Intelligent Multimedia Processing & Communication Systems Journal

IMPCSJ

J IMPCS (2025) 20: 51-63
DOI 10.71856/IMPCS.2025.1212662

Research Paper

# AI-Driven Hyperspectral Image Classification using Low-Rank Representation and Spatial-Spectral Information

**Fatemeh Hajiani***

Assistant Professor, Department of Electrical Engineering, Khorm.C., Islamic Azad University, Khormoj, Iran
**Corresponding Author*, fatemeh.hajiani@iau.ac.ir

**Article Info**

**Article history:**
Received: 10 May 2025
Accepted: 21 June 2025
DOR:

**Keywords**:
Artificial intelligence,
Classification,
Dictionary learning,
Hyperspectral image,
Low-rank representation.

**ABSTRACT**

Hyperspectral image (HSI) classification is one of the most important processes on these images, which artificial intelligence (AI) techniques have recently achieved significant success in this process. Data representation using a low-dimensional subspace is critical in classification task of HSI. By employing the Low-Rank Representation (LLR) approach, low-dimensional representations from data can be effectively extracted. Since this method neglects local information, the extracted features are not sufficiently rich and informative for classification. This paper proposes a machine learning method for hyperspectral image classification, which involves employing the Structure of the Data Regularized LLR with Dictionary Learning (SDLRRDL) model. Our AI-based model, SDLRRDL, presents an approach for learning data structures through a low-rank and sparse representation. Also, to leverage structural data information, a penalty is added to the low-rank representation model. The method can create similar features for data of the same class by combining image class signature and spectral-spatial information. Moreover, image samples are represented through a linear combination of dictionary atoms. Rich and informative features are extracted through the trained dictionary utilizing a training data set that better matches the training content. Then, extracted features are classified using the support vector machine with high accuracy. Simulation results demonstrate that the proposed method has a superior classification accuracy compared to state-of-the-art methods on three popular HSI datasets. The proposed method improves the classification accuracy of the state-of-the-art methods more than 2.13, 0.2, and 0.6 percent on the Indian Pines, Pavia University, and Salinas datasets, respectively.

## I. Introduction

With advancements in remote sensing, high spectral resolution sensors, called hyperspectral sensors, have emerged. These sensors make the Earth surface imaging in several hundred spectral bands possible and provide more details of image classes [1]. The image captured by these sensors in several electromagnetic wavelengths is called a Hyperspectral Image (HSI). The classification is a significant process on the HSI [2]. So far, Various methods have been proposed for HSI classification. Recently, machine learning methods have been popular in the HSI classification. In this regard, numerous Support Vector Machine (SVM) methods have been developed [3]. SVM is a commonly used classifier in most classification problems, achieving significant success in HSI classification. Furthermore, deep learning is recognized as a formidable tool for hyperspectral image processing [4-6].

Hyperspectral images contain valuable information across different wavelengths for the classification task. Methods outlined in [7] and [8] employ spectral information to perform classification. Although rich spectral information for each pixel greatly aids in HSI classification, solely utilizing spectral information is insufficient. The majority of pixels surrounding a central pixel are typically of the same class. Thus, combining spectral and spatial information improves the classification performance. Using spectral and spatial information simultaneously improves classification performance and causes notable success in HSI classification [9]. In [10], the authors propose a Spatial Spectral Transformer Network (SSTN) for HSI classification. The SSTN method employs attention and association modules as spectral-spatial modules for feature extraction. Combining these modules with a structured factorized search framework forms the SSTN to carry out HSI classification. In [11], a Groupwise Separable Convolutional Vision Transformer (GSC-ViT) HSI classification method is proposed as a light vision transformer network. GSC-ViT employs a group-wise separated convolutional module to reduce the number of convolutional kernel parameters and extract local spectral-spatial information from the hyperspectral images. Also, a group-wise separated multi-head associate module is utilized in the vision transformer network. In [12], the authors present the state-space morphological model (SSMM) that exploits morphological analysis to calculate the structure and shape of data and extract spatial and spectral features from HSI. Then, they exploit a multi-head self-attention and a state-space model for HSI classification.

Hyperspectral images are represented as spectral vectors across different wavelengths, containing high amounts of information, and contaminated with noise. Thus, sparse representation approaches can be highly effective in processing the HSI. Accordingly, HSI classification through sparse representation has gained popularity in recent years. These methods typically represent samples as a linear combination of a few dictionary elements [13]. The authors in [14] utilize spatial information to achieve suitable results. They propose the Joint Robust Sparse Representation Classifier (JRSRC) method. This method represents neighboring pixels of a pixel using a sparse linear combination of several training samples. In [15], the authors present a Class-Dependent Sparse Representation Classifier (cdSRC) that utilizes distance and correlation between samples. They combine a sparse representation classifier with a k-nearest neighbor classifier, and propose a class membership function that utilizes Euclidean distance information. Despite the suitable results, sparse representation classification methods are only able to take the local data structure into account.

An effective approach to extract features with lower dimensions is Low-Rank Representation (LRR) model [16]. The primary concept of the LRR is to utilize the most parsimonious information to approximate and reconstruct signals [17-19]. This approach seeks to minimize the rank of each sample by leveraging information across the entire image dataset, essentially taking the global data structure into account. HSI is captured by sensors as signals and represented by spectral vectors across different frequencies. The images contain noise and high amounts of information. Thus, they can be represented sparsely with fewer elements and minimal noise. The low-rank representation method does not consider the local structure and correlation between pixels. Moreover, the features extracted using low-rank representations are not highly distinct. Thus, presenting a model through combining local and global data structures separates the highly distinguishable data by combining low-rank and sparse representation models [20]. In [21], a Low Rank Group Inspired Dictionary Learning (LGIDL) model is proposed to extract features from an image using spatial information. The authors in [20] combine the geometric structure of data with low-rank and sparse representations models and apply an appropriate graph. It causes the extracted features to be highly distinct and informative. The authors in [22] propose a classification approach for HSI called the superpixel-wise low-rank approximation-based partial label learning (SLAP) method. In this method, the low-rank model is combined with the regularized Laplacian, and the model is applied to each superpixel. Then, high-discriminant features are extracted using the affinity graph extraction.

Dictionary learning has a critical role in numerous research fields, including classification, compression, denoising, and super-resolution. Dictionary learning helps extract valuable information for processing from a large amount of redundant and correlated data [23]. A dictionary consists of a set of atoms, with each sample expressed as their linear combination. This representation scheme extends to low-rank methods, where samples are constructed by combining linear combinations of dictionary members.

Selecting an appropriate dictionary has a significant impact on extracting distinctive features from HSI. Most of the low-rank representation approaches consider all training samples as the dictionary. These atoms contain redundant information, which increases the computational complexity. Dictionary learning methods include Structured dictionaries based on mathematical models [24-26] and Trained dictionaries based on training samples [27-29]. Table I shows a comparison of prior works.

As previously mentioned, the majority of neighboring pixels of a pixel are from a similar class, and combining their spectral and spatial information improves classification accuracy. Extracting spatial-spectral information enhances the HSI classification performance. The neighboring pixel is employed to extract its spatial information. In this paper, we employ a shape-adaptive method to calculate pixel neighborhoods. Unlike methods that utilize fixed-size patches, this approach uses an adaptive size to select pixel neighborhoods in each direction. Therefore, the neighborhood is determined based on the spatial information of the pixels. The spectral-spatial information is extracted using statistical moments of the adaptive neighborhoods.

The proposed method presents an optimization algorithm for learning data structure employing a low-rank, sparse representation and appropriate dictionary learning. This method uses a combination of image class information and

TABLE I Comparison of prior works

| Study | method | Key Innovation | limitation | Dataset |
|---|---|---|---|---|
| Camps et al. [3] | Support vector Machine (SVM) | Finding the optimal decision boundary between classes | Sensitive to noisy data, overlapping Classes, and Not Suitable for very large datasets | Indian Pines |
| Li et al. [14] | JRSRC | Combines SRC+ spectral and spatial information | Ignores global structure | Indian Pines, Pavia University |
| Liu et al. [16] | LRR | Global data modeling | Poor local feature extraction | Indian Pines, Pavia University, Salinas |
| He et al. [21] | LGIDL | Super-pixel segmentation | Degradation in global structure preservation and inaccurate local region definitions for LRR in object edges | Indian Pines, Pavia University, Salinas |
| Yang et al. [22] | SLAP | Laplacian graph regularized LRR | The need to build an appropriate adjacency graph and computational complexity in large-scale data | Indian Pines, Salinas |
| Ahmad et al. [12] | SSMM | State space models | Low accuracy for limited and noisy training samples | Indian Pines, Pavia University |
| Zhong et al. [10] | SSTN | Spatial-spectral attention network | Low accuracy for limited and noisy training samples | Indian Pines, Pavia University |
| Zhao et al. [11] | GSC-ViT | Integrating the grouping strategy characteristics into the transformer network | Low accuracy for limited and noisy training samples | Indian Pines, Pavia University, Salinas |
| Proposed | SDLRRDL (AI hybrid) | Combines LRR+ Sparse+ dictionary learning +spectral and spatial information | Computationally intensive | Indian Pines, Pavia University, Salinas |

takes the data structure into account to create similar features for data from the same class. To this end, training sample information is combined with the model as a regularization term for dictionary learning. In contrast to the methods that employ fixed dictionaries or utilize the entire data as dictionaries for data reconstruction, the proposed method uses a set of training samples as dictionary components. This approach applies dictionary learning algorithms and utilizes appropriate constraints to extract highly distinctive features. Subsequently, a support vector machine classifies the features with high accuracy. We call the proposed AI-based method the Structure of the Data Regularized LLR with Dictionary Learning (SDLRRDL) method.

The following sections organize the paper. The SDLRRDL method is discussed in Section II of the paper. In Section III, the optimization algorithm of the SDLRRDL method is outlined. Section IV examines the performance of the SDLRRDL approach, and Section V presents the conclusions of this article.

## II. The Proposed SDLRRDL Method

The pixels in close proximity to one another in remote sensing images are typically composed of the same materials. Therefore, the informative features are extracted by exploiting the information of neighboring pixels. The proposed method employs a shape-adaptive approach to determine the neighborhood of each pixel across different directions. The spectral and spatial features are then extracted using information from neighboring pixels.

Hyperspectral images are represented as spectral vectors across different frequencies. Thus, HSI contains valuable information contaminated with noise. Therefore, sparse representation methods can be quite effective in processing HSI. Furthermore, the high spectral similarities of HSI data are related to the low-rank characteristic. It is important to note that sparse representation considers the data's local structure, while Low-rank representation takes its global structure into account.

In the proposed method, we combine the low-rank and the sparse representation models to extract features, employing spectral-spatial image information. This approach represents the image sample through a linear combination of dictionary atoms. Features with similar representation structures for data from the same classes can be created by adding a structure penalty term as a regularization term to the objective function and utilizing image class information. The proposed method performs dictionary learning using a set of training data, presenting a dictionary that better matches the training content. Ultimately, the extracted features are classified in a supervised manner using a support vector machine classifier. The subsequent section outlines the low-rank and sparse representation model combined with a structure penalty term to apply structural data features and dictionary learning to the model.

### A. Extracting Spatial-Spectral Feature from Adaptive Neighbor

The hyperspectral image consists of the $Y \epsilon R^{d \times n_1 \times n_2}$ dataset with $n_1 \times n_2$ pixels and $d$ bands. Each pixel's data is considered as a vector across different spectral bands. Thus, the hyperspectral image can be considered as the $[y_1 . y_2 . \ldots . y_n] \epsilon R^{d \times n}$ matrix. In this context, $n$ indicates pixels number, and column $y_i$ represents the $i$th pixel's spectral vector. The adjacent pixels in images are typically composed of the same materials. Therefore, valuable features can be extracted by using the information of neighboring pixels. In this paper, we employ a shape-adaptive method for pixel neighborhood selection [30]. This method considers a set of directional convolution kernels with different sizes across eight different directions $\left\{\frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \ldots, 2\pi\right\}$ relative to the central pixel. These kernels are then applied to the first principal component of the HSI to determine the pixel's neighborhood. The set of neighboring pixels of $\boldsymbol{y}_i$ is denoted as $\mathcal{N}_i$, which includes the pixel $\boldsymbol{y}_i$ as well. The statistical information of each pixel's neighborhood is employed to extract its spectral-spatial information. The majority of the neighboring pixels are from similar classes. Calculating these moments provides useful information regarding their corresponding class. To this end, we employ the first and second moments of neighboring pixels as follows:

$$\begin{aligned} \boldsymbol{\mu}_i &= \frac{1}{|\mathcal{N}_i|} \sum_{\boldsymbol{y} \in \mathcal{N}_i} \boldsymbol{y} \\ [\boldsymbol{\sigma}_i^2]_j &= \frac{1}{|\mathcal{N}_i|} \sum_{\boldsymbol{y} \in \mathcal{N}_i} \left([\boldsymbol{y}]_j - [\boldsymbol{\mu}_i]_j\right)^2. \end{aligned} \quad (1)$$

In this context, $|\mathcal{N}_i|$ indicates the number of elements in the set $\mathcal{N}_i$, and $[\boldsymbol{y}]_j$ denotes the j-th element of the vector $\boldsymbol{y}$. Thus, the spatial-spectral feature of the $\boldsymbol{y}_i$ pixel is calculated as follows:

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{\mu}_i \\ \boldsymbol{\sigma}_i^2 \end{bmatrix}. \quad (2)$$

Hence, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$ serves as a spatial-spectral HSI characteristic.

### B. The Proposed Model

This section involves utilizing low-rank representation as an efficient technique to obtain a lower-rank structure within the data. The rank minimization problem approximates the sample matrix with a matrix of lower rank. Generally speaking, most valuable signal information is located in a subspace with fewer dimensions in high-dimensional data. Should the observation matrix become corrupted by noise, it is possible to obtain data with minimal noise using the low-rank matrix denoted by $Z$ [16]. The data consists of multiple subspaces, with each sample expressed by a linear combination of dictionary components as $X = DZ + E$. In this context, the columns of $D = [d_1, d_2, \ldots, d_m] \epsilon R^{d \times m}$ correspond to a set of dictionary bases, $E$ represents the model error, and $Z = [z_1, z_2, \ldots, z_n]$ is the low-rank matrix representing of $\boldsymbol{X}$. Essentially, a dictionary consists of some basis atoms that can represent any data related to that dictionary as a linear combination of its atoms.

Thus, the LLR method calculates the low-rank matrix $Z$ as follows:

$$\min_{Z.E} rank(Z) + \lambda \|E\|_0, \quad s.t. \quad X = DZ + E, \quad (3)$$

where $\lambda$ represents a parameter controlling the noise effect.

Solving this model, containing the rank function and $l_0$ norm, is an NP-hard problem. Thus, the following convex model is employed to solve the optimization problem (3):

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_1, \quad s.t. \ X = DZ + E, \quad (4)$$

where $\|\cdot\|_*$ represents the nuclear norm. This norm is equivalent to the summation of singular values, and $\|\cdot\|_1$ represents the sparsity [16].

As previously stated, the LLR model takes the global structure into account. Consequently, by considering the dictionary, we formulate a joint low-rank and sparse representation through the following objective function to represent the local structure alongside the global structure:

$$\begin{aligned} &\min_{Z.E.D} \|Z\|_* + \beta \|Z\|_1 + \lambda \|E\|_1, \\ &s.t \ X = DZ + E, Z \geq 0, \forall i \ \|d_i\|_2 \leq 1, d_i \geq 0, \end{aligned} \quad (5)$$

where, $\beta$ and $\lambda$ are the parameters balancing the regularization sections, while $\boldsymbol{d}_i$ represents the $i$-th atom in the dictionary. The $\boldsymbol{Z} \geq 0$ constraint is expressed to correctly describe the physical context. Without loss of generality, if the samples are sorted according to their classes, the $Z$ matrix will have a block and diagonal structure as follows:

$$Z = \begin{bmatrix} z_1^* & 0 & \cdots & 0 \\ 0 & z_2^* & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & z_c^* \end{bmatrix}. \quad (6)$$

In this context, c represents the number of classes, while $Z_1$ and $Z_2$ represent the submatrices of the feature matrix that correspond to train and test samples, respectively. This study defines a structured matrix $S$ as a constraint on $Z$, keeping Z as close to matrix S as possible, ensuring the structure of equation (6) is also maintained for $Z_1$ and $Z_2$ matrices. The $S = [S_1, S_2]$ term is composed of two sub-matrices, where $S_1$ corresponds to the training data and $S_2$ corresponds to the test data. It must be noted that $S_1$ utilizes training data information. Thus, the structure in equation (6) is maintained for $Z_1$. However, it is challenging to obtain a similar structure in equation (6) for $Z_2$, as there is a lack of information regarding test sample labels. To this end, the following similarity measure is employed to generate the $S_2$ matrix. If $[p_i, q_i]$ represent the coordinates of the training pixel $\boldsymbol{x_i}$ and $[p_j, q_j]$ represent the coordinates of the test pixel $\boldsymbol{x_j}$, the similarity measure is defined as follows:

$$\begin{aligned} d_{s_{ij}} &= \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2, \\ d_{c_{ij}} &= \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2}, \\ D_{ij} &= {d_{s_{ij}}}^2 + \zeta {d_{c_{ij}}}^2, \\ [S_2]_{ij} &= \begin{cases} \exp(-D_{ij}), & D_{ij}^2 < \beta \\ 0, & D_{ij}^2 > \beta, \end{cases} \end{aligned} \quad (7)$$

where, $d_{s_{ij}}$ and $d_{c_{ij}}$ represent spectral and spatial distances of training and test pixels, respectively, while $D_{ij}$ expresses the spatial-spectral distance of $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ pixels. Furthermore, $\zeta$ is the coefficient of spatial distance impact on the similarity measure. The $\beta$ threshold determines the sparsity level of the $S_2$ matrix.

Thus, the $\|Z - S\|_F^2$ constraint is added to the model to impose the block and diagonal form on the $Z$ matrix. The low-rank representation $Z$ will have an almost similar structure to equation (6), with the final proposed model being expressed as follows:

$$\begin{aligned} &\min_{Z,E,D} \|Z\|_* + \beta\|Z\|_1 + \lambda\|E\|_1 + \alpha\|Z - S\|_F^2 \\ &s.t\ X = DZ + E, Z \geq 0, \forall i\ \|d_i\|_2 \leq 1, d_i \geq 0, \end{aligned} \quad (8)$$

where, $\alpha$ is a penalty parameter that controls the regularization term.

### C. Dictionary Learning

Dictionary selection is critical in feature extraction. A fixed dictionary is ineffective in extracting a representation of highly distinctive coefficients in the various classes. Thus, the present study employed a dictionary learning method. A weakness of the majority of low-rank representation methods lies in the fact that they use all image samples as a dictionary, which creates atoms with redundant information and increases computational complexity. To address this issue, the proposed method solely employs training samples to create the dictionary. This approach extracts highly distinctive features.

The dictionary atoms $\boldsymbol{d}_i$ and data representation coefficients $Z$ constitute the learning parameters in the optimization model. Typically, determining these parameters involves the following two steps. During the first step, a default dictionary is defined in the proposed algorithm using training samples, and $E$ and representation coefficients of $Z$ are calculated. Subsequently, the representation coefficients obtained from the first step are used to update the dictionary atoms $D$. These steps are iterated until the learning algorithm converges.

Now, the highly distinctive features, which contain rich spectral-spatial information, are extracted. A classifier is needed to classify these features. The classifier maps the feature space to a class number. In this paper, we utilized the Support Vector Machine (SVM) for assigning labels to the features. SVM is a binary classifier that separates classes using linear boundaries. This algorithm locates the optimal margin or hyperplane for pixel classification within a multidimensional space. These margins are selected to maximize the distance from the closest pixels, which are called support vectors. This classifier performs well in high-dimensional spaces and can also separate non-linear data using kernels [3].

The following section outlines the model optimization method.

## III. Optimization Algorithm of the Proposed SDLRRDL Method

As previously mentioned, the SDLRRDL method can be defined as follows:

$$\begin{aligned} &\min_{Z,E,D} \|Z\|_* + \beta\|Z\|_1 + \lambda\|E\|_1 + \alpha\|Z - S\|_F^2 \\ &s.t\ \ X = DZ + E, Z \geq 0, \forall i\ \|d_i\|_2 \leq 1, d_i \geq 0 \end{aligned} \quad (9)$$

Initially, the training samples are utilized as a primary dictionary for problem (9). Subsequently, this equation and the initial dictionary are used to calculate $Z$ and $\boldsymbol{E}$. The new dictionary stage is obtained by keeping $Z$ and $E$ constant. $E$, $Z$, and $D$ will be iteratively Calculated until convergence is achieved.

There are various methods to solve low-rank problems [31-33]. The Linearized Alternating Direction Method with Adaptive Penalty (LADMAP) approach [34] uses fewer auxiliary variables compared to the other methods [35], and

achieves faster convergence without any inverse matrix. The present study employs the LADMAP algorithm to solve the problem of SDLRRDL.

A. Optimizing Z and E for a Constant D

The equation (9) is reformulated by defining the auxiliary variable $V$ as follows:

$$\begin{aligned} &\min_{Z.E.D} \|Z\|_* + \beta\|V\|_1 + \lambda\|E\|_1 + \alpha\|V-S\|_F^2 \\ &s.t\ X = DZ + E.V = Z, V \geq 0, \forall i\ \|d_i\|_2 \leq 1, d_i \geq 0 \end{aligned} \tag{10}$$

Subsequently, the augmented Lagrangian function of problem (10) is defined as follows:

$$\begin{aligned} &L(Z.V.E.Y_1.Y_2) = \|Z\|_* + \beta\|V\|_1 + \lambda\|E\|_1 \\ &+\alpha\|V-S\|_F^2 + \langle Y_1, X - DZ - E_c\rangle + \langle Y_2.Z - V\rangle \\ &+\frac{\mu}{2}(\|X - DZ - E_c\|_F^2 + \|Z-V\|_F^2), \end{aligned} \tag{11}$$

where, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ represent Lagrange multipliers, while $\mu > 0$ is the penalty parameter. The other variables are then kept constant, and the Lagrangian complementary function is minimized with iterative updates to obtain each unknown variable. The representation matrix Z is obtained by minimizing the following function:

$$\begin{aligned} L_1 = \|Z\|_* + \frac{\mu}{2}\Bigg(&\left\|X - DZ - E^k + \frac{1}{\mu}Y_1^k\right\|_F^2 \\ &+ \left\|Z - V^k + \frac{1}{\mu}Y_2^k\right\|_F^2\Bigg). \end{aligned} \tag{12}$$

This operation is performed using the LADMAP algorithm as follows:

$$\begin{aligned} Z^{k+1} = \underset{Z}{argmin}\|Z\|_* + \langle \nabla_Z f(Z^k), Z - Z^k\rangle \\ + \frac{\eta\mu_k}{2}\|Z - Z^k\|_F^2, \end{aligned} \tag{13}$$

where, $f$ is calculated as follows:

$$\begin{aligned} &f(Z, E^k, V^k, Y_1^k, Y_2^k) = \\ &\frac{\mu}{2}\Bigg(\left\|X - DZ - E^k + \frac{1}{\mu}Y_1^k\right\|_F^2 \\ &+ \left\|Z - V^k + \frac{1}{\mu}Y_2^k\right\|_F^2\Bigg), \end{aligned} \tag{14}$$

where, $\nabla_Z f$ represents the gradient of $f$ with respect to $Z$. Thus, $Z$ is updated as follows:

$$\begin{aligned} &Z^{k+1} = \Theta_{(\eta\mu_k)^{-1}}[Z^k - \nabla_Z f(Z^k)/\eta\mu_k] \\ &\nabla_Z f(Z^k) = \left[\mu_k\left(Z^k - V^k + \frac{\boldsymbol{Y}_2^k}{\mu_k}\right)\right. \\ &\left. - \mu_k \boldsymbol{D}^T\left(X - \boldsymbol{D}Z - E^k + \frac{\boldsymbol{Y}_1^k}{\mu_k}\right)\right] \end{aligned} \tag{15}$$

The equation $\Theta_\tau(C) = US_\tau(\Sigma)V^T$ is called singular value thresholding (SVT). Furthermore, the equation $S_\tau(x) = sgn(x)\,max(|x| - \tau, 0)$ is the soft thresholding operator, and $U\Sigma V^T$ denotes the singular value decomposition .Thus, $E$ and $V$ are calculated as follows:

$$\begin{aligned} &V^{k+1} = \underset{J\geq 0}{arg\,min}\,\beta\|V\|_1 + \alpha\|V - S\|_F^2 + \\ &+\langle Y_2^k.Z - V\rangle + \frac{\mu_k}{2}\left\|Z^{k+1} - V\right\|_F^2 \\ &= max\left\{S_{\frac{\beta}{2\alpha+\mu}}\left(\frac{2\alpha}{2\alpha+\mu}S + \frac{1}{2\alpha+\mu}Y_2^k + \frac{\mu}{2\alpha+\mu}Z^{k+1}\right), 0\right\} \end{aligned} \tag{16}$$

$$\begin{aligned} &E^{k+1} = \underset{E}{arg\,min}\,\lambda\|E\|_1 \\ &+\frac{\mu_k}{2}\left\|E - \left(X - \boldsymbol{D}Z^{k+1} + \frac{Y_1^k}{\mu_k}\right)\right\|_F^2 \\ &= S_{\frac{\lambda}{\mu_k}}\left(X - DZ^{k+1} + \frac{\boldsymbol{Y}_1^k}{\mu_k}\right) \end{aligned} \tag{17}$$

$Y_1$ and $Y_2$ Lagrange multipliers are updated using the following equations:

$$\begin{aligned} Y_1^{k+1} &= Y_1^k + \mu_k\left(\mathrm{X} - \mathrm{DZ^k} - \mathrm{E^k}\right), \\ Y_2^{k+1} &= Y_2^k + \mu_k(Z^k - V^k), \end{aligned} \tag{18}$$

And convergence is determined as follows:

$$\|X - DZ - E\|_\infty < \varepsilon \quad , \quad \|Z - V\|_\infty < \varepsilon \tag{19}$$

B. Optimizing $D$ for a Constant $Z$

This subsection involves updating the dictionary matrix $D$ while keeping the feature matrix $Z$ constant. The optimization problem is as follows:

$$\begin{aligned} &\min_D \frac{1}{2}\|X - DZ\|_1 \\ &s.t.\ \forall i \quad \|d_i\|_2 \leq 1, d_i \geq 0 \end{aligned} \tag{20}$$

Equation (20) has a quadratic form with respect to the matrix $D$, and constrains the dictionary atoms to a ball with a radius of one. The matrix $D$ is obtained as follows by solving equation (20):

$$D = XZ^T(ZZ^T)^{-1} \tag{21}$$

In equation (21), the inverse calculation of $ZZ^T$ is computationally complex. Thus, a Block Coordinate Descent (BCD) [36] approach is used to solve equation (20) to iteratively update the atoms $\boldsymbol{d}_i$ of the dictionary $D$. $\boldsymbol{R}_i$ is calculated using the equation $\boldsymbol{R}_i = X - \sum_{j\neq i}\boldsymbol{d}_j Z_T^{j^T}$, where $\boldsymbol{Z}_T^j$ represents the $j$-th row of $Z$ in its column form, and the objective function is obtained using the $i$-th dictionary atom as $\left\|\boldsymbol{R}_i - \boldsymbol{d}_i Z_T^{i^T}\right\|_1$. The objective function (20), taking into account the norms concerning $d_i$ is rewriting as follows:

$$\begin{aligned} &\min_{d_i}\frac{1}{2}\left\|d_i - \frac{R_i Z_T^i}{\left\|Z_T^i\right\|_2^2}\right\|_1 \\ &s.t.\ \|d_i\|_2 \leq 1, d_i \geq 0, \end{aligned} \tag{22}$$

where, $d_i$ is calculated as follows:

$$d_i = proj_{l_2^+}\left(\frac{R_i Z_T^i}{\left\|Z_T^i\right\|_2^2}\right), \tag{23}$$

where, $proj_{l_2^+}(\boldsymbol{x})$ represents the projection of $\boldsymbol{x}$ onto a ball with a radius of one in the $l_2$ norm unit. The convergence condition for optimization is as follows:

$$\|D^{k+1} - D^k\|_\infty < \varepsilon \quad (24)$$

The following section outlines the details of simulating the proposed method.

## IV. Experimental result

This section discusses the simulation details and evaluates the effectiveness of the proposed method. First, we introduce the HSI datasets, followed by presenting several comparative methods to evaluate our proposed approach. Finally, the performance of the proposed method will be compared with state-of-the-art methods based on various classification methods. Finally, we comprehensively evaluate SDLRRDL's classification accuracy against state-of-the-art approaches using multiple performance metrics. In this Paper, experiments are executed on a single Intel Core i7 CPU in a MATLAB 2024 b environment. The penalty parameter α is tuned through cross-validation approach. This parameter is selected from {0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5}. The optimal values of α for the Salinas, Pavia University, and Indian Pines datasets are 0.5, 1.1, and 0.7, respectively.

For comparison, a wide range of approaches to classify the HSI, such as deep neural networks, SVM, and sparse and low-rank representations, are considered. Therefore, the accuracy of the proposed SDLRRDL approach is evaluated against SVM [3], JRSRC [14], LRR [16], LGIDL [21], SLAP [22], SSMM [12], SSTN [10], and GSC-ViT [11] methods. In [3], hyperspectral data is classified using the SVM method. In [16], the classification of extracted features using low-rank representation is presented. In [14], JRSRC is proposed to classify extracted features using spatial information and a Sparse model. This method employs a sparse representation of the neighboring pixels of the test pixel using a linear combination of a number of training samples. In [21], an LGIDL model is presented that extracts image features using spatial information. SLAP [22] is a low rank model combined with a regularized Laplacian that is applied to each super pixel separately and extracts features with high discrimination using the created affinity graph. The SSTN method [10] is based on deep neural networks that use attention, association, and structured factorized framework modules to extract spectral-spatial features and classify the HSI. GSC-ViT [11] is another deep learning network-based method that uses a light vision transformer network to classify the HSI. This method uses a group-wise separated convolution and a group-wise separated multi-head associate module to reduce the number of convolutional kernel parameters, extract spectral-spatial information, and classify the HSI. In [12], the SSMM method employs the morphological analysis of HSI, state-space model, and multi-head self-attention module for the HSI classification task.

### A. Hyperspectral Image Dataset

This section introduces the three well-known HSI datasets, Indian Pines, Pavia University, and Salinas, to examine the accuracy of different methods. The Indian Pines HSI dataset was imaged by the AVIRIS sensor on June 12, 1992, over an agricultural-forest region in northeastern Indian Pines. This dataset contains 220 bands that cover the wavelengths in 0.4 μm-2.5 μm, of which twenty bands are eliminated as noisy bands. The Indian Pines dataset contains 145 × 145-dimensional images, where pixels are labeled with sixteen classes. The spatial resolution of Indian Pines is 20 meters per pixel. Moreover, the Pavia University dataset was captured by the ROISIS sensor over the Pavia campus in northern Italy in 2002 and consists of 340 × 610 pixels. The spatial resolution of the Indian Pines dataset is 1.3 meters per pixel. The 103 bands of the Pavia University HSI dataset have wavelengths within the interval of 0.43μm - 0.83μm. Twelve bands are eliminated in this dataset as noisy bands, and 103 bands remain. This image has nine classes. The third dataset is Salinas, which has 224 bands, where 16 noise bands are removed. This image has a 217 x 512-dimensional image. The spatial resolution of the Salinas dataset is 3.7 meters per pixel and contains 16 classes [37].

### B. Performance of the Proposed Method

This subsection evaluates the classification accuracy between the proposed SDLRRDL method and state-of-the-art approaches across three benchmark datasets: Indian Pines (Table II), Pavia University (Table II), and Salinas (Table IV). These methods are evaluated based on three metrics: average accuracy (AA), overall accuracy (OA), and κ coefficient. The classification accuracy of various comparison approaches for the India dataset is presented in Table II. Ten percent of the labeled samples of the Indian Pines HSI dataset are randomly chosen for training, and the remaining are for testing. The presented results show the average of 10 runs. Classification methods based on spectral information have achieved lower accuracy than methods that exploit the spectral-spatial signature of HSI. The classification accuracy results of Table II show that approaches such as LRR and SVM have poor performance in the classification task. The LRR method does not consider the local information. Therefore, the extracted features using this method are not rich enough; hence, it does not achieve proper accuracy. The other methods obtain better accuracy by exploiting the spectral-spatial signature of data. These results indicate that the accuracy of the SDLRRDL method is significantly different from that of other methods. The OA of the proposed SDLRRDL method is at least 2.13 percent higher than the best method and 27.79 percent higher than the LRR method. The proposed method is the most accurate approach in most classes of the Indian Pines dataset. Using

spectral-spatial information and considering the global and local structure within data, and providing appropriate constraints on the data structure and learning a dictionary through a set of training data, the proposed method extracts rich features, which causes high accuracy classification.

Most hyperspectral image classification methods suffer from several common limitations: (1) insufficient training samples, (2) noisy spectral signatures, (3) intra-class spectral variability caused by varying physical properties across different regions, and (4) Imbalanced distribution of class samples. Therefore, methods with strong preprocessing and robust modeling are of great interest. The outlier and noisy data cause high-rank data. Methods based on low-rank representation are robust against noisy and outlier data because by eliminate the higher ranks. Using a combination of spectral and spatial information improves the performance of the classifier and reduces the sensitivity to noise, limited training samples, and the spectral variety of samples in a class. The proposed method reduces these limitations by exploiting spatial and spectral information of samples and using a sparse and low-rank representation.

For instance, in the Indian Pines dataset, SDLRRDL achieves high accuracy in Classes 7 and 9, which have very few training samples. Figure 1 shows the classification map for various state-of-the-art approaches. Based on Figure 1, the methods that use only spectral information often misclassify many pixels in homogeneous areas, and their classification maps are noisy. Therefore, the classification map of the methods that use spectral-spatial information is smoother in areas of the same class. Also, the neighboring pixels are classified more accurately. The classification map of the proposed method in Figure 1 illustrates that with an appropriate spectral-spatial method and the use of data structural information in classes with few training samples, pixels are classified with high accuracy, and areas with the same classes are smoothly labeled. In addition, the boundary pixels between classes are classified with higher contrast than other methods. Figure 2 illustrates OA of various classification methods for different training sample numbers on the Indian Pines dataset. Based on Figure 2, SDLRRDL has high accuracy for a small number of training samples.

TABLE II Classification accuracy (%) of different classifier for Indian Pines dataset

| class | SVM | JRSRC | LRR | LGIDL | SLAP | SSMM | GSC-ViT | SSTN | Proposed method |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 87.80 | 58.54 | 19.15 | 63.41 | 91.22 | 65.79 | 74.63 | 100 | **100** |
| 2 | 78.29 | 92.68 | 63.04 | 96.26 | 94.30 | 91.33 | 93.16 | 95.31 | **97.35** |
| 3 | 64.66 | 95.18 | 56.36 | 91.97 | 95.17 | 92.17 | 98.25 | 94.93 | **98.47** |
| 4 | 77.46 | 92.96 | 21.13 | 87.32 | 95.65 | 77.89 | 94.23 | 95.32 | **98.25** |
| 5 | 91.72 | 88.51 | 75.17 | 90.80 | 95.10 | 91.45 | 94.83 | 90.32 | **99.5** |
| 6 | 97.41 | 87.21 | 86.15 | 99.09 | 97.91 | 98.46 | 99.04 | 98.88 | **99.67** |
| 7 | 64.02 | 72.00 | 47.97 | 84.01 | 97.57 | 77.32 | 89.6 | 86.54 | **100** |
| 8 | 98.14 | 99.07 | 79.53 | 96.98 | 99.32 | 97.12 | 100 | 99.00 | 99.54 |
| 9 | 33.33 | 33.33 | 11.11 | 38.89 | 82.07 | 75.05 | 53.33 | 99.91 | 98.34 |
| 10 | 70.15 | 84.91 | 72.11 | 90.17 | 92.79 | 87.27 | 93.06 | 95.27 | **95.30** |
| 11 | 83.52 | 97.56 | 83.88 | 95.74 | 96.91 | 92.87 | 97.86 | 96.63 | **98.69** |
| 12 | 66.88 | 82.02 | 37.45 | 90.26 | 94.33 | 86.71 | 93.67 | 96.98 | **97.65** |
| 13 | 95.65 | 88.04 | 80.43 | 94.02 | 98.63 | 98.78 | 99.46 | 97.50 | **100** |
| 14 | 94.82 | 95.69 | 90.86 | 99.21 | 98.60 | 97.24 | 98.32 | 98.61 | **99.31** |
| 15 | 59.37 | 95.10 | 24.50 | 94.24 | 97.49 | 88.39 | 89.48 | 96.14 | **98.57** |
| 16 | 94.05 | 80.95 | 17.86 | 89.29 | 95.20 | 89.18 | 95.23 | 93.60 | **98.67** |
| OA | 81.67 | 92.36 | 70.47 | 94.52 | 96.13 | 92.14 | 96.03 | 95.52 | **98.26** |
| AA | 78.60 | 83.99 | 54.19 | 87.60 | 95.14 | 87.94 | 91.51 | 95.93 | **98.70** |
| $\kappa$ | 79.02 | 91.24 | 65.45 | 93.74 | 95.68 | 91.04 | 95.48 | 94.91 | **98.02** |

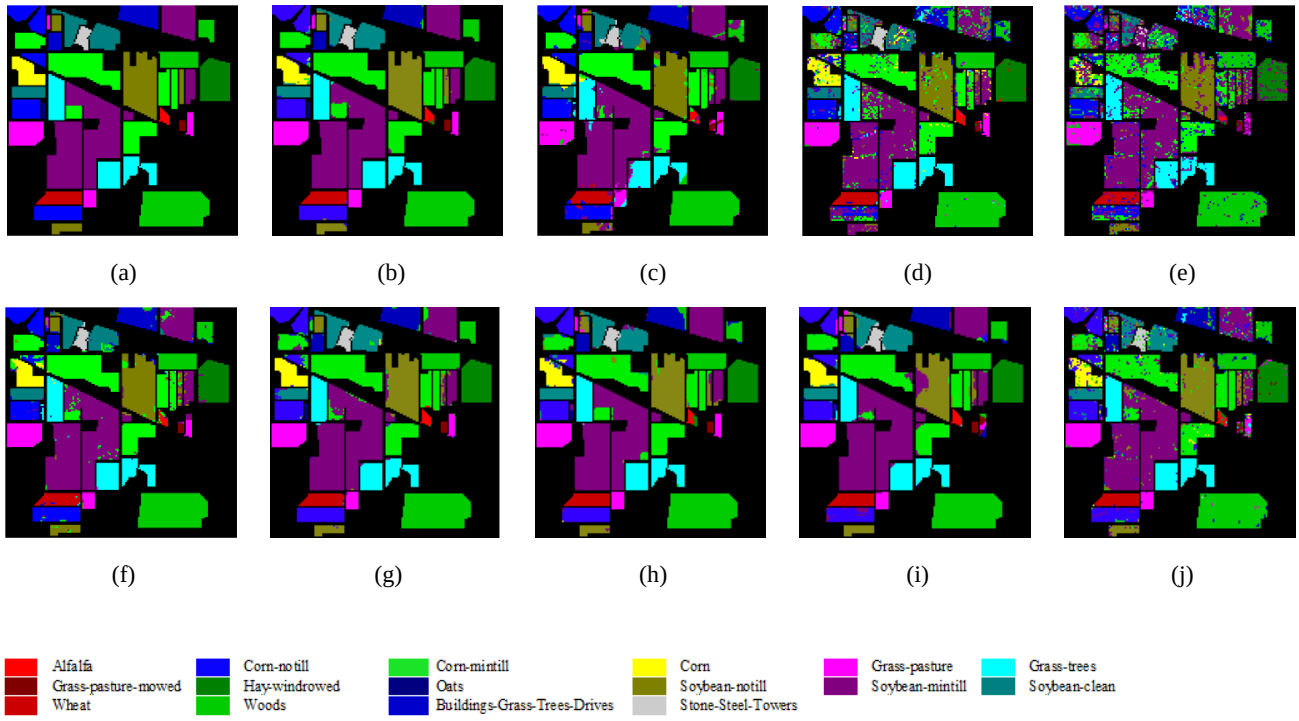


Fig. 1. Classification map of different methods for the Indian Pines dataset. (a)ground truth, (b)The proposed method(SDLRRDL), (c) JRSRC, (d) SVM, (e) LRR, (f) LGIDL, (g) SLAP, (h) GSV, (i) SSTN, (j) SSMM.

TABLE III Classification accuracy (%) of different classifier for Pavia University dataset

| class | SVM | JRSRC | LRR | LGIDL | SLAP | SSMM | GSC-ViT | SSTN | Proposed method |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 94.11 | 95.62 | 88.46 | 96.81 | 99.45 | 98.57 | 99.23 | 98.40 | **99.31** |
| 2 | 96.94 | 99.26 | 97.06 | 99.79 | 99.80 | 98.36 | 99.77 | 98.01 | **99.96** |
| 3 | 81.44 | 88.82 | 72.67 | 89.22 | 85.26 | 89.95 | 95.88 | 91.32 | 95.43 |
| 4 | 94.37 | 91.69 | 74.41 | 98.18 | 98.80 | 98.89 | 96.52 | 98.43 | 98.02 |
| 5 | 99.30 | 99.84 | 68.47 | 100 | 98.87 | 99.88 | 99.74 | 99.71 | 99.88 |
| 6 | 86.73 | 94.91 | 67.54 | 99.35 | 99.10 | 98.84 | 99.55 | 99.90 | 99.81 |
| 7 | 86.30 | 87.89 | 80.36 | 94.70 | 98.30 | 97.49 | 99.38 | 99.49 | 98.93 |
| 8 | 84.02 | 92.68 | 82.68 | 92.17 | 91.91 | 93.82 | 98.17 | 97.11 | 98.13 |
| 9 | 99.89 | 99.59 | 94.56 | 98.89 | 98.92 | 99.21 | 99.31 | 97.59 | 99.28 |
| OA | 93.05 | 96.24 | 86.73 | 97.81 | 98.13 | 97.72 | 99.07 | 97.86 | **99.27** |
| AA | 91.46 | 94.47 | 80.69 | 96.57 | 96.71 | 97.22 | 98.61 | 97.77 | **98.75** |
| $\kappa$ | 90.78 | 94.99 | 81.86 | 97.10 | 97.42 | 96.99 | 98.77 | 97.15 | **99.03** |

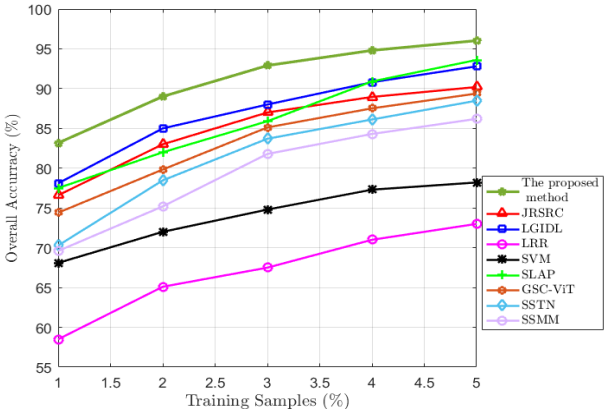


Fig. 2. Overall Accuracy of Various Methods versus Different Numbers of Training Samples for Indian Pines Dataset.

Table III shows the classification performance of various comparison approaches on the Pavia University dataset. Five percent of the HSI-labeled samples are considered for training, and the remaining are for testing. The results are averaged out over ten trials. The SDLRRDL approach outperforms the OA, AA, and κ coefficient of the state-of-the-art approaches. The OA of the proposed SDLRRDL method is at least 0.2 percent higher than that of others. A high improvement in accuracy can be seen compared to methods that only used spectral information. The accuracy of the proposed method is 12.54 percent higher than that of the LRR model. Our method is the most accurate method in classes 1 and 2. Also, it has a negligible accuracy difference with the best accuracy among other methods in the remaining HSI classes. Figure 3 illustrates the OA of various classification approaches versus different numbers of training samples on Pavia University. The proposed SDLRRDL HSI classification approach is accurate in a small number of training samples.

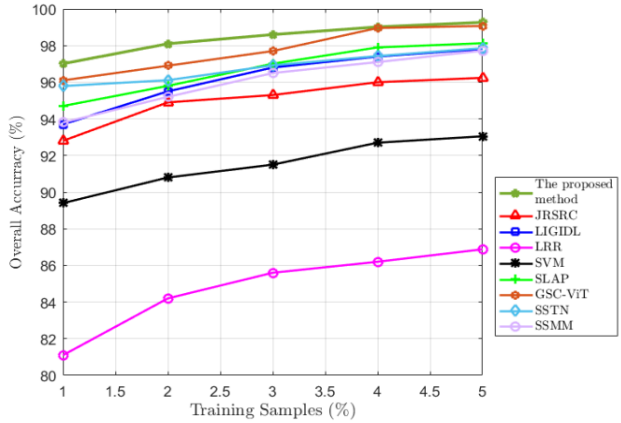


Fig. 3. Overall Accuracy of Various Methods versus Different Numbers of Training Samples for Pavia University Dataset.

For 1 percent of training samples, it has achieved an overall accuracy of 97.01 percent.

The classification performance of various methods on the Salinas dataset is presented in Table IV. For the Salinas dataset, we randomly select 5% of the labeled samples for training purposes, utilizing the remaining 95% for testing. The results are presented based on the average of 10 runs. Based on Table IV, the proposed method improves the OA, AA, and $\kappa$ coefficient of other methods. The OA of our method is at least 0.6 percent higher than that of the other methods. Its accuracy significantly improves the accuracy of methods that rely solely on spectral information. Also, the proposed method is 12.66 percent more accurate than the LRR model. The proposed method has achieved the best accuracy compared to other methods in the first, eighth, 11th, 12th, 13th, 14th, 15th, and 16th classes. Moreover, in other classes, its accuracy has a negligible difference from the best state-of-the-art method.

TABLE IV Classification accuracy (%) of different classifier for Salinas dataset

| class | SVM | JRSRC | LRR | LGIDL | SLAP | SSMM | GSC-ViT | SSTN | Proposed method |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 99.32 | 98.48 | 96.12 | 99.53 | 99.23 | 99.63 | 99.97 | 94.67 | **99.82** |
| 2 | 99.87 | 98.81 | 96.07 | 99.46 | 99.87 | 100 | 99,32 | 99.62 | 99.93 |
| 3 | 99.52 | 99.25 | 94.25 | 99.73 | 99.45 | 99.50 | 99.64 | 99.52 | 99.56 |
| 4 | 98.79 | 98.04 | 92.37 | 99.09 | 98.09 | 98.66 | 99.09 | 98.97 | 98.69 |
| 5 | 97.76 | 97.44 | 93.87 | 98.86 | 99.25 | 99.26 | 99.77 | 97.83 | 99.42 |
| 6 | 99.67 | 98.75 | 94.42 | 99.60 | 99.73 | 99.94 | 99.99 | 99.98 | 99.98 |
| 7 | 99.57 | 99.26 | 96.62 | 99.50 | 99.79 | 100 | 99.97 | 99.96 | 99.76 |
| 8 | 88.35 | 93.59 | 82.73 | 93.99 | 96.61 | 95.71 | 97.66 | 95.77 | **98.88** |
| 9 | 99.86 | 99.44 | 97.88 | 99.41 | 99.99 | 99.80 | 99.99 | 99.95 | 99.97 |
| 10 | 95.41 | 94.84 | 87.32 | 97.85 | 99.02 | 97.11 | 98.74 | 99.68 | 99.46 |
| 11 | 96.45 | 95.86 | 79.70 | 96.75 | 99.92 | 98.71 | 99.50 | 99.25 | **100** |
| 12 | 99.67 | 100 | 94.81 | 99.95 | 99.97 | 99.81 | 99.62 | 99.67 | **100** |
| 13 | 97.74 | 94.48 | 92.64 | 97.82 | 99.81 | 98.91 | 99.26 | 98.67 | **99.96** |
| 14 | 96.95 | 94.39 | 72.44 | 92.62 | 99,43 | 98.48 | 99.58 | 99.62 | **99.97** |
| 15 | 68.79 | 88.20 | 61.87 | 89.04 | 95.08 | 95.96 | 96.89 | 82.20 | **98.89** |
| 16 | 99.30 | 96.97 | 86.49 | 98.66 | 99.58 | 99.66 | 98.74 | 99.77 | **99.88** |
| OA | 92.64 | 95.84 | 86.81 | 96.58 | 98.38 | 98.17 | 98.87 | 94.82 | **99.47** |
| AA | 96.06 | 96.74 | 88.72 | 97.62 | 99.05 | 98.82 | 99.27 | 97.82 | **99.64** |
| $\kappa$ | 91.79 | 95.36 | 85.20 | 0.9619 | 98.20 | 97.96 | 98.75 | 94.25 | **99.41** |

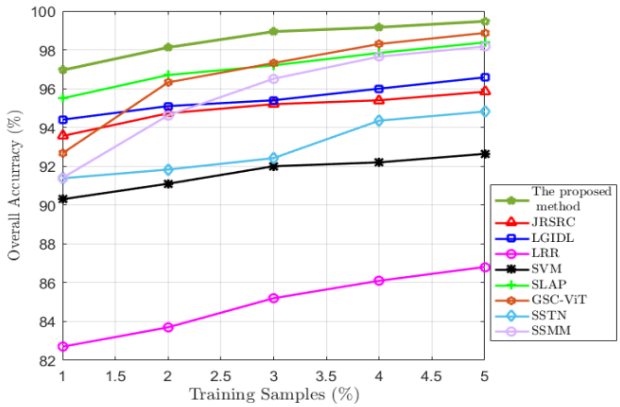


Fig. 4. Overall Accuracy of Various Methods versus Different Numbers of Training Samples for Salinas Dataset.

Figure 4 illustrates the OA of various classification methods for different numbers of training samples on Salinas. As shown in Figure 4, the proposed method achieves high classification accuracy using only a small training set. For only 2 percent of training samples, the proposed method achieves an overall accuracy of 98.02 percent, surpassing the accuracy of most state-of-the-art methods, even when they use 5% of the training data.

TABLE V OA accuracy and computing time (s) for three dataset

| Method | Indian Pines | | Pavia University | | Salinas | |
|---|---|---|---|---|---|---|
| | OA | Time | OA | Time | OA | Time |
| SVM | 81.67 | 5.12 | 93.05 | 6.41 | 92.64 | 8.21 |
| JRSRC | 92.36 | 332.21 | 96.24 | 558.11 | 95.84 | 786.2 |
| LRR | 70.47 | 268.22 | 86.73 | 383.37 | 86.81 | 433.21 |
| LGIDL | 94.52 | 396.13 | 97.81 | 563.21 | 96.58 | 723.12 |
| SLAP | 96.13 | 374.23 | 98.13 | 616.31 | 98.38 | 801.32 |
| SSMM | 92.14 | 148.14 | 97.72 | 283.67 | 98.17 | 401.78 |
| GSC | 96.03 | 206 | 99.07 | 316.25 | 98.87 | 575.23 |
| SSTN | 95.52 | 180 | 97.86 | 365.75 | 94.82 | 475.19 |
| Proposed method | 98.26 | 306.1 | 99.27 | 511.13 | 99.47 | 654.3 |

## C. Runtime Comparison

In this subsection, we evaluate the complexity of the proposed method. The computational complexity of methods can be evaluated using the running time. In this regard, we compare the running times of the proposed method with the state-of-the-art methods. Experiments are executed on a

single Intel Core i7 CPU in a MATLAB 2024 b environment. Table V presents the overall accuracy (OA) and running time of different methods. In the training phase, 10%, 5%, and 5% of the labeled samples are used for the Indian Pines, Pavia, and Salinas datasets, respectively.

Table V indicates that while the proposed method is not a fastest method among the state-of-the-art method, it achieves significant classification accuracy compared to faster methods. However, the running time of the proposed method is not significantly difference with the comparing methods. Based on this table, although the SVM is the fastest method, it has poor classification

.

## V. Conclusions

The current study presents a machine learning method for the HSI classification. This approach proposes the Structure of the Data Regularized LLR with Dictionary Learning (SDLRRDL) model for feature extraction and HSI classification. To enhance low-rank representation, the SDLRRDL method takes the local structure and correlation of data into account by adding sparse regularization to the model and combining spatial and spectral information. Furthermore, the method utilizes a combination of image class information to create similar features for data of the same class. The proposed method represents HSI samples through a linear combination of dictionary atoms. The use of a fixed dictionary and employing all training samples as a dictionary leads to atoms with redundant information, increases computational complexity, and reduces feature Discrimination. SDLRRDL trains the dictionary and utilizes a set of training data to present a dictionary that better matches the training content. The extracted features exhibit high distinctiveness, such that a simple SVM classifier achieves high accuracy. Employing the LADMAP algorithm to solve the objective function has led to faster problem convergence. Experimental results indicate that the SDLRRDL method outperforms state-of-the-art approaches across all three datasets. The proposed method improves the accuracy of state-of-the-art methods by more than 2.13, 0.2, and 0.6 percent on the Indian Pines, Pavia University, and Salinas datasets, respectively.

## References

[1] M. Hamed and F. Hajiani, "A method for segmenting remote sensing images using the Watershed algorithm and Fuzzy C-Means clustering," *J. Commun. Eng.*, vol. 10, no. 37, pp. 65–72, Nov. 2020, doi: 20.1001.1.29809231.1399.10.37.6.0.

[2] D. Ma, S. Xu, Z. Jiang, and Y. Yuan, "Central pixel-based dual-branch network for hyperspectral image classification," *Remote Sens.*, vol. 17, no. 7, p. 1255, Mar. 2025, doi: 10.3390/rs17071255.

[3] Y. Xiao, H. Wang, and W. Xu, "Parameter selection of Gaussian kernel for one-class SVM," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 941–953, May. 2015, doi: 10.1109/TCYB.2014.2340433.

[4] A. Farhadi, M. Mirzarezaee, A. Sharifi, and M. Teshnehlab, "Unsupervised domain adaptation for image classification based on deep neural networks," *Intelligent Multimedia Processing and Communication Systems (IMPCS)*, vol. 4, no. 1, pp. 27–37, Mar. 2023, doi: 10.71856/impcs.2023.903575

[5] Z. Chen, H. Yang, Q. Liu, Y. Liu, M. Zhu, and X. Liang, "Deep learning for hyperspectral image classification: A critical evaluation via mutation testing," *Remote Sens.*, vol. 16, no. 24, pp. 4695–4720, Dec. 2024, doi: 10.3390/rs16244695.

[6] M. Hariri and N. Hydarzadeh, "Face recognition in images using Viola-Jones method and image texture analysis," *Intelligent Multimedia Processing and Communication Systems (IMPCS)*, vol. 4, no. 2, pp. 1–10, Jun. 2023, doi: 10.71856/impcs.2023.903604.

[7] S. G. Azar, S. Meshgini, T. Y. Rezaii, and S. Beheshti, "Hyperspectral image classification based on sparse modeling of spectral blocks," *Neurocomputing*, vol. 407, pp. 12–23, Sep. 2020, doi: 10.1016/j.neucom.2020.04.138.

[8] F. Hajiani and A. Mahmoodzadeh, "Hyperspectral image segmentation using homogeneous area limiting and shortest path algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, May. 2017, doi: 10.14569/IJACSA.2017.080929.

[9] J. Kang, Y. Zhang, X. Liu, and Z. Cheng, "Hyperspectral image classification using spectral–spatial double-branch attention mechanism," *Remote Sens.*, vol. 16, no. 1, p. 193–213, Jan. 2024, doi: 10.3390/rs16010193.

[10] Z. Zhong, Y. Li, L. Ma, J. Li, and W.-S. Zheng, "Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Oct. 2021, doi: 10.1109/TGRS.2021.3115699.

[11] Z. Zhao, X. Xu, S. Li, and A. Plaza, "Hyperspectral image classification using groupwise separable convolutional vision transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1-17, Mar. 2024, doi: 10.1109/TGRS.2024.3377610.

[12] M. Ahmad et al., "Spatial–spectral morphological mamba for hyperspectral image classification," *Neurocomputing*, vol. 636, Jul. 2025, doi: 10.1016/j.neucom.2025.129995.

[13] C. Li, D. Zhu, C. Wu, B. Du, and L. Zhang, "Global overcomplete dictionary-based sparse and nonnegative collaborative representation for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, Mar. 2024, doi: 10.1109/TGRS.2024.3381719.

[14] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 641–645, May. 2016, doi: 10.1109/LGRS.2016.2532380.

[15] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, Nov. 2015, doi: 10.1109/TGRS.2014.2363582.

[16] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, pp. 1615–1622, Jan. 2011, doi: 10.1109/ICCV.2011.6126422.

[17] X. Cheng, Y. Zhang, Y. Yang, L. Jiao, and L. Wang, "A novel low-rank hypergraph feature selection for multi-view classification," *Neurocomputing*, vol. 253, pp. 115–121, Aug. 2017, doi: 10.1016/j.neucom.2016.10.089.

[18] Y. Wang et al., "Self-supervised low-rank representation (SSLRR) for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5658–5672, May. 2018, doi: 10.1109/TGRS.2018.2823750.

[19] F. Hajiani, P. Naser, and A. Keshavarz, "Hyperspectral image classification using low-rank representation and spectral-

spatial information," *J. South. Commun. Eng.*, vol. 11, no. 43, pp. 27–38, Mar. 2022, doi: 10.30495/jce.2022.689206.
[20] F. Hajiani, N. Parhizgar, and A. Keshavarz, "Hyperspectral image classification using cluster based graph regularized low rank representation and dictionary learning," *Neurocomputing*, vol. 462, pp. 208–220, Oct. 2021, doi: 10.1016/j.neucom.2021.07.075.
[21] Z. He, L. Liu, R. Deng, and Y. Shen, "Low-rank group inspired dictionary learning for hyperspectral image classification," *Signal Process.*, vol. 120, pp. 209–221, Mar. 2016, doi: 10.1016/j.sigpro.2015.09.004.
[22] S. Yang, Y. Zhang, Y. Ding, and D. Hong, "Superpixelwise low-rank approximation-based partial label learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, May. 2023, doi: 10.1109/LGRS.2023.3279985.
[23] A. Goel and A. Majumdar, "Semi-supervised graphical deep dictionary learning for hyperspectral image classification from limited samples," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 2108–2114, Sept. 2024, doi: 10.1109/ICIP51287.2024.10647482.
**[24]** R. Gribonval, "Fast matching pursuit with a multiscale dictionary of Gaussian chirps," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 994–1001, May. 2001, doi: 10.1109/78.917803.
[25] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec.1993, doi: 10.1109/78.258082.
[26] S. Fischer, G. Cristóbal, and R. Redondo, "Sparse overcomplete Gabor wavelet representation based on local competitions," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 265–272, Feb. 2006, doi: 10.1109/TIP.2005.860614.
[27] W. Fu, S. Li, L. Fang, and J. A. Benediktsson, "Contextual online dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1336–1347, Oct. 2018, doi: 10.1109/TGRS.2017.2761893.
[28] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015, doi: 10.1109/TGRS.2014.2325067.
[29] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006, doi: 10.1109/TSP.2006.881199.
[30] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May. 2007, doi: 10.1109/TIP.2007.891788.
[31] Z. Lin et al., "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," Coordinated Sci. Lab., Univ. Illinois, Urbana, IL, USA, Tech. Rep. UILU-ENG-09-2214, DC-246, 2009.
[32] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, May. 2010, doi: 10.1137/080738970.
[33] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
[34] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *Adv. Neural Inf. Process. Syst.*, vol. 24, pp. 612–620, Dec. 2011.
[35] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented Lagrangian methods for semidefinite programming," *Math. Program. Comput.*, vol. 2, no. 3, pp. 203–230, Sep. 2010, doi: 10.1007/s12532-010-0017-1.
[36] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 19–60, Mar. 2010. doi: 10.1145/1756006.1756008.
[37] M. Graña, M. A. Veganzons, and B. Ayerdi, "Hyperspectral remote sensing scenes," *Univ. Basque Country*, Accessed: Jun 2018. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.