning to structural modeling. On the other hand, structural consistency constraints can be embedded into deep clustering frameworks, where structural information guides the update of feature representations, thereby enabling the model to learn more stable and discriminative clustering representations under unsupervised conditions.

Acknowledgment

This work was supported in part by the Guangxi Natural Science Foundation (Nos. 2026GXNSFAA00640998 and 2026GXNSFDA00640006), the Scientific Research Basic Ability Improvement Project for Young and Middle-aged Teachers in Universities in Guangxi (No. 2025KY0834), the Special Research Project on the Development Strategy of Characteristic and Advantageous Interdisciplinary Disciplines of Guilin University of Aerospace Technology (No. TS2024452), the National Natural Science Foundation of China (No. 62372119), the Guangxi Bagui Young Talent Training Program, the Education and Teaching Reform Project of Guangxi Normal University (No. 2025JGZ14), and the Joint Cultivation Project of the National Natural Science Foundation of China for

Guangxi Normal University (No. 2025PY039).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

- [1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [2] T. Young, D. Hazarika, S. Poria, and E. Cambria, Recent trends in deep learning based natural language processing [review article], *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, 2018.
- [3] M. W. Libbrecht and W. S. Noble, Machine learning applications in genetics and genomics, *Nat. Rev. Genet.*, vol. 16, no. 6, pp. 321–332, 2015.
- [4] R. Wang, W. Li, K. Shen, T. Zhang, and X. Liao, Evolutionary multi-tasking optimization for high-efficiency time series data clustering, *Tsinghua Science and Technology*, vol. 29, no. 2, pp. 343–355, 2024.
- [5] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton: Princeton University Press, 2015.
- [6] M. Verleysen and D. François, The curse of dimensionality in data mining and time series prediction, in *Computational Intelligence and Bioinspired Systems*, Barcelona, Spain, 2005, pp. 758–770.
- [7] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
- [8] X. Zhang, B. Xie, Y. Wang, and J. Chen, Prototype-based feature representation learning for image deep clustering, *Tsinghua Science and Technology*, vol. 31, no. 5, pp. 2353–2365, 2026.
- [9] X. He and P. Niyogi, Locality preserving projections, in *Proceedings of Advances in Neural Information Processing Systems 16 (NIPS 2003)*, Cambridge, MA, USA, 2003, pp. 153–160.
- [10] C. Zou, G. Lu, L. Du, X. Zeng, and S. Lin, Graph similarity learning for cross-level interactions, *Inf. Process. Manag.*, vol. 62, no. 1, p. 103932, 2025.
- [11] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [12] X. Lv, Z. Su, and B. Chai, Regularized semi-supervised subspace clustering with local-structure-preservation-based soft-MFA, *Tsinghua Science and Technology*, vol. 31, no. 3, pp. 1661–1677, 2026.
- [13] H. Hotelling Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
- [14] Z. Zhao and H. Liu, Spectral feature selection for supervised and unsupervised learning, in *Proc. 24th Int. Conf. Machine Learning*. Corvallis Oregon USA, 2007, pp.

- 1151–1157.
- [15] R. Shang, K. Xu, and L. Jiao, Subspace learning for unsupervised feature selection via adaptive structure learning and rank approximation, *Neurocomputing*, vol. 413, pp. 72–84, 2020.
- [16] T. Yang, L. Zhao, J. Zhao, S. Ma, Q. Xie, and Y. Yuan, AGPVC: Partially view-aligned clustering with anchor graph for multi-view data, *Tsinghua Science and Technology*, doi:10.26599/TST.2025.9010104.
- [17] C. Zou, G. Lu, W. Zhang, X. Zeng, S. Lin, L. Du, and S. Zhang, Enhanced graph similarity learning via adaptive multi-scale feature fusion, in *Proc. Thirty-Fourth Int. Joint Conf. Artificial Intelligence*, Montreal, Canada, 2025, pp. 7309–7317.
- [18] L. Van der Maaten, E. Postma, and H. Van Den Herik, Dimensionality reduction: A comparative review, *J. Mach. Learn. Res.*, vol. 10, pp. 1–41, 2009.
- [19] J. P. Cunningham and Z. Ghahramani, Linear dimensionality reduction: Survey, insights, and generalizations, *J. Mach. Learn. Res.*, vol. 16, no. 89, pp. 2859–2900, 2015.
- [20] J. B. Tenenbaum, V. de Silva, and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [21] H. Abdi and L. J. Williams, Principal component analysis, *Wires Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.
- [22] H. M. Ebied, Feature extraction using PCA and Kernel-PCA for face recognition, in *Proceedings of 2012 8th International Conference on Informatics and Systems (INFOS)*, Giza, Egypt, 2012, pp. MM–72–MM–77.
- [23] G. Wen, Y. Zhu, L. Chen, M. Zhan, and Y. Xie, Global and local structure preservation for nonlinear high-dimensional spectral clustering, *Comput. J.*, vol. 64, no. 7, pp. 993–1004, 2021.
- [24] C. Zhang, Y. Fang, X. Liang, H. Zhang, P. Zhou, X. Wu, J. Yang, B. Jiang, and W. Sheng, Efficient multi-view unsupervised feature selection with adaptive structure learning and inference, in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, Jeju, Korea, 2024, pp. 5443–5452.
- [25] Y. Wang, J. Zhang, T. Zhou, B. Jiang, and J. Chen, Structural feature selection in common spatial patterns using adaptive sparse group Lasso, *CAAI Trans. Intell. Technol.*, vol. 11, no. 2, pp. 367–384, 2026.
- [26] B. Jiang, J. Liu, Z. Wang, C. Zhang, J. Yang, Y. Wang, W. Sheng, and W. Ding, Semi-supervised multi-view feature selection with adaptive similarity fusion and learning, *Pattern Recognit.*, vol. 159, p. 111159, 2025.
- [27] G. Lu, F. Ling, J. Li, L. Zhu, X. Qin, and S. Nong, Graph attention-based dual enhancement for multiview clustering, *IEEE Trans. Comput. Soc. Syst.*, vol. 12, no. 5, pp. 2726–2735, 2025.
- [28] Z. Wu, P. Zhou, G. Wen, and X. Zhu, A pseudo-labeling approach based on knowledge distillation for graph few-shot learning, *Inf. Process. Manag.*, vol. 62, no. 6, p. 104268, 2025.
- [29] F. Nie, Z. Ma, J. Wang, and X. Li, Fast sparse discriminative k-means for unsupervised feature selection, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 9943–9957, 2024.
- [30] J. Wang, H. Wang, F. Nie, and X. Li, Sparse feature selection via fast embedding spectral analysis, *Pattern Recognit.*, vol. 139, p. 109472, 2023.
- [31] S. Xu, X. Xie, and Z. Cao, Graph-regularized consensus learning and diversity representation for unsupervised multi-view feature selection, *Knowl. Based Syst.*, vol. 311, p. 113043, 2025.
- [32] X. Dong, F. Nie, L. Tian, R. Wang, and X. Li, Unsupervised discriminative feature selection with ℓ_2 , p -norm constrained sparse projection, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 10, pp. 8321–8335, 2025.
- [33] K. Jiang, T. Cao, L. Zhu, and Q. Sun, Adaptive and flexible ℓ_1 -norm graph embedding for unsupervised feature selection, *Appl. Intell.*, vol. 54, no. 22, pp. 11732–11751, 2024.
- [34] G. Vardakas, I. Papakostas, and A. Likas, Deep clustering using the soft silhouette score: Towards compact and well-separated clusters, *Mach. Learn.*, vol. 115, no. 4, p. 81, 2026.



published more than 30 academic papers in international reputable academic journals.



Guoqiu Wen received the PhD degree in software engineering from Guangxi Normal University, Guilin, China in 2022 and now is employed as an associate professor with Guangxi Normal University. Her research interests include machine learning, feature learning, graph learning, and data mining. She has

Xinjie Han received the BEng degree from Hebei Normal University, Hebei, China in 2019. She is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, Guilin, China. Her research interests include machine learning and data mining.



published more than 10 academic papers in international reputable academic journals.

Peng Liu received the PhD degree in ecology from Guangxi University, Nanning, China in 2023 and now is employed as a lecturer in aerospace technology with Guilin University, Guilin, China. His research interests include artificial intelligence, spatial information processing, and 3S technology. He has



Jiayang Su received the MS degree in computer science from Guangxi Normal University, Guangxi, China, in 2025. He is currently pursuing the PhD degree at School of Computer Science and Technology, Hainan University, Haikou, China. His research interests include machine learning, graph learning, and data mining.



Yuanpeng Li received the BEng degree in computer science and technology from Hubei University of Education, Wuhan, China in 2025. He is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, Guilin, China. His research interests include large language model applications.



Jiale Liu received the BEng degree in internet of things engineering from Nanyang Institute of Technology, Nanyang, China in 2023. He is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, Guilin, China. His research interests include graph autoencoder clustering methods based on prompt learning.



Shuyu Li received the BEng degree in information security from Guangxi Normal University, Guilin, China in 2023. She is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software / School of Artificial Intelligence, Guangxi Normal University, Guilin, China. Her research interests include multi-view clustering, fuzzy clustering, and data mining.



Cong Lei received the MEng degree from Guangxi Normal University, Guilin, China in 2018 and the PhD degree in engineering from University of Sydney, Sydney, Australia in 2025. His research interests include quantum machine learning



Qinghong Jiang received the MS degree from Guangxi Normal University, Guilin, China in. He is currently a lecturer at School of Electronic and Information Engineering and is pursuing a PhD degree at School of Computer Science and Engineering. His research interests include deep clustering and intelligent control.