

Article

A Novel Technique for Semantic Segmentation of Hyperspectral Images Using Multi-View Features

Reaya Grewal ¹, Geeta Kasana ^{1,*} and Singara Singh Kasana ²

¹ Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala 147004, India; rgrewal_phd19@thapar.edu

² Computer Science and Information Technology, Central University of Haryana, Mahendergarh 123031, India; singarasingh@cuh.ac.in

* Correspondence: gkasana@thapar.edu

Abstract: This research presents an innovative technique for semantic segmentation of Hyperspectral Image (HSI) while focusing on its dimensionality reduction. A unique technique is applied to three distinct HSI landcover datasets, Indian Pines, Pavia University, and Salinas Valley, acquired from diverse sensors. HSIs are inherently multi-view structures, causing redundancy and computation overload due to their high dimensionality. The technique utilizes Canonical Correlation Analysis (CCA) variants, Pairwise CCA (PCCA) and Multiple Set CCA (MCCA), to extract features from multiple views of the input image simultaneously. The performance of PCCA and MCCA is compared with the traditional Principal Component Analysis (PCA) on all datasets. The superior performance of CCA variants, particularly MCCA, is demonstrated in achieving higher Overall Accuracy (OA) for semantic segmentation compared to PCA. The research extends the analysis by integrating machine learning classifiers for per pixel prediction, demonstrating the effectiveness of the proposed techniques i.e., PCCA-SVM and MCCA-SVM.

Keywords: hyperspectral image; multi-view learning; semantic segmentation; canonical correlation analysis; principal component analysis; curse of dimensionality



Citation: Grewal, R.; Kasana, G.; Kasana, S.S. A Novel Technique for Semantic Segmentation of Hyperspectral Images Using Multi-View Features. *Appl. Sci.* **2024**, *14*, 4909. <https://doi.org/10.3390/app14114909>

Academic Editor: Douglas O'Shaughnessy

Received: 15 April 2024

Revised: 26 May 2024

Accepted: 29 May 2024

Published: 5 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hyperspectral Imaging (HSI) has emerged as a pivotal technology in remote sensing, revolutionizing our ability to observe and comprehend the world from above. Unlike traditional RGB images, which capture just three bands of light, hyperspectral sensors are capable of acquiring data across hundreds of contiguous narrow spectral bands across the electromagnetic spectrum. This unique capability enables hyperspectral images to provide a profound level of spectral detail, which is instrumental in a number of applications ranging from vegetation, mineral and forest mapping [1–3], chemical imaging [4], landcover classification [5], and anomaly detection [6] among others.

To comprehend HSIs information and do extended analysis, Semantic segmentation is an advanced computer vision task that involves classifying each pixel in an image into specific semantic categories, thereby assigning a label to every region of interest. It offers a finer level of information by supplying semantic labels to individual pixels in the image. This degree of detail improves the machine's ability to learn and helps it comprehend more intricate concepts, such the boundaries and spatial relationships between several objects in an image. Additionally, manual labeling of images which may be expensive and time-consuming is eliminated by semantic segmentation. It can also be used to identify several things in a picture and is not restricted to a single sort of object.

Due to advancements and continuous research based on Semantic segmentation techniques, they are being used for a variety of data-rich and diversified remote sensing problems [7]. It has played an immense role in various applications such as lane analysis

for autonomous vehicles [8] and geolocalization for Unmanned Aerial Vehicles [9]. Environmental monitoring [10], crop cover and type analysis [11], and land usage analysis in urban areas [12] are a few instances of semantic segmentation of remote sensing imagery. However, semantic segmentation becomes particularly challenging due to the high-dimensional nature of the data, where each pixel is represented by a spectrum across numerous spectral bands [13]. The spectral bands are hundreds in number, redundant and cause Hughes' curse of dimensionality [14] which necessitates efficient techniques that can capture the relevant information while handling the risk of overfitting and computational inefficiency.

To deal with the Curse of Dimensionality, a bunch of Feature Selection (FS) and Feature Extraction (FE) techniques exist. While FS picks from raw features it also loses some of the original features, whereas FE converts the high dimension into meaningful low-dimensional feature space without losing any important feature. FE techniques like Principal Component Analysis (PCA) [15], Independent Component Analysis (ICA) [16], Linear Discriminant Analysis (LDA) [17], Wavelet Transform [18] and more have been implemented on HSI. While some adopted Bilateral filter based feature extraction on Superpixels of HSI [19], others utilised adaptive variation filtering over the principal components acquired through PCA [19]. Minimum Noise Fraction has also played a vital role to reduce dimensions on superpixels of HSI to extract reduced spectral and spatial features [20]. The most important and relevant spectral bands are also found by utilizing a variety of metaheuristic optimization techniques, including Particle Swarm Optimization (PSO) [21], Genetic Algorithm (GA) [22], Gray Wolf Optimization (GWO) [23], and Whale Optimization Algorithm (WOA) [24]. Ranking-based [25] and clustering-based [26] techniques have been explored as well to reduce the dimensions.

The aforementioned dimension reduction techniques focus on the global features and consider HSI as a single dataset and give a singular reduced spectral dataset. However, a Novel perspective emerges when we consider hyperspectral images as a multi-view dataset. In traditional multi-view learning scenarios, multiple sets of features or views of the same data are available. The term 'views' refers to the different ways in which the data is observed, captured, or represented. These views provide complementary information that can enhance the learning process and lead to more robust models. Recent researches have focused on multi-view feature learning and their fusion along with supremacy of Support Vector Machine (SVM). A variety of features like Local Binary Pattern (LBP), Gabor Filters, Gray Level-Gradient Co-occurrence matrix (GLCM) were fused to provide important features to SVM to detect weed and corn in a field [27]. Kernelized SVM helped to fuse different selected features for multi-class classification with recursive feature elimination [28]. Multi-view features were embedded with SVM to detect clouds in remote sensing images [29]. Multiple feature views extracted through GLCM and LBP fused with SVM also improved classification of mitotic and non-mitotic cells in breast cancer images [30]. Hence, Multi-view feature learning forms a strong research area for application on HSI. HSI by their very nature, can be treated as a multi-view dataset by partitioning them on the basis of each wavelength at which the scene has been captured. Each spectral band provides us a distinct view of the same scene. To extract useful patterns from multiple views and extract features simultaneously, Canonical Correlation Analysis (CCA) has been proposed to be a useful technique.

1.1. Motivation

The major motivation of this paper is to highlight the superiority of CCA in comparison with PCA. As discussed above, many researchers have addressed high dimensionality of HSI and subsequently classified it. To the best of our knowledge, this is the first study to employ CCA while treating HSI as multi-view and obtaining reduced spectral features. The reduced features have been projected to SVM for semantic segmentation (dense per pixel prediction), owing to its aforementioned advantages and good performances.

Pairwise CCA (PCCA) and Multi-Set CCA (MCCA) have been employed to extract features in the multi-view fashion. While both techniques share the common objective of un-

covering correlations across multiple views, they differ significantly in their approach and applicability. PCCA is a variant of CCA that operates by computing pairwise correlations between each pair of views and extracts features between two views at a time.

On the other hand, MCCA is generalized CCA, which extends the concept of CCA to more than two views simultaneously. Unlike PCCA, which focuses on pairwise comparisons, MCCA considers the joint correlation structure across multiple views. It aims to uncover common latent factors that are shared across all views, capturing the global relationships among all datasets simultaneously.

The suitability and effectiveness of the CCA-SVM approach in semantic segmentation of landcover HSI, can further enhance the opportunities among the research community in this domain.

1.2. Contribution

The following are the main contributions of the paper to the field of HSI analysis:

- A novel multi-dimensional perspective and feature extraction approach to HSI analysis is introduced, treating each hyperspectral image as multi-view/multi-modal dataset. Each spectral band of the image provides a distinct “view” of the scene, facilitating innovative techniques for simultaneous feature extraction from multiple views. For this, CCA is applied for the first time due to its ability to capture correlated information across all the views simultaneously.
- The extracted multi-view features using CCA are used to predict per pixel class and performed semantic segmentation using SVM.
- A rigorous experimental investigation is conducted to evaluate the performance metrics associated with semantic segmentation of HSIs using multi-view features obtained through CCA. The approach is meticulously compared with conventional feature extraction technique PCA and other machine learning algorithms for segmentation. This comprehensive analysis not only validates the effectiveness of the Proposed Technique but also provides insights into its superiority over existing techniques.

1.3. Organisation of Paper

This paper is structured as follows: Section 1 introduces the concept of HSI, highlighting one of the major challenges of HSI which is Curse of Dimensionality and how semantic segmentation helps in HSI analysis. Section 2 discusses the existing state of the art techniques for HSI dimension reduction and segmentation. Section 3 gives a deeper and mathematical understanding of the techniques proposed in this research paper. Section 4 outlines our Proposed Technique, elucidating the fusion of CCA and SVM within the context of per-pixel prediction i.e., semantic segmentation. Section 5 presents our experimental setup, including the datasets used and performance metrics employed. The results and discussion are provided as well, offering insights into the effectiveness of our approach. Finally, Section 6 encapsulates our conclusions and outlines avenues for future research.

2. Background

Literature suggests for handling huge spectrum, Principal Component Analysis (PCA) has been a commonly employed technique to reduce the spectral dimension for HSI analysis in various applications. In 2002, Rodarmel and Shan [31] introduced the application of statistical approach of PCA on HSI to reduce the number of spectral bands. Their study on two landcover datasets HYDICE and AVIRIS revealed that only first few Principal Components (PCs) gave about 70% correct classification rate. Their results were benchmarked for other researchers who heavily relied on first few PCs in their experiments.

Even till recently, PCA stands as the most widely adopted pre-processing step for dimensionality reduction, which optimizes variance preservation and retains comprehensive spatial information on a global scale. Castaings et al. [32] employed PCA and a gaussian kernel PCA to get informative bands. The reduced spectrum along with morphological operations, was used to classify hyperspectral images. Che et al. [33] segmented the bruises

on apple, where PCA was primarily used to remove redundant information from HSI cube. The first three PCs were used as input for the segmentation task. Torres et al. [34] reduced the abundant spectral information using a combination of PCA and Analysis of Variance (ANOVA) for separating the green oranges from leaves. Fernandez et al. [35] pointed out the complex matrix computations of eigenvectors and eigenvalues on such high dimensional data that were needed for PCA. So they introduced a reconfigurable hardware version of PCA using field-programmable gate arrays for high-speed computing of HSI. Sun et al. [36] selected informative bands from HSI using fast and robust PCA on Laplacian graph. The original data could be cleanly represented in low-rank approximation due to PCA. After double PCA implementation by Li et al. [37], the third PC of HSI gave the best results for segmenting the decayed portions of apples. Due to global variance in data and the dominance of visible and near-infrared bands, PCA is unable to extract the fundamental structure of HSI, and cumulative-variance accumulation is unable to take advantage of non-linear correlations among the transformed features that PCA-based approaches provide. Hence, Uddin et al. [38] used Mutual Information based Minimum Redundance and Maximum Relevance measure based PCA to capture non linear features of HSI. The existing literature is saturated with a variety of combinations of PCA employed on HSI to reduce dimensions.

CCA and PCA can be considered as sister techniques. However, PCA ignores the loss of information and lack of interpretation of the principal components in terms of original features. A brief abstract overview and difference between PCA and CCA has been depicted in Figure 1. Since it can reduce the labeled instance complexity, CCA, which was first developed to measure the linear correlation between two sets of variables, was formally described as a multi-view dimensionality reduction technique by Foster et al. [39]. An additional attractive feature of CCA is that in the event that noise in one view is uncorrelated to the other, the learnt subspace will not contain the noise in the uncorrelated dimensions [40].

CCA is an effective approach for extracting features from multi-view images by uncovering the relationships between different sets of variables. Recent researches show it improves the robustness of classification techniques by combining the strengths of different viewpoints to create a more comprehensive representation. CCA created a hybrid and improved kernel for HSI classification by capturing similarities between spectral and spatial features and hence combining them for SVM [41]. It has also addressed limited training samples issue by performing cross-domain collaborative learning [42]. It assigns labels to unknown samples based on correlation between target clusters created using random walker algorithm. With the advent of big data, CCA has helped to find patterns and associations between hundreds or even thousands of behavioural, neurological, and genetic phenotypic subject characteristics of brain image dataset in neuroscience [43]. CCA has also played a significant role in medical field by combining with deep learning techniques to segment white blood cells and classify blood samples [44]. CCA extracted various features from overlapping nuclei, compressed input images' dimensions and helped network converge faster. Along with dimension reduction, the powerful predictive power of machine learning algorithms has shown great results for the classification and segmentation of such images [45].

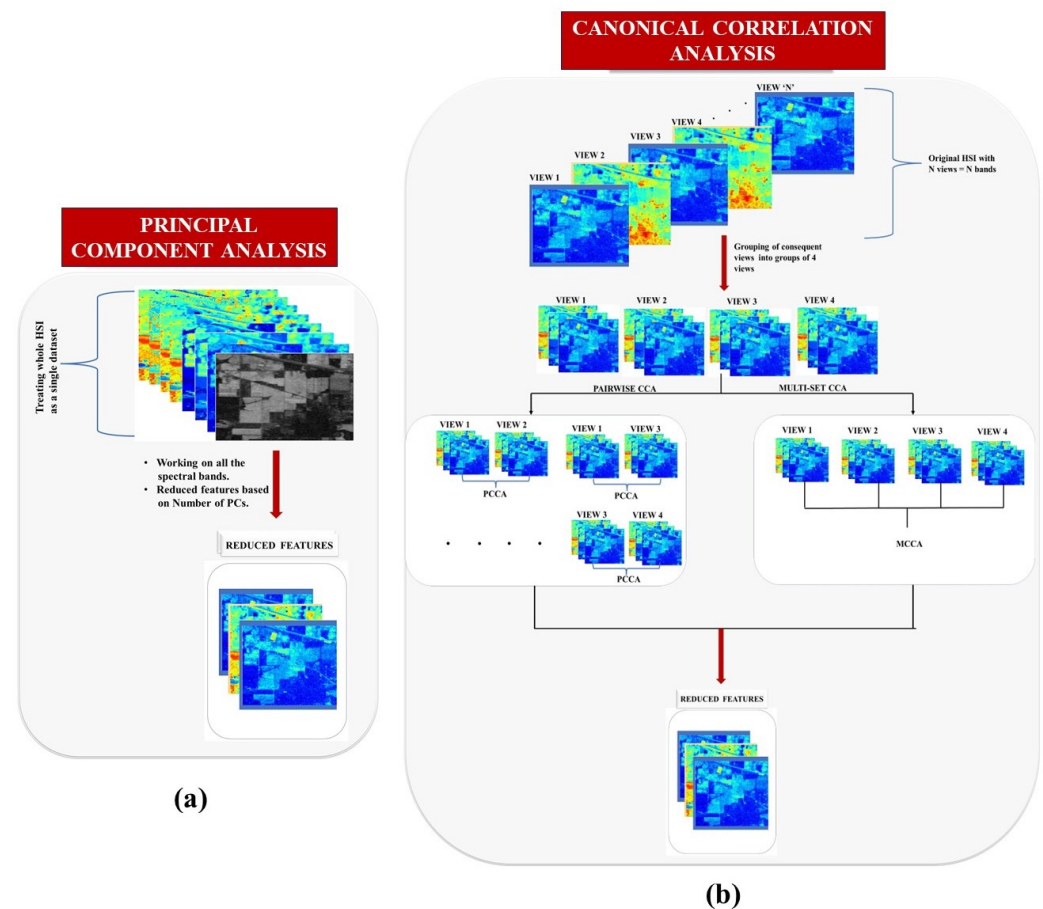


Figure 1. Overview of PCA vs. CCA. (a) A breakdown of major steps in PCA, (b) A breakdown of major steps in CCA.

For the last decade, Deep Learning based techniques have been employed to perform Semantic Segmentation [46]. Some popular Convolutional Neural Network (CNN) architectures have been used on which various researchers have based their semantic segmentation. Their work is inspired from architectures of AlexNet [47], VGGNet [48] Unet [49] and GoogleNet [50]. AlexNet is made up of three fully-connected layers and five convolutional layers. A pooling layer, designed to lower dimensionality, sits between two convolutional layers that are next to each other. Three fully linked layers and several convolutional layers make up VGGNet. VGG-11 and VGG-16 are two examples of a suite of VGGNets that may be made by changing the number of convolutional layers. Three things set GoogLeNet apart from other CNN variations: it has one fully connected layer, has an inception module, and uses auxiliary classifiers during training. The inception model concatenates the filtering results with the max-pooling result after applying filters of three sizes: 1×1 , 3×3 , and 5×5 to the input. The things that set GoogLeNet apart from other CNN variations: it has one fully connected layer and has an inception module which concatenates the filtering results with the max-pooling result after applying filters of three sizes: 1×1 , 3×3 , and 5×5 to the input. Unet has been a frequently applied architecture to semantically segment HSI. It consists of an expanding path (decoder) and a contracting path (encoder) forming a symmetric U-shaped structure. The contracting path gradually reduces spatial dimensions while increasing feature depth, capturing context through the use of convolutional and max-pooling layers. In order to provide accurate localization and comprehensive segmentation outputs, the expanding path upsamples the feature maps using transposed convolutions and concatenates them with equivalent feature maps from the contracting path. Some researchers combined UNet with Transformers to extract spectral spatial features in HSI [51].

Despite their impressive capabilities, deep learning models come with some significant drawbacks. The limited HSI samples are challenging for these models as they are heavily reliant on large volumes of labeled data for effective training, which can be both costly and time-consuming to gather. Another issue is their tendency to overfit, particularly when the HSI dataset is small, resulting in poor generalization to new, unseen data. The training and testing samples tend to overlap which also cause overfitting. Additionally, deep learning models get computationally extensive and maintaining hyperparameters becomes challenging. Machine learning Techniques like SVM [52] become a robust choice to perform semantic segmentation on HSI due to its ability to handle high-dimensional data efficiently by finding optimal hyperplane that maximizes class separation for each pixel.

SVM has been a powerful, supervised learning algorithm for HSI analysis in a variety of applications. It excels in scenarios where the data exhibits complex, nonlinear relationships and is particularly well-suited for hyperspectral images due to their high dimensionality. SVM classification is based on the formation of a decision boundary that owns the largest margin of separation between data samples from several classes. This decision boundary could be linear or non-linear with the help of kernels [53]. The ability of SVM to discern complex, nonlinear relationships and form decision boundaries with the maximum margin of separation between classes makes it invaluable for classification and segmentation tasks. Hu et al. [54] utilized gaussian kernel SVM for spectral spatial classification of Pavia University data along with mathematical morphology. The SVM multi-class results were converted into multiple binary classification results to reduce the noise. Ye et al. [55] applied HSI analysis to detect decay in potatoes where too SVM played a pivotal role. To achieve the best detection accuracy, the parameters of SVM were tuned through the grid search method. In the approach by Ji et al. [56], SVM played a major role in the non-invasive identification of defective potatoes. On being combined with k-means clustering, it gave the desired results. Apart from being in use in agriculture and food quality assessment, SVM has been utilized to identify multi-class landcover scenes as well. Akbari et al. [57] employed SVM on spectral spatial features obtained through Gabor and wavelet filters on Pavia University (PU), obtaining high classification of 96.7% accuracy. SVM has also been applied in multiple kernel learning environment, like Wang et al. [58] utilized spectral, semantic and spatial features infused kernel. This composite kernel SVM was used to classify the landcover Indian Pines and Pavia University dataset. Anand et al. [59] also fed 3D spectral-spatial features of these datasets into SVM for per-pixel prediction. Miclea et al. [18] reduced spectral dimensions through wavelet transform and classified using SVM. As evident from above discussed literature, the ability of SVM to discern complex, nonlinear relationships and form decision boundaries makes it invaluable for classification and segmentation tasks.

3. Materials and Methods

In this section we discuss the datasets and major methods used for multi-view features extraction and performing its semantic segmentation.

3.1. Datasets Description

The following datasets have been used for the implementation of the Proposed Technique.

3.1.1. Indian Pines

Indian Pines is situated in Indiana, USA. This landscape, which includes lush, agricultural land, vegetation, and a natural woodland area, has a total resolution of $145 \times 145 \times 224$. The water's ability to reflect and refract light causes the water's absorption bands to disappear. Bands 104 to 108, 150 to 163, and 200 have been eliminated. As a result, the updated image has 200 bands. As seen in Figure 2, it has 16 distinct classes with a total pixel count ranging from 20 to 2455.

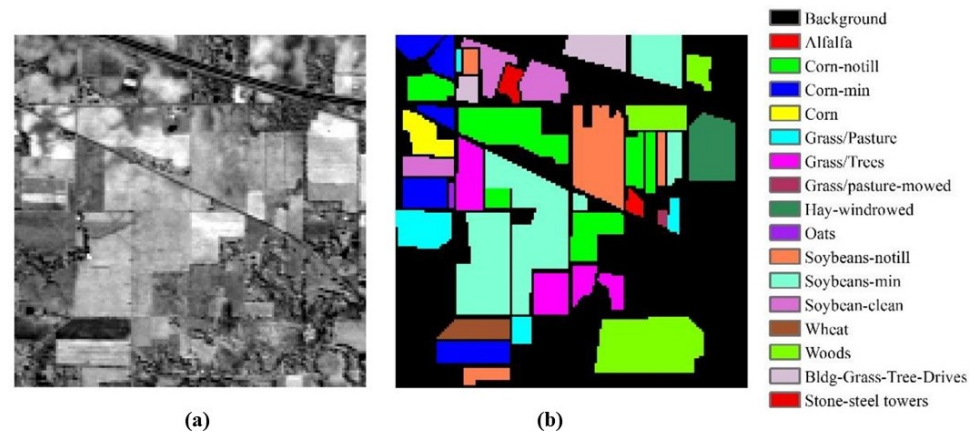


Figure 2. Indian Pines image (a) Sample band image (b) Ground Truth.

3.1.2. Pavia University

The ROSIS sensor uses a flight above Pavia, Italy, to capture the scene at Pavia University. The scene resolution was originally $610 \times 610 \times 115$, however, some areas of the image were removed since they were missing information, leaving $610 \times 340 \times 103$ as the final resolution. As illustrated in Figure 3, it comprises nine distinct classes and has a geometric resolution of 1.3 m. Radiometric adjustments involve the removal of 12 bands.

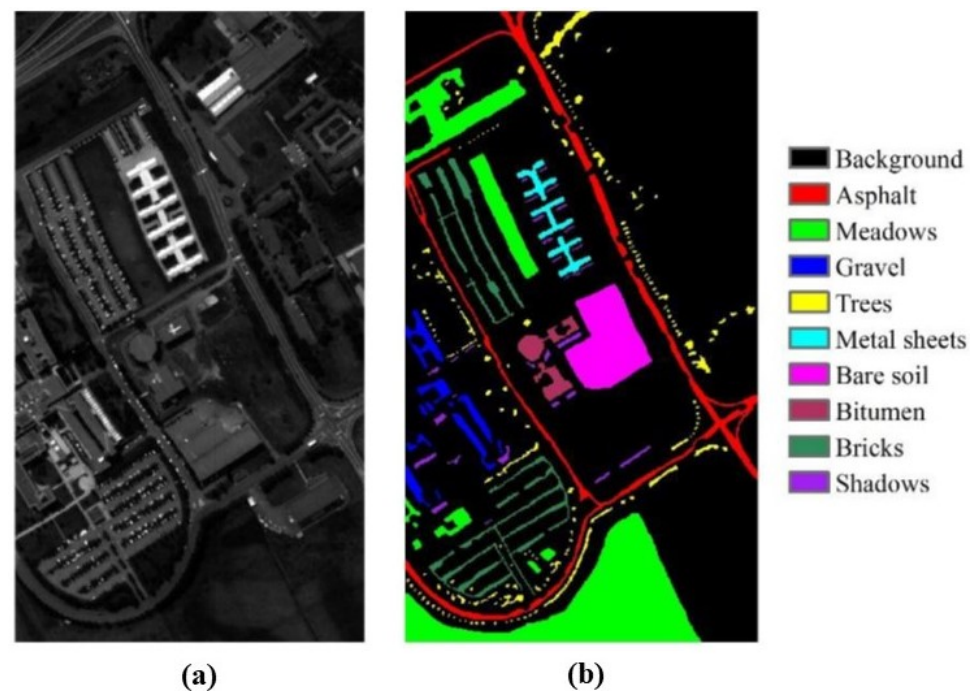


Figure 3. Pavia University image (a) Sample band image (b) Ground Truth.

3.1.3. Salinas Valley

The Salinas scene, with a high spatial resolution of 3.7 m per pixel, is located in California. This scene is recorded by the AVIRIS sensor. It has a resolution of $512 \times 217 \times 224$. Removed are the bands ranging from 108 to 112, 154 to 167, and 224. There are 204 bands on the updated picture. Figure 4 illustrates its 16 classes, which include stubble, Romaine lettuce, green broccoli weeds, and fallow.

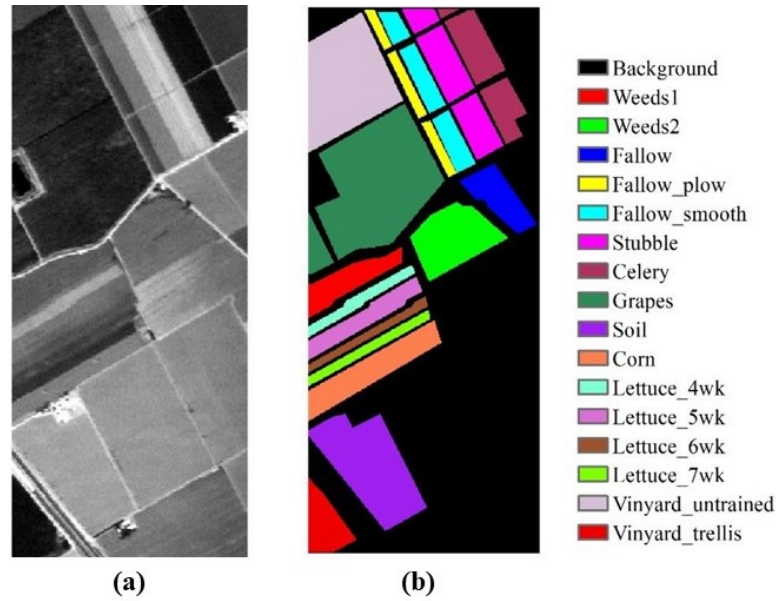


Figure 4. Salinas Valley image (a) Sample band image (b) Ground Truth.

3.2. Canonical Correlation Analysis

CCA is a statistical technique employed to examine the associations between two sets of variables. It seeks linear combinations of variables in each set, referred to as canonical variables, such that the correlation between corresponding canonical variables is maximized. CCA is particularly useful when exploring relationships between multivariate datasets, aiming to uncover latent structures and patterns that exist between them. The primary objective of CCA is :

- **Maximize Correlation:** Find pairs of canonical variables, one from each set, such that the correlation between these pairs is maximized. These canonical variables capture the shared variance between the sets.
- **Sequential Independence:** Subsequent canonical variables are sought in a way that ensures they are uncorrelated with previous canonical variables. This ensures that each new canonical variable captures distinct, independent information.

Let $X = [x_1, x_2, x_3, \dots, x_N] \in R^{d_x \times N}$ and $Y = [y_1, y_2, y_3, \dots, y_N] \in R^{d_y \times N}$ represent two data matrices with number of instances as 'N' and d_x, d_y represent the features of X and Y data matrices, respectively. CCA works towards finding 'K' pairs of linear projections, which are called Canonical Vectors represented as $W_x = [w_{x,1}, w_{x,2}, \dots, w_{x,K}] \in R^{d_x \times K}$ and $W_y = [w_{y,1}, w_{y,2}, \dots, w_{y,K}] \in R^{d_y \times K}$. These vectors are such that the correlation between $W_x^T X$ and $W_y^T Y$ is maximized. If we consider a canonical vector $w_x \in R^{d \times 1}$ for X and a canonical vector $w_y \in R^{d \times 1}$ for Y, the correlation coefficient ρ is maximised by CCA between $X^T w_x$ and $Y^T w_y$ as shown in Equation (1)

$$\rho(X^T w_x, Y^T w_y) = \frac{w_x^T X Y^T w_y}{\sqrt{(w_x^T X X^T w_x)(w_y^T Y Y^T w_y)}} \quad (1)$$

The Equation (1) in the constrained form is represented as:

$$\begin{aligned} & \max_{w_x, w_y} w_x^T X Y^T w_y \\ & s.t. w_x^T X X^T w_x = 1, w_y^T Y Y^T w_y = 1 \end{aligned} \quad (2)$$

The Covariance matrix XX^T or YY^T is singular due to high feature dimensionality which results in undetermined optimization problem. Also, Regularization in CCA helps

to deal with noise in the data. By introducing Equations (3) and (4), regularizations to the covariance matrices can help with the problem.

$$\sum_{xx} = \frac{1}{N} XX^T + r_x I \quad (3)$$

$$\sum_{yy} = \frac{1}{N} YY^T + r_y I \quad (4)$$

where r_x, r_y are regularization coefficients.

It is assumed that the noise vectors are identically distributed, gaussian, and independent in the d_x and d_y columns of X and Y , respectively. As a result, all possible combinations of the covariances of the columns in X and Y will equal zero, except the covariance between a specific column vector and itself. This variance of each column of X and Y is labeled r_x and r_y are called the regularization coefficients. For choosing these parameters, let w_x^i and w_y^i be the weights calculated through CCA when data samples X_i and Y_i are removed, where $i \in \{1, 2, \dots, n\}$. The coefficients r_x and r_y are chosen through Grid Search [60] optimization of the following cost function [61]:

$$\max_{r_x, r_y} [\text{corr}(\{X_i w_x^i\}_{i=1}^n, \{Y_i w_y^i\}_{i=1}^n)] \quad (5)$$

where ‘corr’ refers to the Pearson’s Correlation Coefficient. Whenever a sample i is removed, the above cost function measures changes in w_x^i and w_y^i and seek optimal values of the regularization coefficients where the change is minimized [62].

Eigen Value decomposition (EVD) and Singular Value Decomposition (SVD) are two major approaches for calculating W_x and W_y .

1. Generalised EVD is formulated using:

$$\begin{bmatrix} 0 & \sum_{xy} \\ \sum_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} \sum_{xx} & 0 \\ 0 & \sum_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \quad (6)$$

2. Here, substitutions are made where $\sum$ ’s are Covariance matrices.

$$\begin{aligned} \sum_{xy} &= \frac{1}{N} XY^T, \\ \sum_{yx} &= \frac{1}{N} YX^T. \end{aligned} \quad (7)$$

3. The top ‘ K ’ generalised eigen vectors are: $\{[w_x, k; w_y, k]\}_{k=1}^K$.

4. The K th generalised eigen value is equal to the correlation $\rho(w_{x,k}^T X, w_{y,k}^T Y)$.

The second solution to calculate W_x, W_y involves performing SVD.

1. SVD is formulated on a matrix T .

2. The matrix T is given by the expression, $T = \sum_{xx}^{-1/2} \sum_{xy} \sum_{yy}^{-1/2}$.

3. Let W_x, W_y be the K th leading left and right singular vectors of matrix T .

4. This gives us canonical matrices which will be $\sum_{xx}^{-1/2} W_x, \sum_{yy}^{-1/2} W_y$.

5. The K th leading singular vector T is equal to the correlation $\rho(w_{x,k}^T X, w_{y,k}^T Y)$.

On obtaining W_x, W_y , Canonical Variables i.e., the new projected features are computed as: $Z_x = W_x^T X$ and $Z_y = W_y^T Y$.

However, CCA works for two views at a time. To utilise multiple views in a single time, MCCA can be employed. Let’s denote XX^T as the Covariance matrix C_{xx} . Similarly, XY^T would be C_{xy} and so on. Using the transformations w_k, w_q MCCA can be computed,

where k, q are the set of indices of the views varying as $k, q = 1, 2, 3 \dots K$ and 'K' are the total number of views. It satisfies the following:

$$\begin{aligned} \arg \max_{w_k, w_q} \frac{\sum_{kq} (X_k w_k)^T (X_q w_q)}{\sqrt{\sum_k (X_k w_k)^T (X_k w_k)} \sqrt{\sum_q (X_q w_q)^T (X_q w_q)}} \\ = \frac{\sum_{kq} w_k^T C_{kq} w_q}{\sqrt{\sum_k w_k^T C_{kk} w_k} \sqrt{\sum_q w_q^T C_{qq} w_q}} \end{aligned} \quad (8)$$

The denominators can be fixed as $w_k^T C_{kk} w_k = 1$, since we are interested in the direction of w_k and we get :

$$\begin{aligned} \arg \max_{w_k, w_q} \sum_{kq} w_k^T C_{kq} w_q \\ \text{s.t. } \sum_k w_k^T C_{kk} w_k = 1 \sum_{kq} w_k^T C_{kq} w_q \end{aligned} \quad (9)$$

By using Lagrange multipliers λ and equating the partial derivatives to zero, above optimization can be solved. This sums up the MCCA as:

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ C_{21} & C_{22} & \dots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \dots & C_{KK} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} \\ = \lambda \begin{bmatrix} C_{11} & 0 & \dots & 0 \\ 0 & C_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{KK} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} \end{aligned} \quad (10)$$

To obtain CCA for just two views would be a special case of the above equations.

3.3. Support Vector Machine

SVM constitute a powerful class of supervised machine learning algorithms widely employed in classification and regression tasks. Developed by Cortes et al. [63], SVM has since gained popularity for their effectiveness in handling high-dimensional data and achieving robust generalization performance. For a binary classification problem with two classes, let's denote the classes as +1 and -1. The goal of SVM is to find a hyperplane represented by the Equation (11) :

$$f(x) = w \cdot x + b \quad (11)$$

where w is the weight vector, x is the input vector and b is the bias term. The decision rule for classification is based on the sign of $f(x)$:

$$\text{Predict class } y = \begin{cases} +1 & \text{if } f(x) \geq 0 \\ -1 & \text{if } f(x) < 0 \end{cases} \quad (12)$$

The goal of SVM is to maximise the margin i.e., the distance between the two closest points of different classes, which are the support vectors. The distance from a data point to the decision boundary is given by the equation:

$$\text{Distance} = \frac{|f(x)|}{\|w\|} \quad (13)$$

where w represents the Euclidean magnitude of the weight vector. The objective of SVM is to maximise the margin while reducing the classification errors. It is achieved by solving optimization problem as:

$$\text{Minimise } \frac{1}{2} \|\mathbf{w}\|^2 \quad (14)$$

subjected to the constraints,

$$y_i(w \cdot x_i + b) \geq 1 \quad (15)$$

for all data points (x_i, y_i) where, y_i is the class label of the i -th data point. Whenever the data is not linearly separable, SVM uses Kernel Trick i.e., it uses a mathematical function to transform the data into higher dimensions where it is separable [64].

4. Proposed Technique

The hyperspectral image is envisioned to be segmented by utilizing the hyperspectral data's block structure in a multi-view manner. The entire hypercube is automatically divided into disjoint sub-blocks along the spectral dimension. An overview of the proposed technique to perform semantic segmentation on HSI is shown in Figure 5.

With each spectral band essentially representing a distinct “view” of the same scene. These views encapsulate valuable information pertaining to the unique spectral characteristics of the materials and objects present in the scene. By treating each spectral band as an individual view, CCA can be employed to uncover meaningful relationships and correlations among these views. The key insight behind this approach lies in the fact that different spectral bands may convey complementary information. While some bands may excel in capturing the spectral signatures of specific materials, others may be more adept at highlighting spatial or structural details. By applying CCA to these multi-views, it becomes possible to extract features that are not only representative of the spectral information but also incorporate spatial characteristics. However, considering each spectral band as a different view for CCA hikes the computational complexity.

In practical terms, hyperspectral images are divided into subsets or views, with each subset corresponding to a specific wavelength or band. CCA is then applied to identify linear combinations of bands that exhibit the highest correlation. These linear combinations effectively serve as new features, encapsulating both spectral and spatial information from the original dataset.

4.1. Multi-View Feature Extraction

CCA is a multivariate statistical technique used to explore relationships across two datasets at a time. The algorithm begins with the pre-processing step, standardizing each dataset to have zero mean and unit variance. This normalization ensures that each dataset contributes equally to the analysis, regardless of differences in scale. CCA works on two views at a time, and is referred as PCCA in our paper.

Alongwith PCCA, we have also utilised MCCA because the former considers the information locally among two datasets at a time only. Whereas, MCCA provides simultaneous exploration of shared information across all the datasets via reduced features, thus also reducing the overload and time. In the context of MCCA, PCCA is a special case when there are only two datasets.

The next step involves formulating covariance matrices for within-dataset and between-dataset relationships. These matrices, denoted as C_{ii} and C_{ij} , capture the statistical associations within each dataset and the relationships between different datasets. Solving the generalized eigenvalue problem leads to the identification of canonical vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ and their corresponding canonical correlations $\lambda_1, \lambda_2, \dots, \lambda_m$. These canonical variables represent linear combinations of the original features that maximize the associations between datasets.

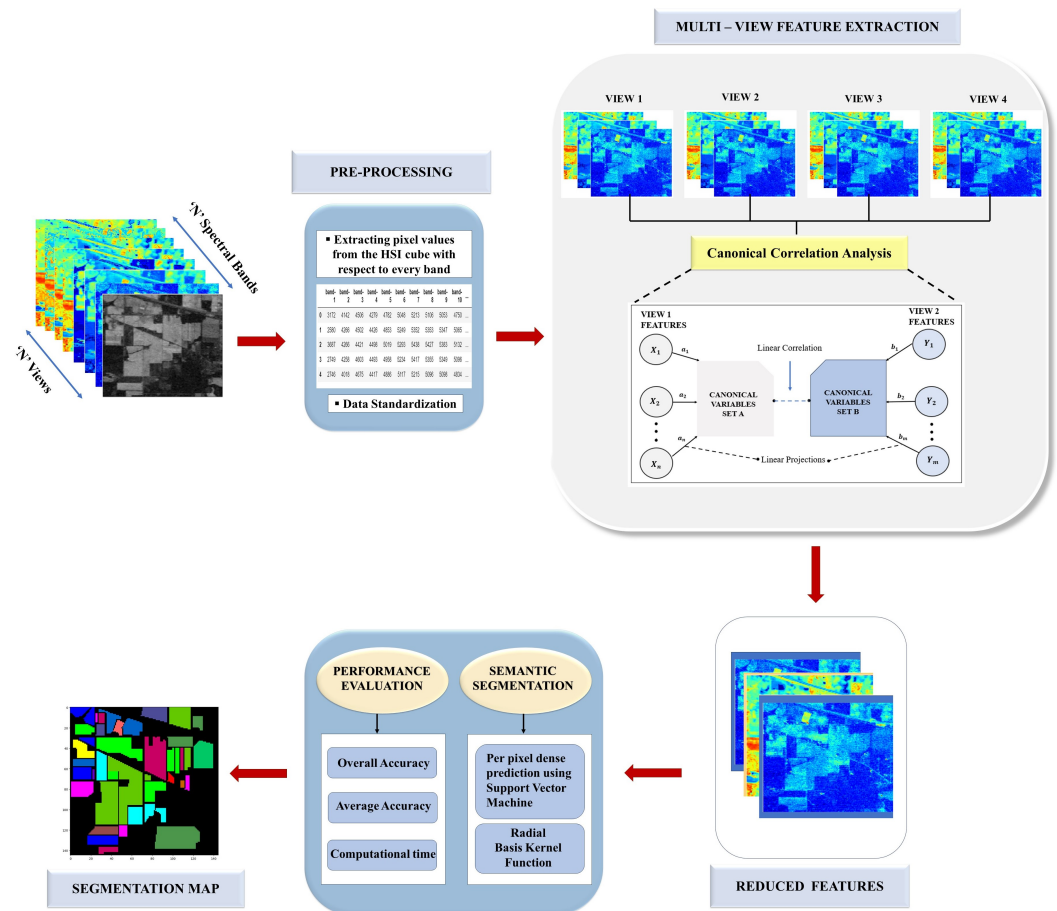


Figure 5. Overview of the Proposed Technique for Semantic Segmentation of HSI. (The same image scene captured at different spectral bands is shown as views with different colors).

The output of CCA, including the canonical vectors and canonical correlations, can be further utilized in various applications. In this study, the exploration of hyperspectral image data involves the application of CCA to unveil intricate relationships among various views and reduce dimension of HSI. The derived canonical variables, representing consolidated information from different spectral views, serve as reduced features that encapsulate the underlying multi-view structure. These canonical variables, extracted through the synergy of MCCA, are instrumental in capturing the shared information across multiple datasets. Leveraging these reduced features, subsequent analyses, such as semantic segmentation using SVM, gained a comprehensive understanding of the inter-dataset relationships. The Canonical variables provide a compact representation that enhances the efficiency and interpretability of downstream tasks, thereby contributing to the advancement of hyperspectral image analysis techniques.

Treating each spectral band as a different view, HSI gives us the automatic opportunity to exploit it for CCA and reduce the dimension of original HSI cube. Rather than picking every band as a view for CCA, we divided the whole dataset into subsequent groups of views. This reduced the computational overload for CCA and gave better results. For one single HSI dataset with hundreds of bands, $K = 4$ number of views were generated by combining together spectral bands from the consecutive wavelengths. This became four different datasets X_1, X_2, X_3, X_4 for CCA to act upon.

4.2. Semantic Segmentation Using Support Vector Machine

SVM is a powerful machine learning algorithm commonly used for classification tasks. In this paper, we focus on two fundamental tasks: per-pixel prediction and semantic segmentation. Per-pixel prediction involves the classification of each individual pixel

within an HSI dataset into specific land-cover categories. Semantic segmentation, on the other hand, extends this concept by partitioning the entire image into coherent regions, each corresponding to a distinct class or category. The algorithm begins by splitting the dataset into training and testing sets to evaluate the model's performance. To ensure consistent and meaningful comparisons, the feature matrix obtained from CCA is normalized to have zero mean and unit variance. This step is crucial for preventing certain features from dominating the learning process due to their larger scales.

The selection of an appropriate kernel function and regularization parameter, C is a pivotal aspect of SVM which has been selected through Grid Search Optimization. The kernel computes the similarity between data points in the transformed space, where the choice of the kernel bandwidth parameter, σ , is crucial in capturing the relationships between the features. The SVM model is trained using the normalized feature matrix, incorporating the kernel matrix that quantifies the pairwise similarities between data points. C balances the trade-off between achieving a low training error and a simple model, helping prevent overfitting.

Once trained, the SVM model is capable of predicting the class labels of new data points. This is achieved by computing the decision function, which evaluates the weighted sum of the Lagrange multipliers (α_i) and the corresponding labels (y_i) for the support vectors. The model's performance is then assessed on the testing set, using standard metrics such as accuracy. Overall, SVM is a versatile and effective algorithm for semantic segmentation tasks, providing a robust solution when combined with feature reduction techniques like CCA.

The Proposed Technique of extracting features from the multi-view dataset and performing Semantic Segmentation using SVM has been discussed in Algorithm 1.

Algorithm 1 Semantic Segmentation using Multi-View Features Algorithm

Require: Hyperspectral image dataset X_n^m with n pixel samples and m spectral bands, Number of Subsets/Views K .

Ensure: Predicted Classes of each Pixel $C = 1, 2, 3 \dots C_i$.

- 1: **begin**
 - 2: Divide the hyperspectral image dataset X into K subsets, each containing adjacent spectral bands. Denote these subsets as $\{X_i\}_{i=1}^K$.
 - 3: **for** j from 3 to 10 **do**
 - 4: Standardize the data in every X_i to have zero mean and unit variance.
 - 5: Perform CCA to extract Canonical Variables U_j .
 - 6: Combine the least correlated U_j to form a feature matrix U .
 - 7: **end for**
 - 8: Hyper-tune the different Parameters of SVM like kernel function, C and γ using Grid Search Optimization.
 - 9: Compute the kernel matrix M based on the chosen best values of kernel function, C and γ .
 - 10: Train the SVM classifier on U .
 - 11: Predict the class labels for each pixel in the hyperspectral image using the trained SVM model.
 - 12: **end**
-

5. Experimental Results and Discussion

The proposed technique has been implemented using Python 3.10.9 on Jupyter Notebook IDE running on the Lenovo Legion 5 Pro system with an Intel Core i7 processor and 16 GB RAM, manufactured in India. The experiments have been performed on real Hyperspectral images which are landcover scenes i.e., IP, Pu and SA. To tackle such high dimension of spectral features, HSI has been employed in a new light of having multiple views and extracted features using CCA variants.

5.1. Evaluation Parameters

The Semantic Segmentation i.e., per pixel classification is done with the help of SVM. The performance of the proposed technique is measured through parameters listed below.

- **Overall Accuracy (OA):-** It measures the percentage of correctly classified samples. It is summation of accurate recognition of samples divided by total number of samples.

$$OA = \frac{\sum_{i=1}^C M_{ii}}{N} \quad (16)$$

where C is the total number of labels/classes. M_{ii} represents the samples that actually belonged to i th class and were predicted to belong to i th class.

- **Average Accuracy (AA):-** It measures the average percentage of correctly classified samples for an individual class.

$$AA = \frac{(\sum_{i=1}^C (M_{ii} / \sum_{i=1}^C M_{ij}))}{C} \quad (17)$$

where C is the total number of labels/classes. M_{ij} represents the samples that actually belonged to i th class and were predicted to belong to j th class.

5.2. Comparative Analysis of PCA and CCA

The analysis of hyperspectral images is very challenging due to many complexities. One such complexity is Hughes phenomena or the curse of dimensionality which when handled beforehand aids in smooth HSI analysis. As discussed earlier, PCA has been the most common technique for dimension reduction. However, it does not focus on HSI as a composition of hundreds of views or single-band datasets. PCA works on single dataset and gives reduced features, capturing maximum variance. With CCA, we have been able to reduce dimension by treating HSI in its original habitat i.e., multiple views. By applying PCCA which works locally between two views at a time, we have achieved higher OA than PCA for semantic segmentation of IP, PU and SA. The paper also implements an improved and faster version of CCA which is MCCA that works on views globally and gives higher OA on all the three datasets. The detailed comparison between PCA, PCCA and MCCA has been highlighted in Tables 1–3.

Tables 1–3 bring out the better performance of CCA as compared to PCA for semantic segmentation with SVM using reduced features of IP, PU and SA datasets resp. The features reduced using the inherent multi-view nature of HSI have given a better performance than the regularly reduced features using PCA. In the earlier studies of PCA implementation for dimension reduction, researches have usually selected the value of Number of Components = 3. Here, the Number of Components means the number of reduced features or Principal Components being selected. For CCA, Number of Components selects the number of Canonical Variables which is a linear combination of original features. Table 1 presents the results of feature extraction with 3 number of components on the IP dataset with the highest OA of 77.02% and 85.75% for PCCA and MCCA, respectively as compared to OA 62.32% by PCA. For the same number of components on PU dataset PCA records an OA of 78.83% whereas a significant jump in OA is observed by PCCA and MCCA, giving an OA of 89.95% and 90.90 %,resp. which indicates better dense pixel prediction. The proposed technique of PCCA and MCCA records the highest OA on SA as compared to IP and PU. For 3 number of components, an OA of 92.15% and 93.88% is recorded for PCCA and MCCA, resp.

Table 1. Comparison of Overall Accuracy of PCA and CCA on IP dataset.

No. of Components	PCA-SVM (%)	PCCA-SVM (%)	MCCA-SVM (%)
3	62.32	77.02	85.75
4	63.25	73.46	89.41
5	64.14	75.07	89.56
6	64.42	73.31	88.92
7	64.61	72.53	88.24
8	64.87	71.02	88.09
9	65.44	70.90	88.48
10	65.75	69.25	87.46

Table 2. Comparison of Overall Accuracy of PCA and CCA on PU dataset.

No. of Components	PCA-SVM (%)	PCCA-SVM (%)	MCCA-SVM (%)
3	78.83	89.95	90.90
4	81.34	88.56	92.66
5	81.98	88.38	95.07
6	84.86	87.94	94.92
7	86.22	87.45	94.79
8	86.75	87.39	94.55
9	86.83	86.66	94.50
10	87.62	86.43	94.43

Table 3. Comparison of Overall Accuracy of PCA and CCA on SA dataset.

No. of Components	PCA-SVM (%)	PCCA-SVM (%)	MCCA-SVM (%)
3	86.00	92.15	93.88
4	87.80	91.20	94.45
5	88.12	90.33	94.67
6	88.07	89.05	94.64
7	88.38	88.36	94.62
8	89.16	88.01	94.88
9	89.18	87.51	95.08
10	89.22	87.61	95.04

Although, choosing No. of Components = 3 can be the criteria according to earlier researches too but for experimental and comparison point of view, the performance of PCA and CCA has been compared against a range of 3, 4, 5 . . . 10. It has shown superiority of multi-view features extraction using CCA as compared to PCA. Figures 6–8 show the resultant segmentation maps obtained from the PCA-SVM, PCCA-SVM and MCCA-SVM. These have been compared together with the Ground Truth images of the datasets. These figures show comparison of ground truth, PCA, PCCA and MCCA on IP, PU and SA, resp. where MCCA showcases the best segmentation map with lesser misclassified pixels (seen as noise in the map) followed by PCCA. It is evident from the images of better segmentation and lesser misclassified pixels using the Proposed Technique with CCA variants.

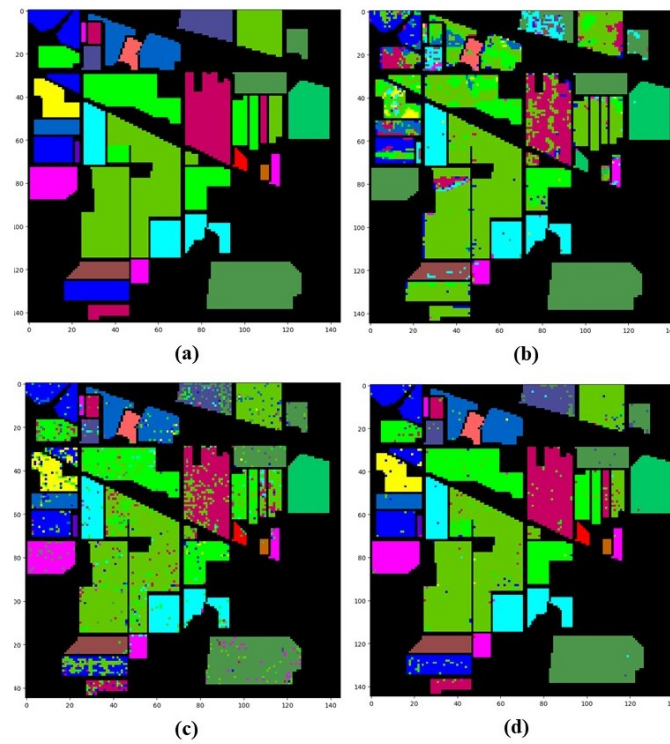


Figure 6. Comparison of PCA and CCA on IP dataset (a) Ground truth, (b) Segmentation map after PCA-SVM (c) Segmentation map after PCCA-SVM, (d) Segmentation map after MCCA-SVM.

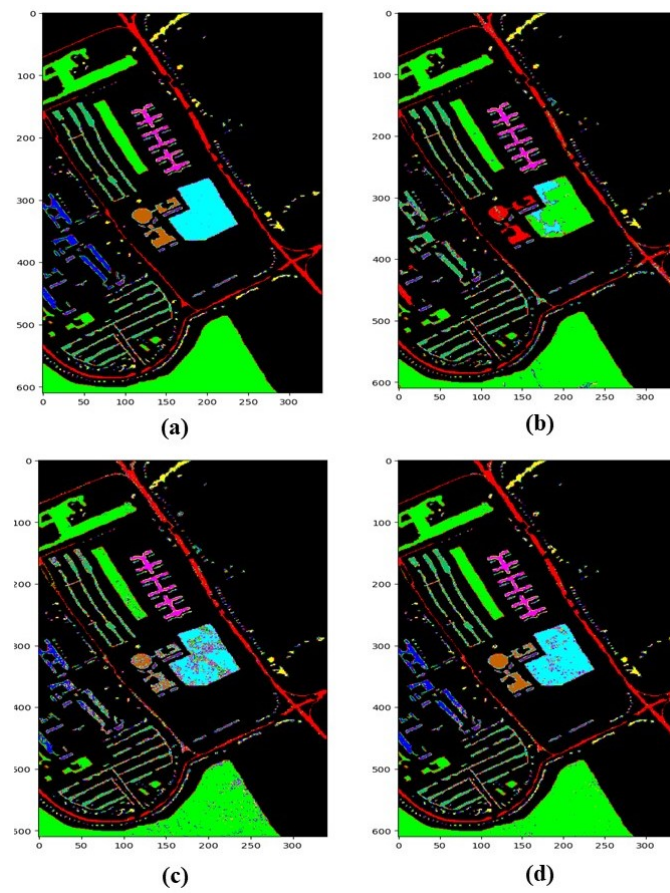


Figure 7. Comparison of PCA and CCA on PU dataset (a) Ground truth, (b) Segmentation map after PCA-SVM (c) Segmentation map after PCCA-SVM, (d) Segmentation map after MCCA-SVM.

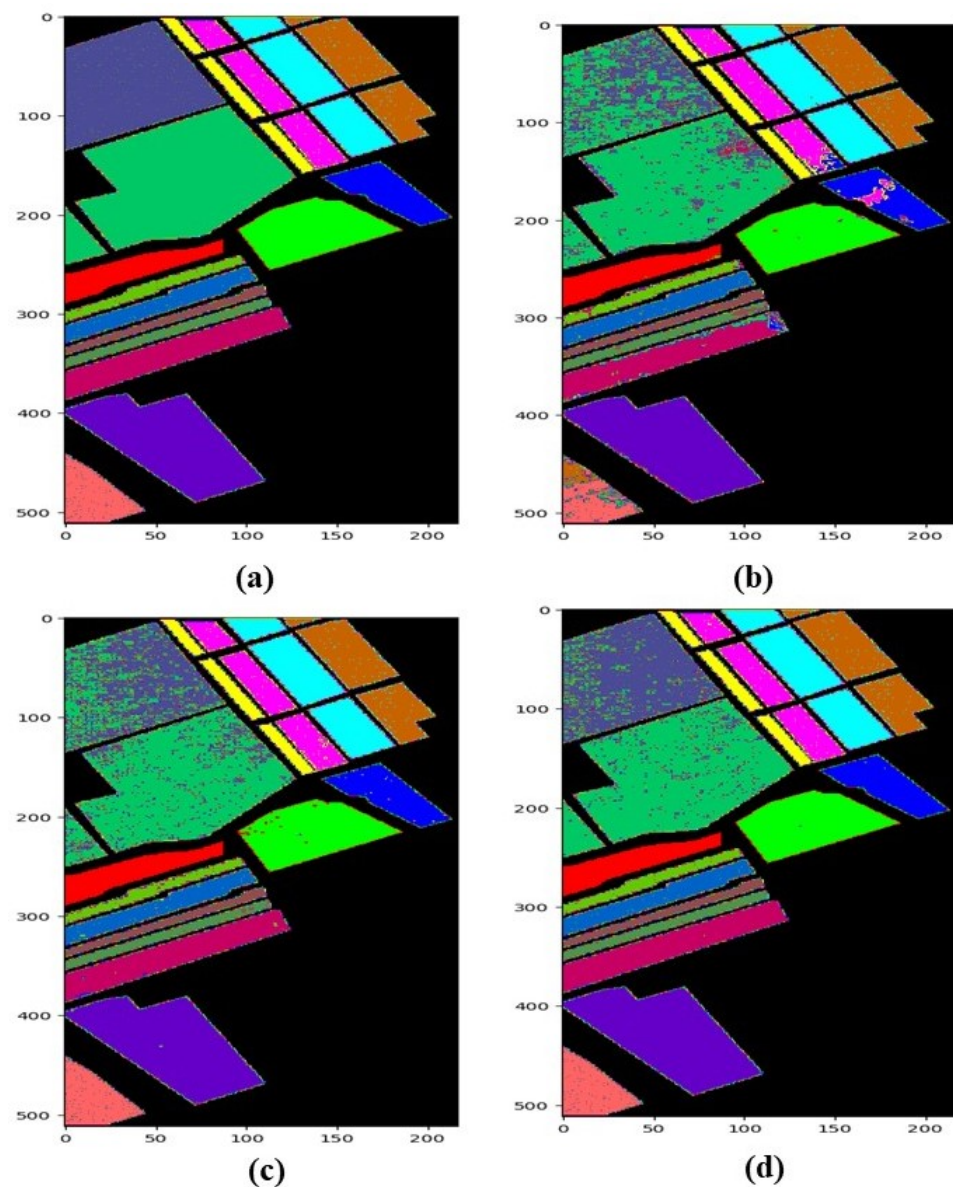


Figure 8. Comparison of PCA and CCA on SA dataset (a) Ground truth, (b) Segmentation map after PCA-SVM (c) Segmentation map after PCCA-SVM, (d) Segmentation map after MCCA-SVM.

5.3. Comparison of Proposed Technique with Existing Techniques

The performance of the proposed technique is compared with other traditional machine learning algorithms i.e., Naive Bayes ([65]), Random Forest ([66]) and k-NN ([67]).

5.3.1. Pairwise Canonical Correlation Analysis

In PCCA, the analysis is conducted independently for each pair of views of dataset, seeking canonical variables that maximize the correlation between the two views. The comparison of different ML algorithms has been evaluated against different number of Canonical Variables i.e., No. of Components. This approach explores the relationships between pairs of datasets individually, providing insights into the shared information between each distinct combination. It has aided us in dimension reduction which enabled semantic segmentation of SVM through reduced, extracted features. Tables 4–6 show the robustness of SVM to handle high dimensionality and with appropriate kernel selection enhancing its performance compared to other machine learning classifiers. Table 4 compares OA of PCCA with different ML techniques on IP where it is evident that SVM when combined with PCCA gives the highest OA of 77.02% followed by RF with OA of 71.80% for number

of components = 3. Similarly, Table 5 highlights supremacy of SVM for PU, with an OA of 89.95% followed by RF with OA of 86.58%. For SA, SVM with CCA recorded the highest OA of 92.15%.

Table 4. Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on IP dataset.

No. of Components	PCCA-SVM (%)	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)
3	77.02	71.80	54.68	69.21
4	73.46	71.95	54.97	69.36
5	75.07	67.36	52.97	63.02
6	73.31	53.90	42.14	47.70
7	72.53	51.95	40.68	47.90
8	71.02	43.75	36.92	40.48
9	70.90	42.97	34.34	38.39
10	69.25	44.5	40.14	37.60

Table 5. Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on PU dataset.

No. of Components	PCCA-SVM (%)	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)
3	89.95	86.58	62.76	83.62
4	88.56	71.85	57.81	66.88
5	88.38	67.91	55.71	66.88
6	87.64	47.76	46.24	41.19
7	87.45	45.05	45.87	38.20
8	87.39	44.91	45.47	37.75
9	86.66	43.25	44.21	35.88
10	86.43	46.13	46.5	38.62

Table 6. Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on SA dataset.

No. of Components	PCCA-SVM (%)	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)
3	92.15	91.32	81.36	89.54
4	91.20	88.92	79.26	85.61
5	90.33	85.10	68.53	78.59
6	89.05	78.93	66.57	73.37
7	88.36	66.92	63.22	67.02
8	88.01	62.95	62.77	61.28
9	87.51	62.09	59.39	60.34
10	87.61	50.66	59.99	57.39

The superiority of the fusion of PCCA and SVM is also evident in the Class-wise accuracies of the HSI dataset for No. of Components = 3 as shown in Tables 7–9. The proposed technique of PCCA-SVM records an AA of 78.00% on IP as shown in Table 7, which is the highest as compared to AA of 70%, 64.00% and 70.50% of RF, Naive bayes and KNN. For PU in Table 8, PCCA-SVM give the highest OA of 87.00% in comparison with the AA of 80%, 72% and 78% of RF, Naive Bayes and KNN. Amongst all the datasets used in the research, highest AA of 96.00% is obtained for SA as shown in Table 9. RF records the second best AA of 95.00% followed by KNN with an AA of 94.00%.

Table 7. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on IP dataset.

Class	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)	Proposed Technique (PCCA-SVM (%))
1	67.00	100.00	83.00	69.00
2	65.00	16.00	68.00	73.00
3	51.00	38.00	55.00	61.00
4	47.00	65.00	41.00	58.00
5	71.00	53.00	79.00	86.00
6	87.00	69.00	85.00	91.00
7	78.00	100.00	100.00	80.00
8	100.00	98.00	100.00	99.00
9	25.00	75.00	25.00	100.00
10	57.00	54.00	63.00	69.00
11	81.00	21.00	73.00	79.00
12	80.00	40.00	70.00	75.00
13	86.00	93.00	93.00	95.00
14	83.00	74.00	65.00	87.00
15	35.00	27.00	29.00	39.00
16	100.00	100.00	100.00	94.00
AA	70.00	64.00	70.50	78.00
OA	71.80	54.68	69.21	77.02

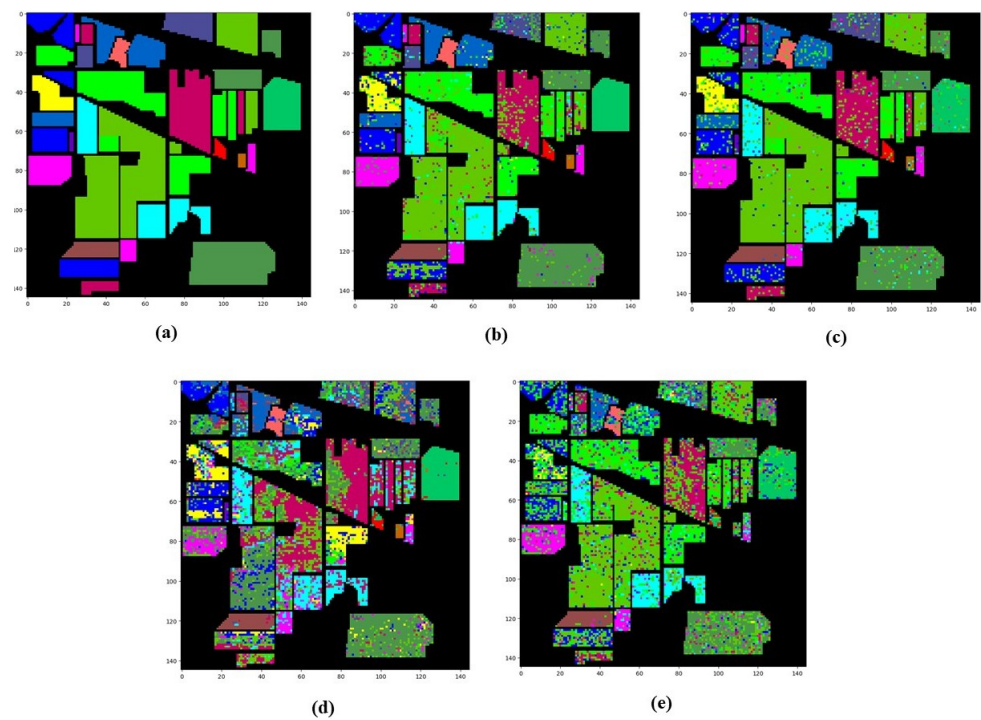
Table 8. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on PU dataset.

Class	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)	Proposed Technique (PCCA-SVM (%))
1	92.00	77.00	86.00	92.00
2	96.00	54.00	95.00	96.00
3	59.00	62.00	60.00	63.00
4	82.00	76.00	75.00	87.00
5	100.00	100.00	100.00	100.00
6	73.00	60.00	64.00	82.00
7	45.00	60.00	55.00	73.00
8	77.00	59.00	68.00	86.00
9	98.00	96.00	98.00	100.00
AA	80.00	72.00	78.00	87.00
OA	86.58	62.76	83.62	89.95

The segmentation maps generated after application of PCCA and different machine learning techniques has been shown in Figures 9–11. The proposed technique gives better segmentation maps in comparison to other techniques. In Figure 9 PCCA-SVM produces the least distorted segmentation map of IP. The segmentation map of Naive Bayes and KNN produce the noisiest segmentation results with highest misclassified pixels. Similarly, for PU and SA in Figures 10 and 11, the proposed technique produces segmentation map as closest to the Ground Truth indicating more accurate semantic segmentation of the landcover.

Table 9. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on SA dataset.

Class	PCCA-RF (%)	PCCA-Naive Bayes (%)	PCCA-KNN (%)	Proposed Technique (PCCA-SVM (%))
1	96.00	83.00	95.00	99.00
2	98.00	96.00	100.00	99.00
3	96.00	80.00	95.00	99.00
4	99.00	98.00	100.00	99.00
5	97.00	84.00	98.00	99.00
6	99.00	96.00	100.00	100.00
7	98.00	97.00	100.00	100.00
8	88.00	55.00	83.00	89.00
9	98.00	93.00	98.00	99.00
10	97.00	89.00	97.00	98.00
11	90.00	91.00	87.00	93.00
12	95.00	80.00	92.00	99.00
13	96.00	97.00	96.00	100.00
14	98.00	93.00	97.00	97.00
15	74.00	74.00	62.00	75.00
16	98.00	97.00	99.00	99.00
AA	95.00	88.00	94.00	96.00
OA	91.32	81.36	89.54	92.15

**Figure 9.** Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on IP dataset (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

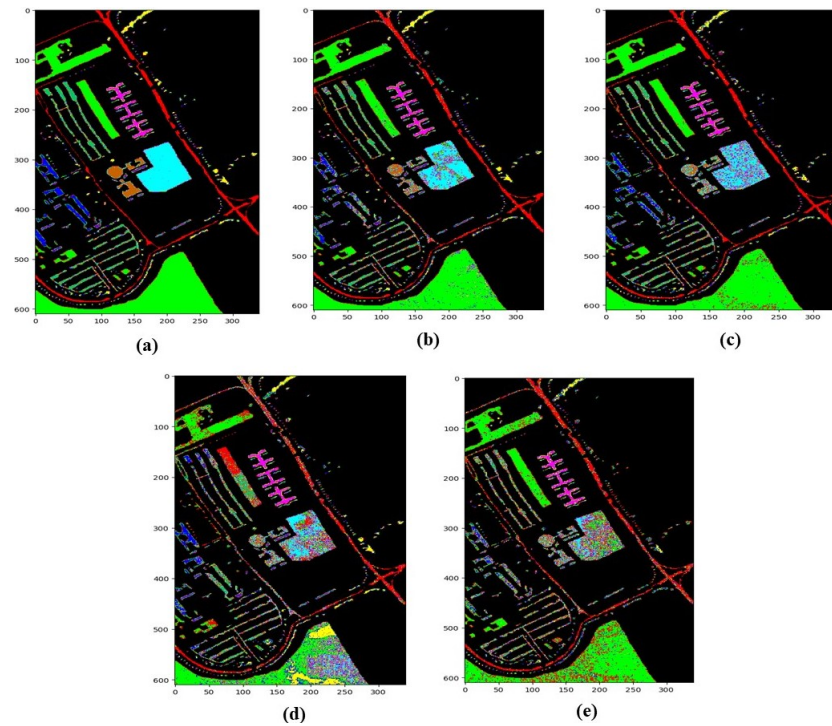


Figure 10. Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on PU dataset (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

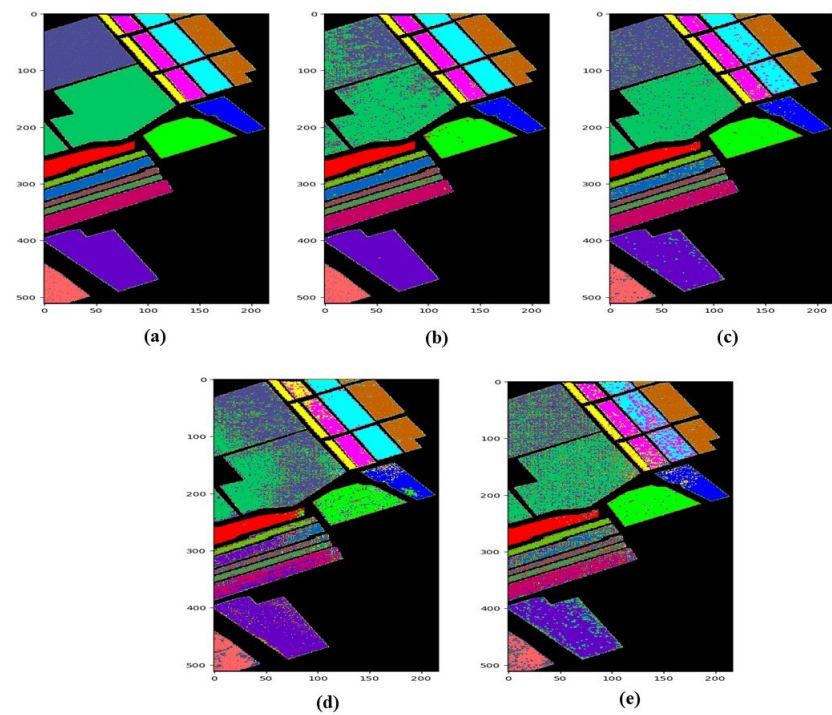


Figure 11. Comparison of Proposed Technique (PCCA-SVM) with Existing Techniques on SA dataset (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

5.3.2. Multiple-Set Canonical Correlation Analysis

The PCCA has given better performance than PCA as evident in Section 5.2. We have applied PCCA on four views of every HSI dataset by selecting two views at a time and reduced the dimensions. On pairing the reduced features with SVM for per pixel prediction i.e., semantic segmentation, it outperforms other machine learning algorithms and gives good performance. However, when working with more than two views, sequential application of pairwise CCA might lead to information loss, as the joint interdependencies among all datasets are not considered simultaneously. In such cases, MCCA becomes advantageous, offering a more comprehensive exploration of relationships in multi-view datasets. The comparison of different existing ML techniques has been evaluated against different number of Canonical Variables i.e., Number of Components. Tables 10–12 highlight the proposed technique i.e., the fusion of MCCA-SVM robustness over other existing techniques.

Table 10. Comparison of Proposed Technique (MCCA-SVM) with Existing Techniques on IP dataset.

No. of Components	MCCA-SVM (%)	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)
3	85.75	83.41	59.12	79.07
4	89.41	87.51	63.17	82.92
5	89.56	87.65	63.27	85.02
6	88.92	87.85	63.07	83.95
7	88.24	87.24	62.24	84.09
8	88.09	87.12	64.48	84.14
9	88.48	87.31	63.65	83.41
10	87.46	86.63	63.75	82.97

Table 11. Comparison of Proposed Technique (MCCA-SVM) with Existing Techniques on PU dataset.

No. of Components	MCCA-SVM (%)	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)
3	90.90	84.59	71.76	82.24
4	92.66	88.94	74.66	86.61
5	95.07	92.85	79.13	91.07
6	94.92	93.27	80.31	91.70
7	94.79	92.97	80.58	90.76
8	94.55	91.87	81.42	89.60
9	94.50	91.85	82.46	89.25
10	94.43	91.68	82.41	88.53

Table 12. Comparison of Proposed Technique (MCCA-SVM) with Existing Techniques on SA dataset.

No. of Components	MCCA-SVM (%)	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)
3	93.88	92.60	82.84	89.59
4	94.45	93.54	85.53	90.57
5	94.67	93.78	85.82	90.61
6	94.64	94.41	86.50	90.21
7	94.62	94.61	88.34	90.66
8	94.88	94.69	88.98	90.95
9	95.08	95.26	88.25	91.09
10	95.04	95.60	89.34	91.02

Table 10 compares OA of MCCA with different ML techniques on IP where it is evident that SVM when combined with MCCA gives the highest OA of 85.75% followed by RF with OA of 83.41% for number of components = 3. Similarly, Table 11 highlights supremacy of SVM for PU, with an OA of 90.90% followed by RF with OA of 84.59%. For SA, SVM with MCCA recorded the highest OA of 93.88%.

The superiority of the fusion of MCCA and SVM is also evident in the Class-wise accuracies of the HSI dataset for No. of Components = 3 as shown in Tables 13–15. The proposed technique of MCCA-SVM records an AA of 85.75% on IP as shown in Table 13, which is the highest as compared to AA of 80%, 64.00% and 79.50% of RF, Naive bayes and KNN. For PU in Table 14, MCCA-SVM give the highest OA of 88.50% in comparison with the AA of 79%, 73% and 78% of RF, Naive Bayes and KNN. Amongst all the datasets used in the research, highest AA of 97.00% is obtained for SA as shown in Table 15. RF records the second best AA of 96.00% followed by KNN with an AA of 94.00%.

Table 13. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on IP dataset.

Class	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)	Proposed Technique (MCCA-SVM (%))
1	67.00	83.00	100.00	100.00
2	79.00	47.00	76.00	80.00
3	66.00	36.00	64.00	70.00
4	84.00	37.00	71.00	92.00
5	96.00	30.00	96.00	94.00
6	97.00	83.00	97.00	96.00
7	67.00	100.00	89.00	89.00
8	100.00	89.00	99.00	97.00
9	50.00	50.00	50.00	50.00
10	76.00	56.00	70.00	83.00
11	85.00	53.00	77.00	85.00
12	76.00	44.00	68.00	83.00
13	100.00	95.00	100.00	100.00
14	98.00	98.00	96.00	97.00
15	47.00	29.00	32.00	62.00
16	94.00	100.00	94.00	100.00
AA	80.00	64.00	79.50	85.75
OA	83.41	59.12	79.07	86.00

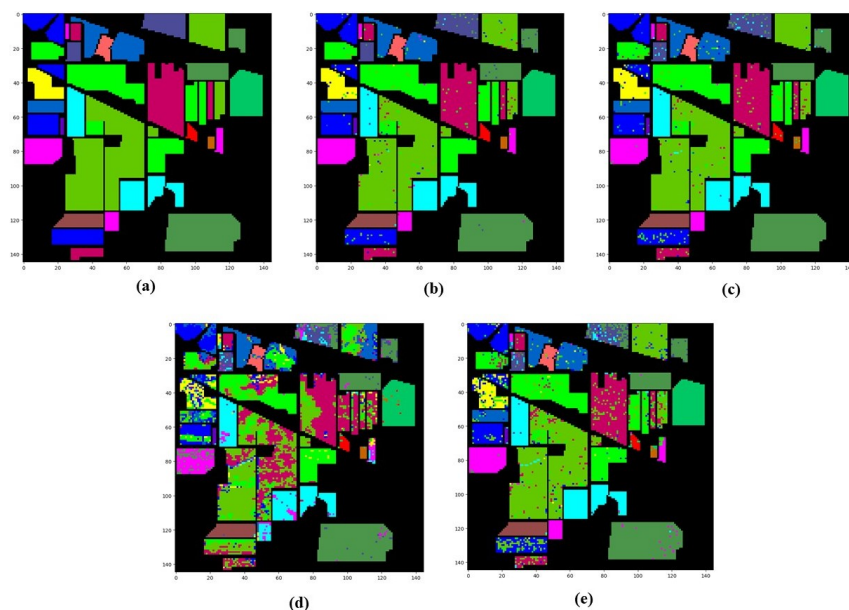
Table 14. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on PU dataset.

Class	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)	Proposed Technique (MCCA-SVM (%))
1	90.00	70.00	84.00	90.00
2	97.00	78.00	94.00	96.00
3	53.00	28.00	54.00	78.00
4	90.00	88.00	84.00	93.00
5	100.00	100.00	100.00	100.00
6	41.00	31.00	45.00	63.00
7	54.00	80.00	65.00	82.00
8	87.00	86.00	78.00	88.00
9	100.00	99.00	100.00	99.00
AA	79.00	73.00	78.00	88.50
OA	84.59	71.76	82.24	90.90

Table 15. Comparison of Class-wise Accuracy of the Proposed Technique with Existing Techniques on SA dataset.

Class	MCCA-RF (%)	MCCA-Naive Bayes (%)	MCCA-KNN (%)	Proposed Technique (MCCA-SVM (%))
1	99.00	92.00	97.00	99.00
2	99.00	97.00	98.00	100.00
3	97.00	73.00	93.00	99.00
4	100.00	99.00	100.00	100.00
5	98.00	97.00	97.00	98.00
6	99.00	96.00	100.00	100.00
7	100.00	99.00	100.00	100.00
8	90.00	53.00	82.00	88.00
9	99.00	93.00	97.00	99.00
10	98.00	93.00	98.00	99.00
11	95.00	90.00	91.00	93.00
12	100.00	96.00	99.00	100.00
13	98.00	98.00	97.00	97.00
14	99.00	94.00	96.00	99.00
15	75.00	75.00	62.00	73.00
16	99.00	97.00	99.00	100.00
AA	96.00	90.00	94.00	97.00
OA	92.60	82.84	89.59	93.88

The segmentation maps generated after application of the Proposed Technique (MCCA-SVM) and existing techniques have been shown in Figures 12–14. The proposed MCCA-SVM gives better segmentation maps in comparison to other techniques with MCCA. In Figure 12 MCCA-SVM produces the best segmentation map of IP closest to the Ground Truth as compared to other techniques. The segmentation map of Naive Bayes produced the noisiest segmentation results with highest misclassified pixels. Similarly, for PU and SA in Figures 13 and 14, the proposed technique produces segmentation map as closest to the Ground Truth indicating more accurate semantic segmentation of the landcover.

**Figure 12.** Comparison of MCCA and different ML techniques on IP (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

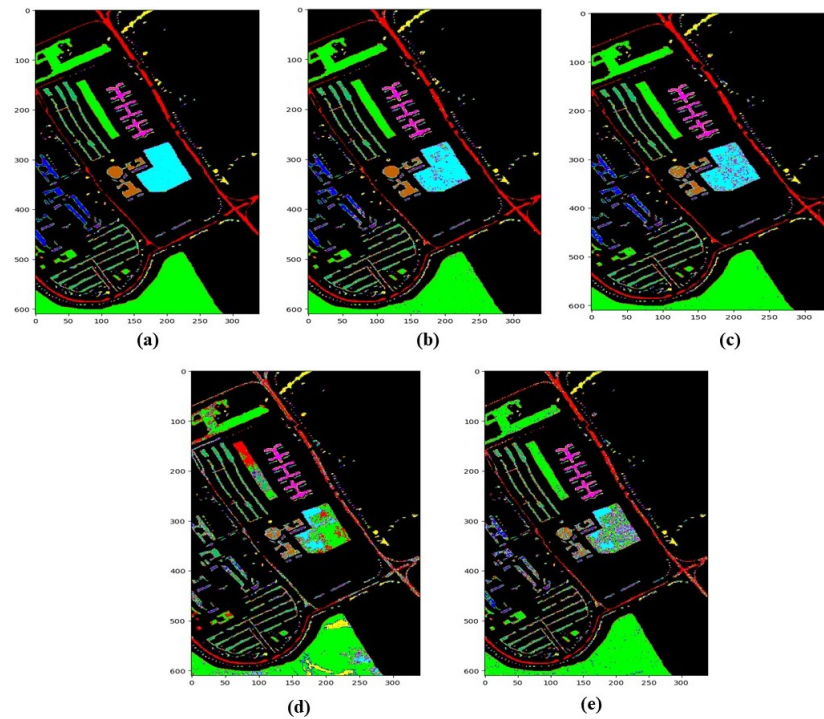


Figure 13. Comparison of MCCA and different ML techniques on PU (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

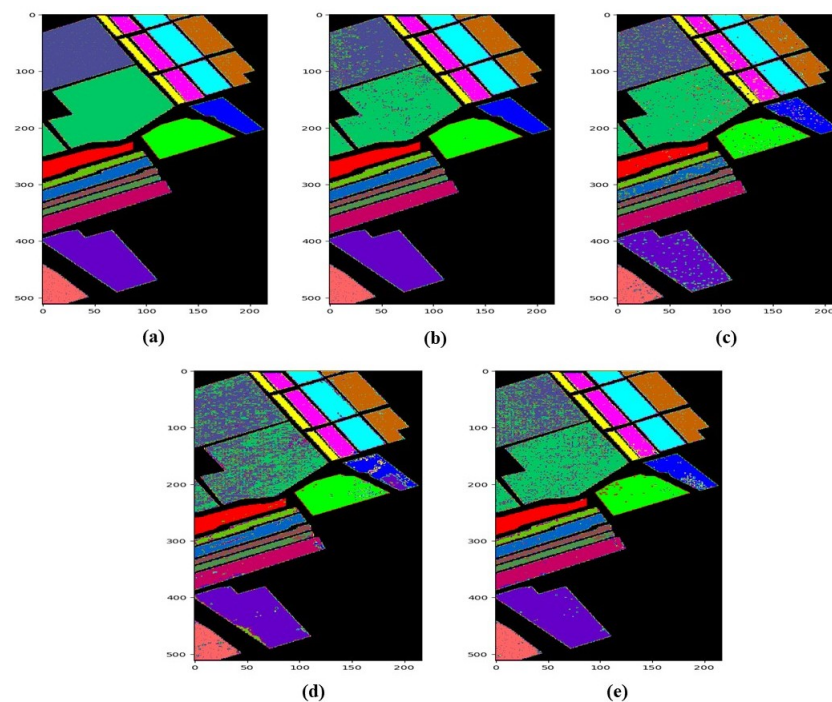


Figure 14. Comparison of MCCA and different ML techniques on SA (a) Ground Truth, (b) Segmentation map after SVM, (c) Segmentation map after RF, (d) Segmentation map after Naive Bayes, (e) Segmentation map after KNN.

It is evident from Tables 13–15 that the Proposed Technique has performed better than existing techniques in terms of the metric ‘Average Accuracy (AA)’. For IP, PU and SA dataset, Highest AA of 85.75%, 88.50% and 97.00% is given by the Proposed Technique (MCCA-SVM).

5.3.3. Comparison between Semantic Segmentation Performed Using the Proposed Techniques PCCA-SVM and MCCA-SVM

Understanding HSI as a multi-view dataset and application of CCA has given better results than PCA-SVM for semantic segmentation of HSI. With the implementation of MCCA in conjunction with SVM has notably improved computational results in HSI segmentation when compared to its counterpart, PCCA. PCCA involves multiple analyses, each focusing locally on a pair of views, potentially leading to scalability issues as the number of views increases. In contrast, MCCA addresses the dimensionality challenge by jointly analyzing all views in a unified framework. MCCA excels in handling multiple sets of data simultaneously, enabling a more comprehensive exploration of spectral relationships within hyperspectral scenes. The enhanced computational efficiency of MCCA arises from its parallelism and intrinsic capacity to consider correlations across multiple sets of spectral bands, leading to a better understanding of the intricate relationships present in hyperspectral data. This extension not only refines the characterization of spectral information but also contributes to the optimization of the SVM-based classification process.

By incorporating a broader spectrum of correlations, MCCA provides a richer feature space for SVM, facilitating more accurate and discriminative classification outcomes. Figure 15 highlights the improved OA and AA of MCCA with respect to PCCA.

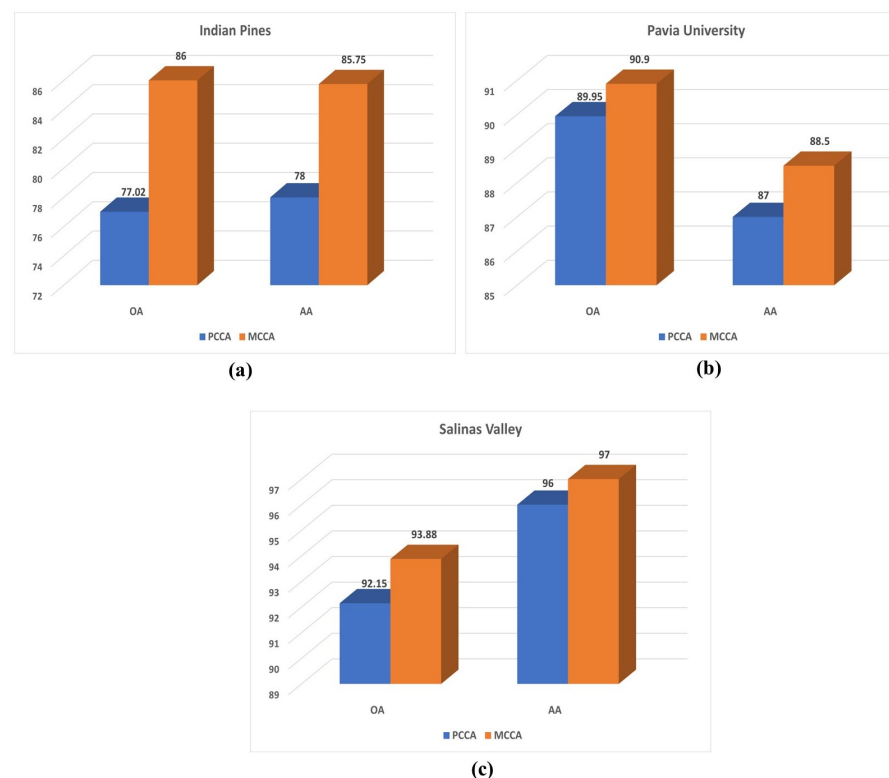


Figure 15. Comparison of PCCA-SVM and MCCA-SVM in terms of OA and AA on (a) Indian Pines, (b) Pavia University and (c) Salinas Valley.

Table 16 highlights the improvement in computational time of MCCA-SVM as compared to PCCA-SVM. For simplicity, No. of Components is specified as ‘n_comps’ in the Table 16. The results in Table 16 and Figure 16 have been rigorously studied and compared by varying No. of Components for all the three HSI datasets where MCCA-SVM outshines PCCA-SVM.

It is evident in Table 16 that MCCA-SVM has given results in lesser computational time as compared to PCCA-SVM. The lowest time is recorded as 3.02 s for IP dataset, 6.23 s for PU dataset and 8.59 s for SA dataset by Semantic Segmentation using MCCA-SVM.

Additionally, MCCA-SVM takes an average of 3.58 s, 8.19 s, and 12.22 s on IP, PU, and SA, respectively. Due to variations in spatial resolution, the time curves for the three datasets differ in Figure 16. IP has the lowest computational time average followed by PU and SA. PU and SA have approximately 4.2 and 5.4 times more pixel samples than IP, respectively, which suggests their higher time curves than IP's.

Table 16. Time comparison between PCCA-SVM and MCCA-SVM on HSI datasets.

n_comps	Indian Pines		Pavia University		Salinas Valley	
	PCCA-SVM	MCCA-SVM	PCCA-SVM	MCCA-SVM	PCCA-SVM	MCCA-SVM
3	29.34 s	3.38 s	14.57 s	6.23 s	23.60 s	8.59 s
4	38.98 s	3.23 s	48.35 s	6.60 s	32.62 s	9.83 s
5	43.48 s	3.02 s	71.45 s	7.33 s	49.06 s	9.96 s
6	60.33 s	3.36 s	114.63 s	7.49 s	70.77 s	10.63 s
7	62.44 s	3.73 s	210.43 s	8.33 s	108.98 s	11.77 s
8	68.24 s	3.80 s	250.99 s	9.35 s	135.11 s	13.27 s
9	73.83 s	4.07 s	273.71 s	9.99 s	154.00 s	15.86 s
10	86.90 s	4.08 s	300.11 s	10.26 s	201.33 s	17.89 s

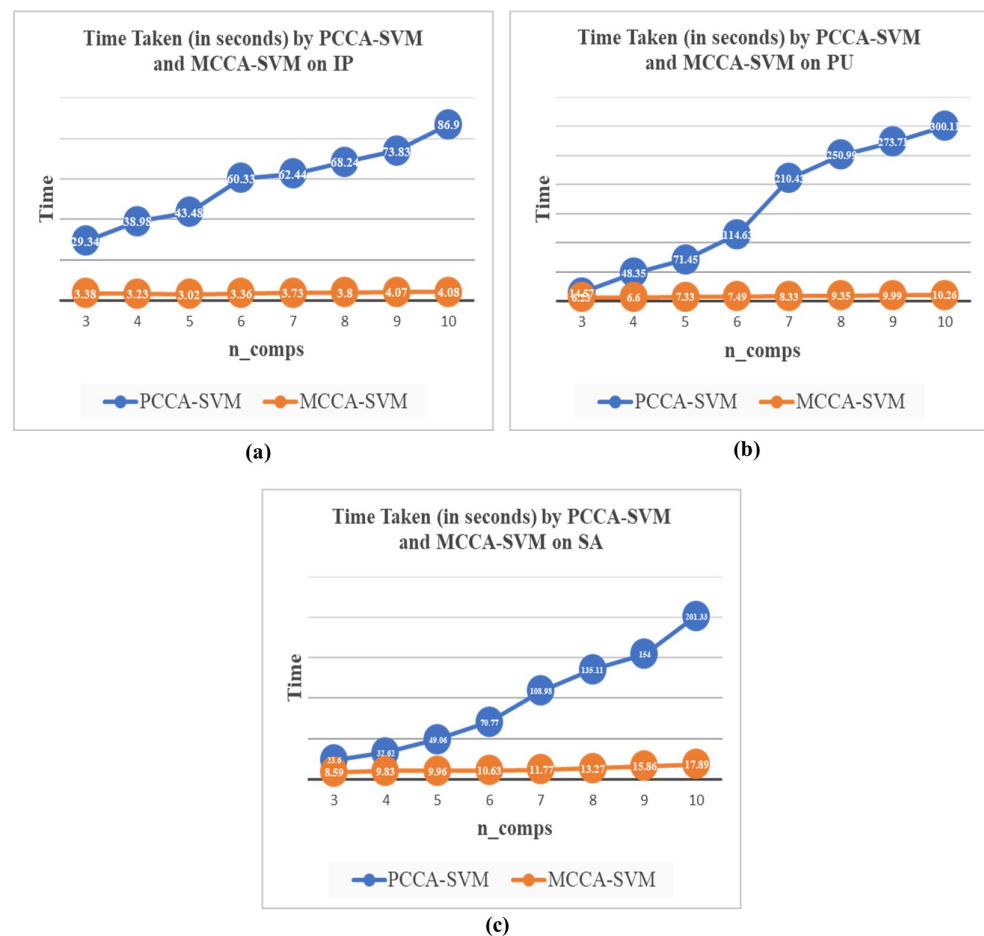


Figure 16. Comparison of the time taken by PCCA-SVM and MCCA-SVM on (a) Indian Pines, (b) Pavia University and (c) Salinas Valley.

6. Conclusions and Future Scope

In conclusion, this research has presented a comprehensive exploration of Canonical Correlation Analysis variants, namely PCCA and MCCA, as powerful techniques in fusion with SVM for hyperspectral image semantic segmentation. Through rigorous experimentation and evaluation on benchmark datasets, the efficacy of these CCA variants in capturing meaningful spectral relationships has been demonstrated. The approach exhibits notable performance in preserving fine-grained details and accurately delineating land cover classes within hyperspectral scenes. The Proposed Technique's ability to mitigate the curse of dimensionality, effectively exploit spectral information, and adapt to varying scene complexities positions it as a promising solution for HSI segmentation challenges. The fusion of PCCA and MCCA facilitates the capture of pairwise and collective spectral correlations, offering a more holistic representation of hyperspectral data. Furthermore, the scalability of the approach has been evidenced through experiments on datasets with varying spatial and spectral resolutions. The consistent performance across different scenarios underscores the robustness of the CCA-based technique. As the research community endeavours to advance remote sensing capabilities, the suggested future directions emphasize the potential for further refinement and extension of the Proposed Technique. Integration with deep learning models, adaptation to larger datasets, exploration of unsupervised learning scenarios, and incorporation of multimodal data present exciting opportunities for enhancing the accuracy and versatility of hyperspectral image segmentation. In essence, the findings presented in this research contribute to the evolving landscape of remote sensing applications, providing a foundation for future innovations and improvements in hyperspectral image analysis. The demonstrated capabilities of CCA variants pave the way for more sophisticated and effective techniques, fostering advancements in environmental monitoring, agriculture, and resource management.

Author Contributions: R.G.; Conceptualization, Methodology, Result analysis, Visualization, Writing—original draft. G.K.; Investigation and Supervision. S.S.K.; Investigation and Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found at <https://rslab.ut.ac.ir/data> (accessed on 8 August 2023).

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

References

1. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [CrossRef]
2. Yang, C.; Everitt, J.H.; Bradford, J.M. Yield estimation from hyperspectral imagery using spectral angle mapper (SAM). *Trans. ASABE* **2008**, *51*, 729–737. [CrossRef]
3. Fagan, M.E.; DeFries, R.S.; Sesnie, S.E.; Arroyo-Mora, J.P.; Soto, C.; Singh, A.; Townsend, P.A.; Chazdon, R.L. Mapping species composition of forests and tree plantations in Northeastern Costa Rica with an integration of hyperspectral and multitemporal Landsat imagery. *Remote Sens.* **2015**, *7*, 5660–5696. [CrossRef]
4. Burger, J.; Gowen, A.A. The interplay of chemometrics and hyperspectral chemical imaging. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–4.
5. Manian, V.; Velez-Reyes, M. Support vector classification of land cover and benthic habitat from hyperspectral images. *Int. J. High Speed Electron. Syst.* **2008**, *18*, 337–348. [CrossRef]
6. Wang, X.; Luo, G.; Tian, L. Application of hyperspectral image anomaly detection algorithm for Internet of things. *Multimed. Tools Appl.* **2019**, *78*, 5155–5167. [CrossRef]
7. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]

8. Fischer, P.; Azimi, S.M.; Roschlaub, R.; Krauß, T. Towards hd maps from aerial imagery: Robust lane marking segmentation using country-scale imagery. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 458. [\[CrossRef\]](#)
9. Nassar, A.; Amer, K.; ElHakim, R.; ElHelw, M. A deep CNN-based framework for enhanced aerial imagery registration with applications to UAV geolocalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1513–1523.
10. Yuan, X.; Sarma, V. Automatic urban water-body detection and segmentation from sparse ALSM data via spatially constrained model-driven clustering. *IEEE Geosci. Remote Sens. Lett.* **2010**, *8*, 73–77. [\[CrossRef\]](#)
11. Jadhav, J.K.; Singh, R. Automatic semantic segmentation and classification of remote sensing data for agriculture. *Math. Model. Eng.* **2018**, *4*, 112–137. [\[CrossRef\]](#)
12. Fang, B.; Li, Y.; Zhang, H.; Chan, J.C.W. Hyperspectral images classification based on dense convolutional networks with spectral-wise attention mechanism. *Remote Sens.* **2019**, *11*, 159. [\[CrossRef\]](#)
13. Grewal, R.; Kasana, S.S.; Kasana, G. Hyperspectral image segmentation: A comprehensive survey. *Multimed. Tools Appl.* **2023**, *82*, 20819–20872. [\[CrossRef\]](#)
14. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [\[CrossRef\]](#)
15. Datta, A.; Ghosh, S.; Ghosh, A. PCA, kernel PCA and dimensionality reduction in hyperspectral images. In *Advances in Principal Component Analysis: Research and Development*; Springer: Singapore, 2018; pp. 19–46.
16. Datta, D.; Mallick, P.K.; Bhoi, A.K.; Ijaz, M.F.; Shafi, J.; Choi, J. Hyperspectral image classification: Potentials, challenges, and future directions. *Comput. Intell. Neurosci.* **2022**, *2022*. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Jia, S.; Zhao, Q.; Zhuang, J.; Tang, D.; Long, Y.; Xu, M.; Zhou, J.; Li, Q. Flexible Gabor-based superpixel-level unsupervised LDA for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10394–10409. [\[CrossRef\]](#)
18. Miclea, A.V.; Terebes, R.M.; Meza, S.; Cislariu, M. On spectral-spatial classification of hyperspectral images using image denoising and enhancement techniques, wavelet transforms and controlled data set partitioning. *Remote Sens.* **2022**, *14*, 1475. [\[CrossRef\]](#)
19. Li, Q.; Zheng, B.; Tu, B.; Wang, J.; Zhou, C. Ensemble EMD-based spectral-spatial feature extraction for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5134–5148. [\[CrossRef\]](#)
20. Liang, N.; Duan, P.; Xu, H.; Cui, L. Multi-View Structural Feature Extraction for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 1971. [\[CrossRef\]](#)
21. Paul, A.; Chaki, N. Band selection using spectral and spatial information in particle swarm optimization for hyperspectral image classification. *Soft Comput.* **2022**, *26*, 2819–2834. [\[CrossRef\]](#)
22. Paul, A.; Chaki, N. Dimensionality reduction of hyperspectral image using signal entropy and spatial information in genetic algorithm with discrete wavelet transformation. *Evol. Intell.* **2021**, *14*, 1793–1802. [\[CrossRef\]](#)
23. Wang, Y.; Zhu, Q.; Ma, H.; Yu, H. A hybrid gray wolf optimizer for hyperspectral image band selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [\[CrossRef\]](#)
24. Phaneendra Kumar, B.L.; Manoharan, P. Whale optimization-based band selection technique for hyperspectral image classification. *Int. J. Remote Sens.* **2021**, *42*, 5105–5143. [\[CrossRef\]](#)
25. Datta, A.; Ghosh, S.; Ghosh, A. Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2814–2823. [\[CrossRef\]](#)
26. Beirami, B.; Mokhtarzade, M. Supervised and unsupervised clustering based dimensionality reduction of hyperspectral data. *Int. J. Eng.* **2021**, *34*, 1407–1412.
27. Chen, Y.; Wu, Z.; Zhao, B.; Fan, C.; Shi, S. Weed and corn seedling detection in field based on multi feature fusion and support vector machine. *Sensors* **2020**, *21*, 212. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Guo, Y.; Zhang, Z.; Tang, F. Feature selection with kernelized multi-class support vector machine. *Pattern Recognit.* **2021**, *117*, 107988. [\[CrossRef\]](#)
29. Zhang, W.; Jin, S.; Zhou, L.; Xie, X.; Wang, F.; Jiang, L.; Zheng, Y.; Qu, P.; Li, G.; Pan, X. Multi-feature embedded learning SVM for cloud detection in remote sensing images. *Comput. Electr. Eng.* **2022**, *102*, 108177. [\[CrossRef\]](#)
30. Rehman, M.U.; Akhtar, S.; Zakwan, M.; Mahmood, M.H. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images. *Biomed. Signal Process. Control* **2022**, *71*, 103212. [\[CrossRef\]](#)
31. Rodarmel, C.; Shan, J. Principal component analysis for hyperspectral image classification. *Surv. Land Inf. Sci.* **2002**, *62*, 115–122.
32. Castaings, T.; Waske, B.; Atli Benediktsson, J.; Chanussot, J. On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile. *Int. J. Remote Sens.* **2010**, *31*, 5921–5939. [\[CrossRef\]](#)
33. Che, W.; Sun, L.; Zhang, Q.; Tan, W.; Ye, D.; Zhang, D.; Liu, Y. Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging. *Comput. Electron. Agric.* **2018**, *146*, 12–21. [\[CrossRef\]](#)
34. Torres, I.; Sánchez, M.T.; Cho, B.K.; Garrido-Varo, A.; Pérez-Marín, D. Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging. *Comput. Electron. Agric.* **2019**, *167*, 105070. [\[CrossRef\]](#)
35. Fernandez, D.; Gonzalez, C.; Mozos, D.; Lopez, S. FPGA implementation of the principal component analysis algorithm for dimensionality reduction of hyperspectral images. *J. Real-Time Image Process.* **2016**, *16*, 1395–1406. [\[CrossRef\]](#)
36. Sun, W.; Du, Q. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3185–3195. [\[CrossRef\]](#)

37. Li, J.; Luo, W.; Wang, Z.; Fan, S. Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method. *Postharvest Biol. Technol.* **2019**, *149*, 235–246. [\[CrossRef\]](#)
38. Uddin, M.P.; Mamun, M.A.; Afjal, M.I.; Hossain, M.A. Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification. *Int. J. Remote Sens.* **2021**, *42*, 286–321. [\[CrossRef\]](#)
39. Foster, D.P.; Kakade, S.M.; Zhang, T. *Multi-View Dimensionality Reduction Via Canonical Correlation Analysis*; Toyota Technical Institute: Chicago, IL, USA, 2008.
40. Andrew, G.; Arora, R.; Bilmes, J.; Livescu, K. Deep canonical correlation analysis. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; PMLR: Atlanta, GA, USA, 2013; pp. 1247–1255.
41. Chen, H.; Liu, J.; Xiao, L. An improved composite kernel framework for hyperspectral image classification using canonical correlation analysis. *Remote Sens. Lett.* **2019**, *10*, 411–420. [\[CrossRef\]](#)
42. Qin, Y.; Bruzzone, L.; Li, B.; Ye, Y. Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3952–3966. [\[CrossRef\]](#)
43. Wang, H.T.; Smallwood, J.; Mourao-Miranda, J.; Xia, C.H.; Satterthwaite, T.D.; Bassett, D.S.; Bzdok, D. Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *NeuroImage* **2020**, *216*, 116745. [\[PubMed\]](#)
44. Patil, A.; Patil, M.; Birajdar, G. White blood cells image classification using deep learning with canonical correlation analysis. *Irbm* **2021**, *42*, 378–389. [\[CrossRef\]](#)
45. Grewal, R.; Singh Kasana, S.; Kasana, G. Machine Learning and Deep Learning Techniques for Spectral Spatial Classification of Hyperspectral Images: A Comprehensive Survey. *Electronics* **2023**, *12*, 488. [\[CrossRef\]](#)
46. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [\[CrossRef\]](#)
47. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1–9. [\[CrossRef\]](#)
48. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
49. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
50. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
51. Niu, B.; Feng, Q.; Chen, B.; Ou, C.; Liu, Y.; Yang, J. HSI-TransUNet: A transformer based semantic segmentation model for crop mapping from UAV hyperspectral imagery. *Comput. Electron. Agric.* **2022**, *201*, 107297. [\[CrossRef\]](#)
52. Kaul, A.; Raina, S. Support vector machine versus convolutional neural network for hyperspectral image classification: A systematic review. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6945. [\[CrossRef\]](#)
53. Scholkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2018.
54. Hu, L.; Qi, C.; Wang, Q. Spectral-spatial hyperspectral image classification based on mathematical morphology post-processing. *Procedia Comput. Sci.* **2018**, *129*, 93–97. [\[CrossRef\]](#)
55. Ye, D.; Sun, L.; Tan, W.; Che, W.; Yang, M. Detecting and classifying minor bruised potato based on hyperspectral imaging. *Chemom. Intell. Lab. Syst.* **2018**, *177*, 129–139. [\[CrossRef\]](#)
56. Ji, Y.; Sun, L.; Li, Y.; Li, J.; Liu, S.; Xie, X.; Xu, Y. Non-destructive classification of defective potatoes based on hyperspectral imaging and support vector machine. *Infrared Phys. Technol.* **2019**, *99*, 71–79. [\[CrossRef\]](#)
57. Akbari, D. Improving spatial-spectral classification of hyperspectral imagery by using extended minimum spanning forest algorithm. *Can. J. Remote Sens.* **2020**, *46*, 146–153. [\[CrossRef\]](#)
58. Wang, Y.; Yu, W.; Fang, Z. Multiple kernel-based SVM classification of hyperspectral images by combining spectral, spatial, and semantic information. *Remote Sens.* **2020**, *12*, 120. [\[CrossRef\]](#)
59. Anand, R.; Veni, S.; Aravindh, J. Robust classification technique for hyperspectral images based on 3D-discrete wavelet transform. *Remote Sens.* **2021**, *13*, 1255. [\[CrossRef\]](#)
60. Guo, Y.; Hastie, T.; Tibshirani, R. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* **2007**, *8*, 86–100. [\[CrossRef\]](#) [\[PubMed\]](#)
61. González, I.; Déjean, S.; Martin, P.G.; Baccini, A. CCA: An R package to extend canonical correlation analysis. *J. Stat. Softw.* **2008**, *23*, 1–14. [\[CrossRef\]](#)
62. Golugula, A.; Lee, G.; Master, S.R.; Feldman, M.D.; Tomaszewski, J.E.; Speicher, D.W.; Madabhushi, A. Supervised regularized canonical correlation analysis: Integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC Bioinform.* **2011**, *12*, 483. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Cortes, C.; Vapnik, V. Support vector machine. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
64. Tharwat, A. Parameter investigation of support vector machine classifier with kernel functions. *Knowl. Inf. Syst.* **2019**, *61*, 1269–1302. [\[CrossRef\]](#)

65. Everingham, M.; Zisserman, A.; Williams, C.K.; Van Gool, L.; Allan, M.; Bishop, C.M.; Chapelle, O.; Dalal, N.; Deselaers, T.; Dorkó, G.; et al. The 2005 pascal visual object classes challenge. In Proceedings of the Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment: First PASCAL Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, 11–13 April 2005; Revised Selected Papers; Springer: Berlin/Heidelberg, Germany, 2006; pp. 117–176.
66. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning: Methods and Applications*; Springer: New York, NY, USA, 2012; pp. 157–175.
67. Deng, Z.; Zhu, X.; Cheng, D.; Zong, M.; Zhang, S. Efficient kNN classification algorithm for big data. *Neurocomputing* **2016**, *195*, 143–148. [[CrossRef](#)]

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