PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification

Article *in* IETE Technical Review · March 2020

DOI: 10.1080/02564602.2020.1740615

3 authors:

Md Palash Uddin
Deakin University

Md. Al Mamun
Rajshahi University of Engineering & Technology

Dr Md Ali Hossain
UNSW Sydney

IETE Technical Review

ISSN: 0256-4602 (Print) 0974-5971 (Online) Journal homepage: https://www.tandfonline.com/loi/titr20

PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification

Md. Palash Uddin, Md. Al Mamun & Md. Ali Hossain

IETE TECHNICAL REVIEW
https://doi.org/10.1080/02564602.2020.1740615

REVIEW ARTICLE

PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification

Md. Palash Uddin [1,2], Md. Al Mamun [1] and Md. Ali Hossain[1]

[1]Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh; [2]Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh

ABSTRACT
The hyperspectral remote sensing images (HSIs) are acquired to encompass the essential information of land objects through contiguous narrow spectral wavelength bands. The classification accuracy is not often satisfactory in a cost-effective way using the entire original HSI for practical applications. To enhance the classification result of HSIs the band reduction strategies are applied which can be divided into feature extraction and feature selection methods. PCA (Principal Component Analysis), a linear unsupervised statistical transformation, is frequently adopted for the extraction of features from HSIs. In this paper, PCA and SPCA (Segmented-PCA), SSPCA (Spectrally Segmented-PCA), FPCA (Folded-PCA) and MNF (Minimum Noise Fraction) as linear variants of PCA together with KPCA (Kernel-PCA) and KECA (kernel Entropy Component Analysis) as nonlinear variants of PCA have been investigated. The top transformed features were picked out using accumulation of variance for all other feature extraction methods except for MNF and KECA. MNF uses SNR (Signal-to-Noise Ratio) values and KECA employs Renyi quadratic entropy measurement for this purpose. The studied approaches are equated and analyzed for Indian Pine agricultural and urban Washington DC Mall HSI classification using SVM (Support Vector Machine) classifier. The experiment illustrates that the costly effective and improved classification performance of the feature extraction approaches over the performance using the entire original dataset. MNF offers the highest classification accuracy and FPCA offers the least space and time complexity with satisfactory classification result.

KEYWORDS
Feature reduction; feature selection; feature extraction; hyperspectral image; FPCA; KECA; KPCA; MNF; PCA; Segmentation-based PCA

1. INTRODUCTION

The hyperspectral remote sensing images (HSIs) are usually taken at massive amount of contiguous narrow spectral wavelengths for the better analysis of the earth objects. The wavelengths range to capture the same ground surface is typically from 400 *nm* to 2500 *nm* covering the visible light to infrared region of the Electromagnetic Spectrum (EM) for acquiring the land objects. These immense bands for obtaining the earth classes facilitate HSIs to employ in various applications such as agriculture, mining, military surveillance, verification counterfeit goods and documents, etc. [1,2]. As the spectral resolution can now be in *nm*, hyperspectral sensors can offer great discernment facility in data analysis [2] for numerous humanitarian tasks such as precision agriculture for better farming practices, discrimination among vegetation categories for individual's better treatments and so on [3–5].

1.1 Hyperspectral Remote Sensing Data

In general, hyperspectral data is represented by a hypercube (X^*Y^*F) in which two dimensions (X and Y) denote spatial information and the third dimension (F) denotes the spectral information for the ground covers. Here, F is the spectral information, and X and Y are the spatial information of the ground covers. However, some common preprocessing operations such as geometric correction, radiometric correction and atmospheric correction are accomplished before further analysis of HSIs [1,3,6]. For the further analysis of the HSI hypercube to provide the intended remote sensing applications, the HSI is usually transformed into a data matrix as shown in Figure 1 [2]. In this conversion, a pixel's spectral signature or vector, represented as $\mathbf{x}_n$, in the data matrix or in hypercube is denoted as $\mathbf{x}_n = [x_{n1}\, x_{n2}\, \ldots\ \ldots\, x_{nF}]^T$, where, $n\epsilon$ [1, S], and $S = X^*Y$.

1.2 HSI Classification and Dimensionality Reduction

In HSI context, the classification has the general objective to automatically label the pixels (spectral patterns or signatures) into some predefined classes. The efficient HSI classification is a challenging stage for providing most of the applications using the HSI. The classification can be

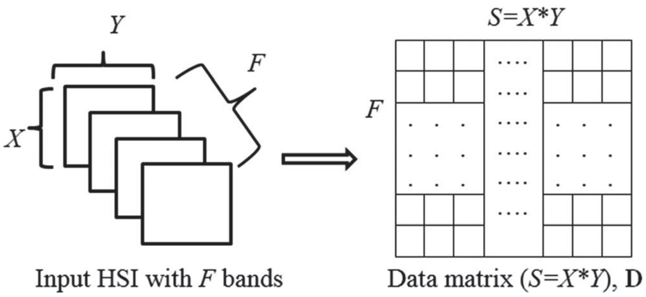


Figure 1: Conversion of a typical HSI hypercube into 2D data matrix [2]. Every band image of the HSI of size $X \times Y$ is converted into a row matrix. Then, the total F row matrices are placed one after another forming an $F \times S$ dimensional data matrix, **D**

performed either using the original features or using the transformed features. Since the geometrical resolution (length) of a pixel may typically range from 1.0 to 80 m, a single pixel is able to represent one or more ground objects as well as several neighboring pixels can also represent a single object [3,7,8]. Consequently, the classification method can be divided into three categories: (i) per-pixel or one pixel or point or hard or crisp classification, where individual pixel is assigned to a single ground class, (ii) sub-pixel or soft or fuzzy classification, where a single pixel can be assigned to more than one earth class and (iii) texture or super resolution or contextual or spatial classification, where neighboring pixels may belong to a single class. These three categories can be performed either in supervised way or in unsupervised fashion. Moreover, the classification method can be either non-parametric or parametric [3,7,8]. For instance, the SCM (spectral correlation mapper), an enhancement on the SAM (spectral angle mapper), is a supervised per-pixel parametric classifier which finds the correlation between testing spectra and ground spectra requiring high computational time [9]. On other hand, the MLC (maximum likelihood classifier) is a supervised sub-pixel or per-pixel parametric classifier [10]. SVM is a supervised per-pixel or sub-pixel non-parametric classifier which is implemented in this experiment for accomplishing the per-pixel classification. The SVM classifies a pixel based on a maximum margin between the testing and training spectra requiring an optimizer and plenty of kernel functions [11].

On the other hand, there are some quiet difficulties for HSI classification for obtaining effective result although it offers beneficiary 2D spatial and 1D spectral information [12]. It needs excessively more computational resources and cost [13]. Additionally, all the bands of the HSI may not encompass same proportion of statistics as well as several bands may have less discriminatory information [14]. Moreover, as it is evidenced in [15–17] that it is not the general practice to use the complete set of bands of the HSI in the classification algorithms. Therefore, feature extraction and feature selection as band reduction techniques are typically employed to address the aforementioned difficulties for the effective HSI classification. The feature extraction extracts the intrinsic features through transformation and feature selection selects the relevant bands from the HSI. The band reduction techniques also fix the Hughes phenomenon or curse of dimensionality problem [18]. Consequently, various feature extraction and feature selection are significantly presented to alleviate these classification challenges for better classification result [2,19–43].

However, as the feature selection uses the original dataset to choose a set of significant bands [38–46], it suffers from high computational cost and local minima problem [47,48]. In this manner, the band selection strategy is acceptable to work in certain circumstances. But this is thus unable to offer the complete advantage of the hyperspectral data. Additionally, this strategy needs a full inspection of the entire band combinations to obtain an optimal subset of the spectral bands. Therefore, it will not be next easily possible to find the best feature selections from the HSI dataset as the amount of band combinations rises exponentially. Consequently, this phenomenon leads to a better alternative approach, called feature extraction, to decrease the dimensionality. The feature extraction tries to reveal the most useful and vital information in the HSI [13,49], which is the topic of this paper. Feature extraction, both unsupervised and supervised, methods embrace non-linear and linear transformations. Supervised methods such as DAFE (discriminant analysis feature extraction) involve some *a priori* knowledge. These approaches could therefore be computationally problematic and the result is mostly reliant on the training set. Instead, unsupervised feature extraction approaches such as PCA are used for extracting effective features without using any *a priori* information about the HSI [50]. However, some attempts are done to develop semi-supervised methods for dimensionality reduction of hyperspectral data [51–55]. To this end, a typical nomenclature of the dimensionality reduction techniques for HSI is illustrated in Figure 2 [44] where this paper aims to provide a general overview and characterization of the PCA-based unsupervised linear and non-linear feature extraction techniques for real HSI classification.

1.3 PCA-based Feature Reduction

PCA is commonly adopted as a linear unsupervised feature extraction method in reducing the dimensionality of

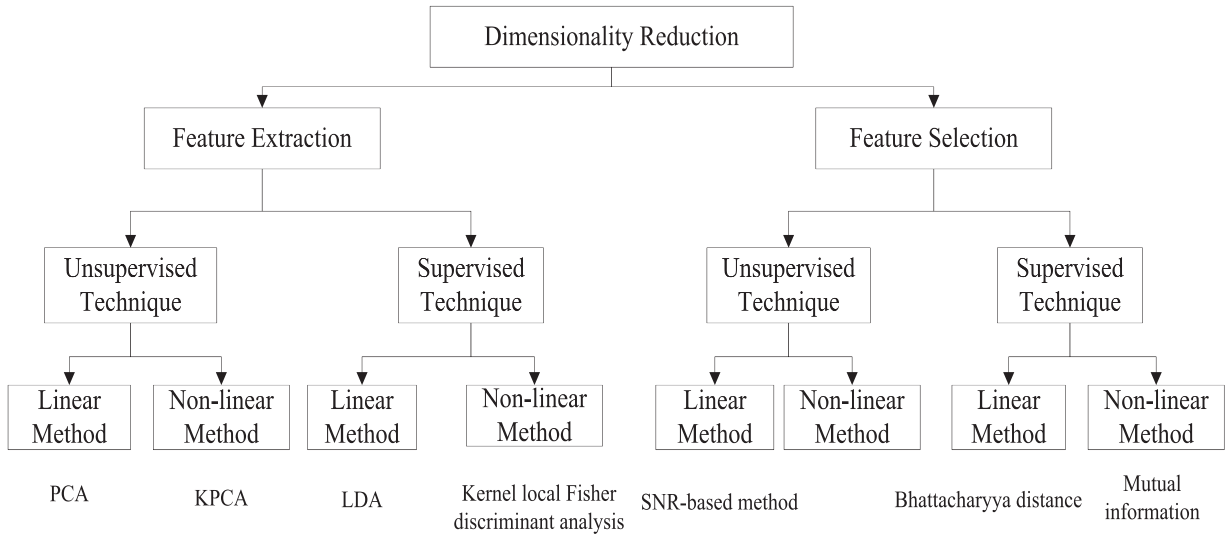


Figure 2: Typical nomenclature of dimensionality reduction techniques for HSI [44]

remote sensing data such as HSI. However, it is argued in [56] that PCA may not be operative for HSI classification. The reasons are as follows: (i) it may be unable to extract subtle information from certain distributions of data as global variance of the HSI is considered, (ii) the highest principal components might not surely contain estimated diverse information and (iii) PCA is controlled visible-light and near-infrared regions of the HSI. However, the conventional PCA could be used as a straightforward background to develop a comparatively more operative feature extraction approach. This newly developed PCA-based method can be able to collect the most useful and subtle information for classification purposes [13]. Consequently, the PCA-based unsupervised linear feature extraction method named SPCA is proved more efficient than the conventional PCA [57]. Another PCA-based unsupervised linear feature extraction method called SSPCA is evidenced superiorly effective than SPCA and conventional PCA especially for plant covers' classification task [13]. Onwards, it has been demonstrated that the PCA-based unsupervised linear feature extraction method named FPCA is efficient over PCA and SPCA [2,58]. Besides these, the PCA-based unsupervised linear feature extraction method MNF can produce superior classification result than PCA when the HSIs inevitably comprise noises due to the sensor error and the influences of other environmental factors [59]. However, the unsupervised nonlinear feature extraction method KPCA can also offer better classification performance fixing nonlinearity in the HSI at a greater computational cost and more input features over PCA and independent component analysis (ICA) [60]. In [61], we studied PCA, SPCA, FPCA, KPCA and KECA in the area of Indian Pine agricultural HSI dataset. In [61], it has been exposed that KPCA and KECA produced better classification accuracy than the studied linear feature extraction methods as they can fix the nonlinearity in the dataset finely. However, the nonlinear methods required the most memory space and comparatively more input features to the SVM classifier. Moreover, it has been observed that the linear method FPCA produced the best accuracy among the linear methods and also very near to that of the nonlinear methods offering the least memory complexity and less input features. In this paper, the PCA-based linear unsupervised methods SPCA, SSPCA, FPCA and MNF together with nonlinear unsupervised KPCA and KECA are critically studied. The investigated methods are then compared and analyzed by various performance measure indexes for Indian Pine HSI and the Washington DC Mall [62] HSI classification using per-pixel SVM classifier. To this end, the general flowchart diagram of the work-done in this paper is illustrated in Figure 3.

The rest of this paper is organized into the following sections. Section II is an explanation and exploration of the present PCA-based methods implemented for HSI. Section III focuses on the experimental arrangement and outcome investigation of the methods for classifying Indian Pine and Washington DC Mall HSIs whereas Section IV recapitulates the interpretations and finalizes the paper.

2. FEATURE EXTRACTION

2.1 PCA for HSI

In hyperspectral remote sensing imagery, PCA identifies the correlation among the bands for extracting the essential characteristics of the HSI. In its implementation [63,64], the zero-mean image $\mathbf{I} = [\mathbf{I}_1 \mathbf{I}_2 \ldots\ldots \mathbf{I}_n]$

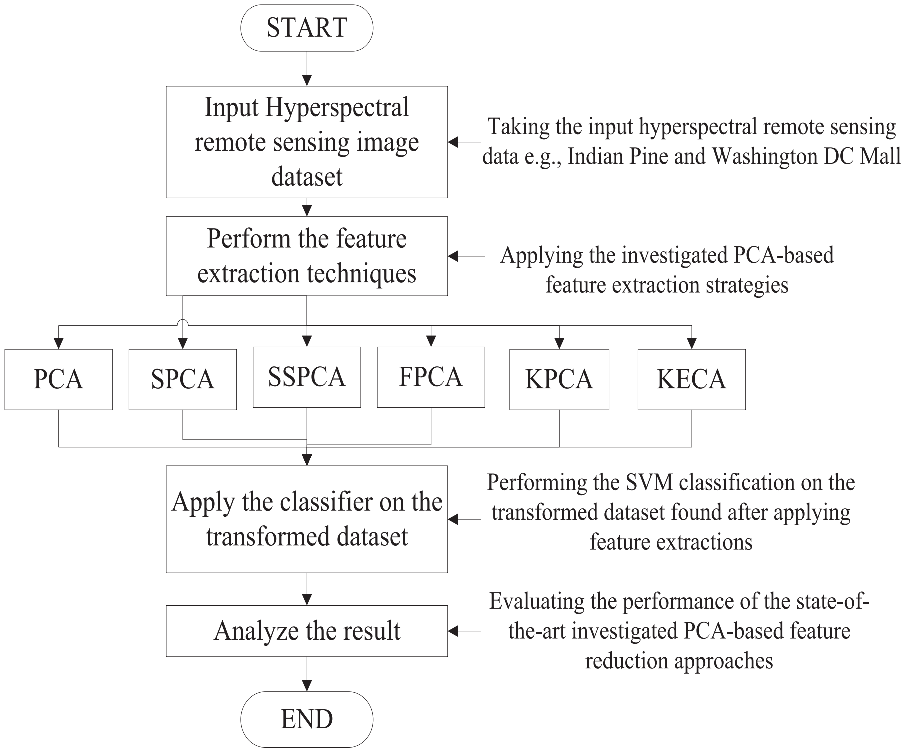


Figure 3: Flowchart of the proposed work

is calculated from the data matrix $\mathbf{D}$. Here, $\mathbf{M} = (1/S)\sum_{n=1}^{S} \boldsymbol{x}_n$ is the mean-image vector and $\boldsymbol{I}_n = \boldsymbol{x}_n - \mathbf{M} = [I_{n1}\ I_{n2}\ \ldots\ \ldots\ I_{nF}]^T$ is the mean-adjusted spectral vector. Now, the covariance matrix $\mathbf{C} = (1/S)\mathbf{I}\mathbf{I}^T$, of the HSI is computed for the following Eigendecomposition operation.

$$\mathbf{C} = \mathbf{V}\mathbf{E}\mathbf{V}^T, \quad (1)$$

where $\mathbf{E}$ and $\mathbf{V}$ denotes the matrix of all respective eigenvalues and eigenvectors, also called principal components (PCs), respectively. Now, q eigenvectors are usually picked out to form an F×q dimensional matrix, $\mathbf{w}$, where $\mathrm{q} \leq \mathrm{F}$ and often $\mathrm{q} \ll \mathrm{F}$. For this purpose, the conventional methodology is to order the eigenvalues from maximum to minimum to pick-up the top q PCs. Other approaches to rank the PCs for further operations include discriminant analysis [44], mutual information-based methods [14,48] divergence analysis, *e.g.* Bhattacharyya distance [49], etc. To end, the projection matrix, $\mathbf{Y}$ can be calculated as $\mathbf{Y} = \mathbf{w}^{T*}\mathbf{I}$.

2.2 SPCA for HSI

Conventional PCA can extract the operative features from the highly correlated original bands of the HSI. The PCA also manipulates a big-size $(F\times F)$ covariance matrix that involves greater computation [57]. As PCA considers global variance of the HSI, it may ignore certain useful subtle information of the dataset. Furthermore, it can be seen that neighboring bands of HSI exploit higher correlations. Therefore, SPCA has been proposed to enhance the application of PCA through omitting highly-correlated blocks' low correlations [57]. In this way, SPCA may be capable to mine the native features of the entire dataset by performing PCA on the highly correlated bands' groups.

In the implementation of SPCA for HSI [57], based on the correlation among the bands the full zero mean HSI data matrix is first divided into L subgroup datasets. Usually, the highly correlated bands are selected as the members of a subgroup for extracting the local characteristics of the entire dataset. Then, conventional PCA is applied on each subgroup dataset which is illustrated in Figure 4 [57]. Mathematically, let $\mathbf{D}_t$ denote the segmented data matrices, where $t\epsilon\ [1, L]$ and n_t be the amount of the successive bands for each segmented data matrix. Then, $\mathbf{D}_1 = [\mathbf{I}'_1\mathbf{I}'_2\ldots\mathbf{I}'_n]$, $\mathbf{D}_2 = [\mathbf{I}''_1\mathbf{I}''_2\ldots\mathbf{I}''_n]$, etc. where, $\mathbf{I}'_j$ contains first n_1 rows of the respective $\mathbf{I}_j$ and $\mathbf{I}''_j$ contains n_2 rows skipping first n_1 rows of the respective $\mathbf{I}_j$ with $j,\ n\ \epsilon\ [1, S]$. After that, Eigendecomposition operation is performed on every computed covariance matrix

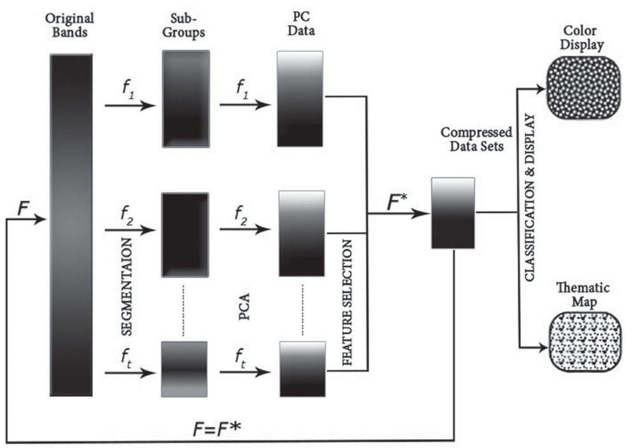


Figure 4: Working procedure of SPCA and SSPCA

of every $\mathbf{D}_t$. At last, the final projection matrix of the whole HSI is found by combining the distinct projection matrices.

2.3 SSPCA for HSI

As stated earlier, SPCA can extract the overall characteristics of each segmented dataset. These global structures supply the local characteristics of the entire dataset for obtaining better classification result or color display. However, to detect the plant targets this strategy may not also be fruitful as it divides the HSI based on the correlations among the bands which may entail to a loss of essential characteristics for the same [13]. On the other hand, it is evidenced that in comparison to the conventional PCA, the piecewise-PCA (PPCA) algorithm presented in [65] can be more efficient for data compression and rebuilding of the plant's spectral databases. The PPCA executes PCA on the groups of the different wavelength regions independently to achieve the enhanced result. Thus, from the impression of classification of SPCA and the spectral region based segmentation of PPCA, the spectrally segmented-PCA (SSPCA) is presented for extracting subtle and most useful information with a view to classifying agricultural HSI more effectively [13]. Like SPCA, SSPCA applies the conventional PCA on the subgroup datasets which are constructed through analyzing the different regions of capturing spectral wavelengths of the HSI. The spectral regions in this segmentation strategy include visible wavelength (VIS: usually 0.4–0.7 μm), near infrared wavelength (NIR: usually 0.7–1.1 μm) and short-wavelength infrared (SWIR: usually 1.1–3.0 μm). The SWIR group can also be broken as SWIR-1 (SWIR-1: usually 1.1–1.35 μm) and SWIR-2 (SWIR-2: usually 1.35–3.0 μm) [13]. The working stages can also be illustrated through Figure 4 in which only the segmentation principles are grounded on different spectral regions of the hyperspectral data.

2.4 FPCA for HSI

All the bands in the entire dataset and in the segmented dataset are treated equal to obtain the covariance matrix by conventional PCA, and SPCA and SSPCA respectively. Thus, these feature extraction methods can be failure to pick up the distinct influences of the featured F spectral bands [2]. Also, these methods produce a big covariance matrix [2]. Consequently, FPCA is presented to mine the native subtle characteristics from the HSI data in which the correlation between bands along with band groups are extracted effectively. In this manner, FPCA reduces the size of the covariance matrix [2]. The FPCA is motivated by the procedure of 2D-PCA [66] which was essentially developed for the facial recognition task.

In the implementation of FPCA for HSI [2], first every mean-adjusted spectral vector, also called spectral signature, $\mathbf{I}_n$ is converted into a 2D matrix of size $H \times W$, where the total F bands are folded into H $(< W)$ groups or segments as shown in Figure 5. Mathematically, let $\mathbf{A}_n$ be the converted matrix with size $H \times \mathrm{W}$, where $F = H{*}W$. Then, $\mathbf{A}_n = [\,\mathbf{a}_{n1}\,\mathbf{a}_{n2\,\ldots}.\mathbf{a}_{nH}]^T$, in which $\mathbf{a}_{nh} = [I_{n(1+W(h-1))} I_{n(2+W(h-1))} \;\cdots\;\; \cdots\; I_{n(W+W(h-1))}]$ and $h\epsilon$ [1, H]. Now, the covariance matrix $\mathbf{C}_n$ for every $\mathbf{A}_n$ is calculated as $\mathbf{C}_n = \mathbf{A}_n^T \mathbf{A}_n$. The complete covariance matrix $\mathbf{C}_{FPCA}$ to perform the Eigendecomposition operation for the entire HSI is computed as follows:

$$\mathbf{C}_{FPCA} = \frac{1}{S} \sum_{n=1}^{S} \mathbf{C}_n = \frac{1}{S} \sum_{n=1}^{S} \mathbf{A}_n^T \mathbf{A}_n \tag{2}$$

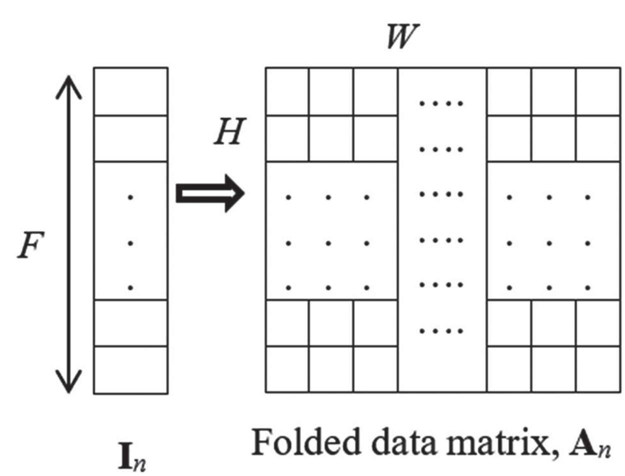


Figure 5: Folding of a spectral vector. W $(> H)$ pixels of each spectral vector is placed row wise consecutively

2.5 MNF for HSI

Although the variance-based PCA is frequently used for feature extraction of remote sensing images, the variances may not essentially reveal the real SNR because of the un-equal noise changes acquired in the spectral bands. Consequently, the band having trivial variance does not imply the poor image feature. Moreover, it can ensure a high SNR in comparison to the other bands with big variances but small SNRs [67]. To cope with this, a PCA-based maximum noise fraction (MNF) transformation is presented in which the transformed features or principal components are ranked on the basis of the maximization of SNR thus not on the variance as in case of PCA [68]. Later, the MNF is reinterpreted using a two-stage procedure: one is the noise-whitening process and another is the PCA to provide what the original MNF does [69]. This newly computed transform is named as Noise-Adjusted Principal Component (NAPC) which in turns is derived as minimum noise fraction (MNF). In this way, the MNF is basically two cascaded PCA transformations that use singular value decomposition (SVD) internally for the support of PCA operation [70]. In the implementation of MNF for remote sensing image such as HSI [12,67–72], first the noise in the input HSI, $\mathbf{X}_1 = \mathbf{D}^T$ is estimated for calculating its covariance matrix, $\mathbf{C}_N$. After that, the $\mathbf{C}_N$ is diagonalized using SVD as follows:

$$\mathbf{E}_{1N} = \mathbf{V}_{1N}{}^T \mathbf{C}_N \mathbf{V}_{1N}, \tag{3}$$

where $\mathbf{E}_{1N}$ is the diagonal matrix of the eigenvalues of $\mathbf{C}_N$ in a decreasing order and $\mathbf{V}_{1N}$ is an orthogonal matrix of the eigenvectors of $\mathbf{C}_N$. The above equation can also be rewritten as follows:

$$\mathbf{I}_N = \mathbf{P}^T \mathbf{C}_N \mathbf{P}, \tag{4}$$

where $\mathbf{I}_N$ is an identity matrix and $\boldsymbol{P} = \mathbf{V}_{1N}\, \mathbf{E}_{1N}{}^{-1/2}$ is the transformation matrix that turns the noise covariance matrix into identity. Now, the original data, $\mathbf{X}_1$ is projected onto a new space as $\mathbf{Y} = \mathbf{X}_1\mathbf{P}$ in which the noise in the data is whitened. In another words, the noise in $\mathbf{Y}$ is thus white with zero-mean and unit variance. In the second step of MNF for HSI, the SVD is applied on the covariance matrix, $\mathbf{C}_Y$ of the noise adjusted data, $\mathbf{Y}$ as follows:

$$\mathbf{E}_{2Y} = \mathbf{V}_{2Y}{}^T \mathbf{C}_Y \mathbf{V}_{2Y}, \tag{5}$$

where $\mathbf{E}_{2Y}$ is the diagonal matrix of the eigenvalues of $\mathbf{C}_Y$ in a decreasing order and $\mathbf{V}_{2Y}$ is an orthogonal matrix of the eigenvectors of $\mathbf{C}_Y$. After this, the MNF transformation matrix is found as $\mathbf{T}_{MNF} = \mathbf{P}\mathbf{V}_{2Y}$. Finally, the projection matrix for the original dataset $\mathbf{X}_1$ can be obtained as follows:

$$\mathbf{Y}_{MNF} = \mathbf{X}_1 \mathbf{T}_{MNF} \tag{6}$$

Also, the SNR-value matrix of all the projected eigenvectors can be calculated as $\mathbf{SNR} = \textit{diagonal}\ (\mathbf{E}_{2Y})\text{-}1$. Usually, the first few components with the highest SNR value are selected for the subsequent operations. Moreover, the original data matrix can be reproduced as, $\tilde{\mathbf{X}}_1 = (\mathbf{T}_{MNF}^T)^{-1}\mathbf{Y}_{MNF}^T$. However, to re-project the projected matrix with first p components into its originary space the complete projected matrix is approximated by keeping the first p bands and by equating zero to the remaining F-p bands.

2.6 KPCA for HSI

Sometimes, due to the lack of scale invariant and complicated structures of the HSI, these linear feature extraction methods cannot assure good class separation [50,73]. Fortunately, KPCA allows generalizing the traditional PCA for band reduction in a nonlinear way. For this, each spectral vector or signature $\mathbf{x}_i$ in the HSI is estimated to a new vector $\emptyset(\boldsymbol{x}_i)$ in a new higher dimensional feature space in which $\emptyset$ is a non-linear function. Then, PCA can be accomplished on the transformed dataset, but this can tremendously be expensive and inefficient. Consequently, kernel strategy can be used to shorten the calculation for KPCA. In the implementation of KPCA for HSI using kernel method, the kernel matrix $\mathbf{K}$ is first built from the $\{\mathbf{x}_i\}$ as $\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$,where, $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function that can be defined as $k(\mathbf{x}_i, \mathbf{x}_j) = \emptyset(\mathbf{x}_i)^T\emptyset(\mathbf{x}_j)$. Then, the Gram matrix $\tilde{\mathbf{K}}$ is calculated as follows.

$$\tilde{\mathbf{K}} = \mathbf{K} - 1_S\mathbf{K} - \mathbf{K}1_S + 1_S\mathbf{K}1_S, \tag{7}$$

where 1_S is of size $S{\times}S$ whose all components are $1/S$. Now, the following calculation is done to prepare the vectors $\mathbf{a}_i$ ($\mathbf{K}$ is replaced by $\tilde{\mathbf{K}}$ if the transformed dataset $\{\emptyset(\mathbf{x}_i)\}$ does not offer zero-mean.).

$$\mathbf{K}\mathbf{a}_k = \boldsymbol{E}_k N \mathbf{a}_k, \tag{8}$$

where $\mathbf{a}_k$ is the S dimensional column vector of a_{ki}, *i.e.* $\mathbf{a}_k = [a_{k1}a_{k2} \ldots\ldots a_{kS}]^T$. Finally, the kernel principal components $\mathbf{y}_k(\mathbf{x})$ are computed using the following equation.

$$\mathbf{y}_k(\mathbf{x}) = \emptyset(\mathbf{x})^T\mathbf{v}_k = \sum_{i=1}^{S} a_{ki}k(\mathbf{x}, \mathbf{x}_i) \tag{9}$$

Here, the power of the kernel methods is that $\emptyset(\mathbf{x}_i)$ should not be calculated explicitly for the KPCA transformation.

The kernel matrix can directly be constructed from the dataset $\{\mathbf{x}_i\}$. The detail discussion on the mostly used kernels such as polynomial, Gaussian kernel and Radial Basis Function (RBF) for KPCA is in [66].

2.7 KECA for HSI

KECA is a succeeding nonlinear unsupervised transformation of KPCA that exposes structure through Renyi Quadratic Entropy (RQE) of the input dataset. Thus, KECA does not depend on the eigenvectors and eigenvalues of $\mathbf{K}$ directly [74,75]. This technique achieves band reduction through projecting onto those KPCA axes that contribute more to the entropy estimation. In this way, KECA can create extremely dissimilar converted data than KPCA.

In its implementation for HSI [74,75], $\mathbf{K}$ from the $\{\mathbf{x}_i\}$ is built. Then, $V(p) = \int p^2(\mathbf{x})d\mathbf{x}$ is in concentration to calculate the RQE since the following logarithm of the original RQE estimation is a monotonic function.

$$H(p) = -\log \int p^2(\mathbf{x})d\mathbf{x}, \tag{10}$$

where $p(\mathbf{x})$ is the *pdf* making the dataset, $\mathbf{D} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ of dimensionality d. Next, the RQE estimator, $\hat{V}(p)$ is calculated which may be used for the Eigendecomposition as $\mathbf{K} = \mathbf{E}\mathbf{D}_d\mathbf{E}^T$, where, $\mathbf{D}_d$stores the eigenvalues $\lambda_1, \ldots .., \lambda_N$ diagonally and $\mathbf{E}$ contains the corresponding eigenvectors $\mathbf{e}_1, \ldots .,\mathbf{e}_N$ as column-wise using the following equation:

$$\hat{V}(p) = \frac{1}{N^2}\sum_{i=1}^{N}\left(\sqrt{\lambda_i}\mathbf{e}_i^T\mathbf{1}\right)^2 \tag{11}$$

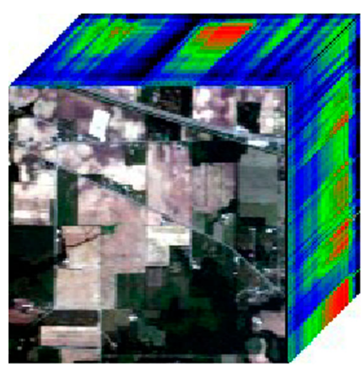
Subsequently, KECA is signified as a k-dimensional transformation found from projecting Φ onto a subspace U_k spanned by those k KPCA axes that contribute most to the RQE estimate of the data. Therefore, a sub-set of KPCA axes constitutes the U_k. Finally, the transformed data, Φ_{eca} for the KECA is manipulated as follows:

$$\Phi_{eca} = P_{U_k}\boldsymbol{\varphi} = \sqrt{\mathbf{D}_k}\mathbf{E}_k^T \tag{12}$$

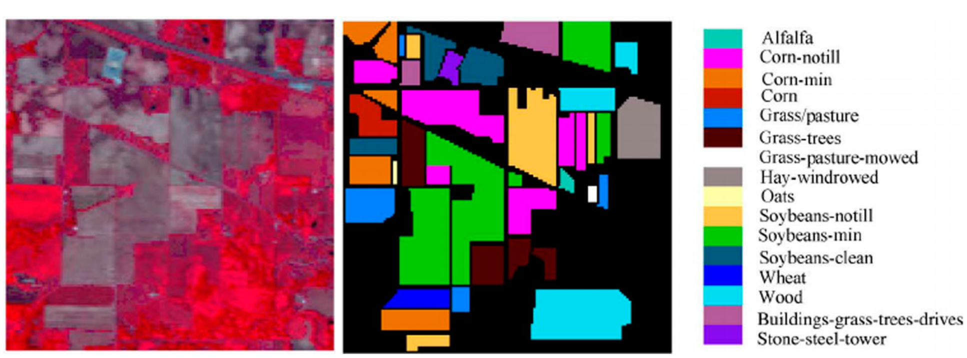
3. EXPERIMENTAL RESULT ANALYSIS

3.1 Dataset Description

The Indian Pine hyperspectral dataset and the Washington DC Mall hyperspectral dataset have been

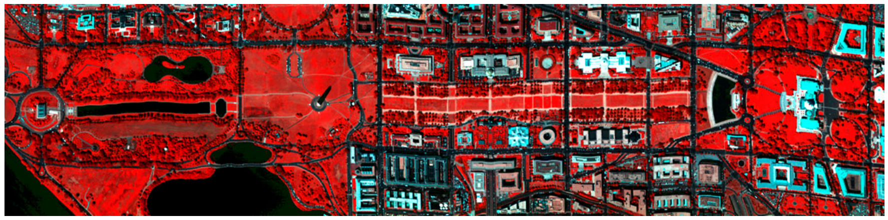


a) Indian Pine HSI hypercube

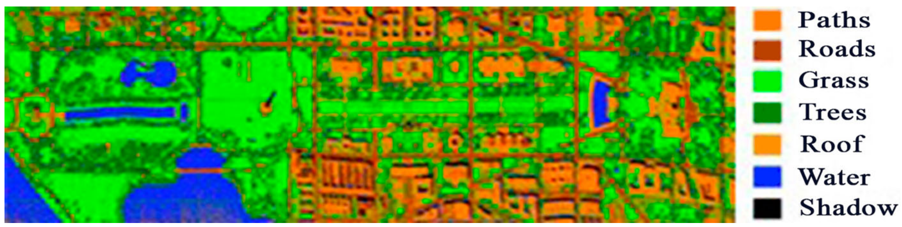


b) Ground truth image

Figure 6: Indian Pine HSI dataset [6]. (a) is the hypercube of the HSI and (b) is a sample false color image with the ground cover classes and thematic mapping



a)False color composite of the original hyperspectral images of the Washington DC Mall dataset



b)Related ground truth

Figure 7: Washington DC dataset [62]. (a) is false color composite of the original hyperspectral images of the Washington DC Mall dataset and (b) is the related ground truth

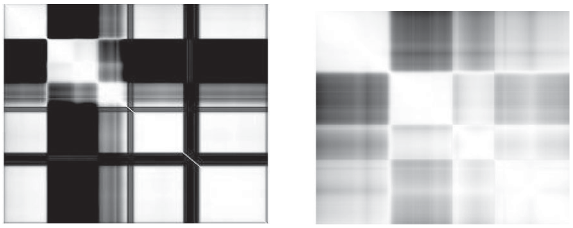
experimented in this paper to analyze the investigated feature extraction techniques and their variants. The agricultural Indian Pine HSI is shown in Figure 6. This HSI is acquired by AVIRIS, a NASA airborne sensor, at the Indian Pine test site in North-western Indiana in June 1992. It has 220 spectral bands (spectral resolution: 10 nm) that cover visible light to near infrared region (0.4–2.5 µm) to capture the agricultural ground objects at the site. The size of each band image is 145×145 having 20 m geometric resolution or GSD (Ground Sampling Distance) with 16 ground cover classes [6]. Also, this Indian Pine dataset is used by many researchers for the classification task in this context [2,16,44,47,76–88]. Then again, the Washington DC dataset is acquired by HYDICE sensor as the airborne hyperspectral data flight-line representation of an urban area [62]. The HSI has originally 210 bands in 0.4–2.4 µm wavelengths region of visible and infrared spectrum. Some bands with water absorption were removed which results in 191 spectral bands in the dataset. Each band image is of size 1280×307 and GSD is 3 m. The dataset contains seven ground classes of the test site. Figure 7 shows a sample false color image with the thematic mapping and ground cover classes of the dataset.

3.2 Experimental Setup and performance measure indexes

The cumulative variance information, peak SNR (PSNR), space requirement and computational cost have been manipulated as the performance measures for the investigated feature extraction approaches. But, the OA (Overall-classification Accuracy) has been mostly considered as the objective measurement to quantitatively assess the approaches. The former parameters are only calculated with Indian Pine dataset and the latter is calculated for both datasets. For quicker processing and well training purposes, 2127 pixels of from the Indian Pine HSI of 14 different classes and 3296 pixels from the Washington DC HSI of 6 different classes have been used for the succeeding tasks.

The upper 12 principal components with largest eigenvalues have been selected accumulating about 100% variance for PCA, SPCA, SSPCA and FPCA to deliberate the variance statistics and PSNR. However, KPCA offers 63.34% cumulative variance using 12 features. Also, the top twelve MNF components with highest SNR values have been used to calculate the PSNR for this method. Moreover, the top twelve KECA components with highest Renyi quadratic entropy have been used for calculating its PSNR.

The band-to-band correlation matrices of the datasets as shown in Figure 8 through image representation are analyzed for finding the number of ways to apply SPCA. In case of Indian Pine dataset, the dataset is divided into 3 different ways to apply SPCA. Firstly, the dataset is divided into 3 different highly correlated subgroup datasets to apply SPCA and this is denoted as SPCA3. Similarly, the dataset is divided into 4 and 5 highly correlated subgroup datasets to perform SPCA independently



a) Correlation matrix of Indian Pine HSI

b) Correlation matrix of Washington DC Mall HSI

Figure 8: Image representation of the band-to-band correlation matrices (white $= 1$ or -1; black $= 0$). (a) Correlation matrix for Indian Pine HSI and (b) Correlation matrix for Washington DC Mall dataset

Table 1: Segmentation of Bands of Indian Pine for SPCA

		Number of segments		
Sub-group	Factor	3 (SPCA3)	4 (SPCA4)	5 (SPCA5)
1	Range of bands	1–36	1–36	1–36
	Diagonal average correlation	0.9365	0.9365	0.9365
	Picked-up PCs	1–6	1–5	1–5
2	Range of bands	37–102	37–79	37–79
	Diagonal average correlation	0.6263	0.9421	0.9421
	Picked-up PCs	1–4	1–3	1–3
3	Range of bands	103–220	80–102	80–102
	Diagonal average correlation	0.7024	0.943	0.943
	Picked-up PCs	1–2	1–2	1–2
4	Range of bands	N/A	103–220	103–162
	Diagonal average correlation		0.7024	0.5196
	Picked-up PCs		1–2	1
5	Range of bands	N/A	N/A	163–220
	Diagonal average correlation			0.941
	Picked-up PCs			1

and these are denoted as SPCA4 and SPCA5 respectively. Table 1 shows the parameters for SPCA3, SPCA4 and SPCA5 of varying band-subgroups.

As the wavelengths of the Indian Pine HSI is 0.4–2.5 μm, SSPCA has been applied through dividing the entire dataset into 3 different ways. The first variant is applied on the two subgroup datasets that contain the bands of the VIS and reflected infrared (RIR) regions respectively. The second variant is applied on the three subgroup datasets which comprise the bands of the VIS, NIR and SWIR regions respectively while the third variant is applied on the four subgroup datasets containing the bands of the VIS, near infrared (NIR), SWIR-1 and SWIR-2 regions respectively. The detail of the variants and the segments are summarized in Table 2. On the other hand, as $F = 220$, five (5) probable different ways with five different folding options are considered

Table 2: Spectrally segmentation of Bands of Indian Pine for SSPCA

		Number of segments		
Sub-group	Factor	2 (SSPCA2)	3 (SSPCA3)	4 (SSPCA4)
1	Name of spectral group	VIS	VIS	VIS
	Range of Wavelength (μm)	0.4–0.7	0.4–0.7	0.4–0.7
	Range of bands	1–32	1–32	1–32
	Diagonal average correlation	0.9582	0.9582	0.9582
	Picked-up PCs	1–6	1–5	1–4
2	Name of spectral group	RIR	NIR	NIR
	Range of Wavelength (μm)	0.7–2.5	0.7–1.1	0.7–1.1
	Range of bands	33–220	33–72	33–72
	Diagonal average correlation	0.2866	0.7182	0.7182
	Picked-up PCs	1–6	1–4	1–4
3	Name of spectral group	N/A	SWIR	SWIR-1
	Range of Wavelength (μm)		1.1–2.5	1.1–1.35
	Range of bands		73–220	73–102
	Diagonal average correlation		0.5472	0.6357
	Picked-up PCs		1–3	1–2
4	Name of spectral group	N/A	N/A	SWIR-2
	Range of Wavelength (μm)			1.35–2.5
	Range of bands			103–220
	Diagonal average correlation			0.7024
	Picked-up PCs			1–2

Table 3: Folding options to apply FPCA on the Indian Pine

Folding Option	*H*	*W*	Representation
1	11	20	FPCA11/20
2	10	22	FPCA10/22
3	5	44	FPCA5/44
4	4	55	FPCA4/55
5	2	110	FPCA2/110

to perform FPCA on the HSI. The folding options for performing the FPCA on the dataset are shown in Table 3.

Moreover, to apply MNF since the noise is not directly given, it is calculated explicitly through subtracting each spectral signature to its next spectral signature [89]. This noise calculation strategy is adopted for both HSIs in hope to achieve better feature extraction. Mathematically, the noise, denoted as $\mathbf{X}1_N$, for the HSI dataset $\mathbf{X}1$ is found as follows:

$$\mathbf{X}1_N(i,:) = X1(i,:) - \mathbf{X}1(i+1,:), \tag{13}$$

where $\mathbf{X}1_N(i,:)$ represents the noise between the current spectral signature $\mathbf{X}1\ (i,:)$ and its next spectral signature $\mathbf{X}1(i+1,:)$. Finally, RBF kernel has been used to apply KPCA and KECA.

Table 4: Segmentation of Bands of Washington DC Mall for SPCA

		Number of segments	
Sub-group	Factor	2 (SPCA2)	3 (SPCA3)
1	Range of bands	1–55	1–55
	Diagonal average correlation	0.9733	0.9733
	No. of bands	55	55
2	Range of bands	56–191	56–133
	Diagonal average correlation	0.8911	0.9511
	No. of bands	136	78
3	Range of bands	N/A	134–191
	Diagonal average correlation		0.9872
	No. of bands		58

As like Indian Pine HSI the mechanism of applying SPCA, SSPCA and FPCA is followed for the Washington DC Mall HSI dataset. The dataset has been divided into 2 different ways to apply SPCA grounded on the correlation matrix as shown in Figure 8(b). The parameters for the 2 different SPCA of varying band-subgroups are illustrated in Table 4. On the other hand, as the wavelength of the Washington DC dataset is 0.4–2.4 µm, SSPCA has been applied through dividing the entire dataset into 3 different ways as illustrated in Table 5. However, as F is 191, the band 16 is removed as a noisy band using MultiSpec to set the number of bands even for evenly applying the FPCA. Thus, two (2) probable different ways with two different folding options are considered to perform FPCA on the HSI. The folding options for performing the FPCA on the dataset are shown in Table 6.

Now, the PCs versus variances (%) of the Indian Pine HSI for PCA, 3 variants of SPCA, 3 variants of SSPCA, 5 variants of FPCA, and KPCA are shown in Figure 9 while the PSNR values using all the feature extraction approaches are illustrated in Figure 10. In case of SPCA and SSPCA, the average of the segments' PSNR has been taken for the assessment. Therefore, from Figure 9 it can be seen that PCA converges very fast to about 100% cumulative variance while from Figure 10 it can also be said that PCA produces better PNSR than the other methods. However, the linear methods SPCA, SSPCA, FPCA and MNF also obtain near PSNR to that of PCA and better PSNR than that of the nonlinear methods.

Nevertheless, the space constraint for the approaches over various phases is equated in Table 7, where it can clearly be seen that the nonlinear methods KPCA and KECA require more memory for data, covariance and projection matrices whereas the linear method FPCA requires the least memory space in all stages. Here, $F = H^{*}W$, q is the number of principal components selected, S is the total number of pixels in the HSI and q' is the number of eigenvalues chosen ($q = Hq$')

Table 5: Spectrally segmentation of Bands of Washington DC for SSPCA

		Number of segments		
Sub-group	Factor	2 (SSPCA2)	3 (SSPCA3)	4 (SSPCA4)
1	Name of spectral group	VIS	VIS	VIS
	Range of Wavelength (µm)	0.4–0.7	0.4–0.7	0.4–0.7
	Range of bands	1–55	1–55	1–55
	No. of bands	55	55	55
	Diagonal average correlation	0.9733	0.9733	0.9733
2	Name of spectral group	RIR	NIR	NIR
	Range of Wavelength (µm)	0.7–2.47	0.7–1.1	0.7–1.1
	Range of bands	56–191	56–91	56–91
	No. of bands	136	36	36
	Diagonal average correlation	0.8911	0.9865	0.9865
3	Name of spectral group	N/A	SWIR	SWIR-1
	Range of Wavelength (µm)		1.1–2.47	1.1–1.35
	Range of bands		92–191	92–102
	No. of bands		100	11
	Diagonal average correlation		0.9338	0.9949
4	Name of spectral group	N/A	N/A	SWIR-2
	Range of Wavelength (µm)			1.35–2.47
	Range of bands			103–191
	No. of bands			89
	Diagonal average correlation			0.9584

Table 6: Folding options to apply FPCA on the Washington DC HSI

Folding Option	*H*	*W*	Representation
1	10	19	FPCA10/19
2	2	95	FPCA2/95

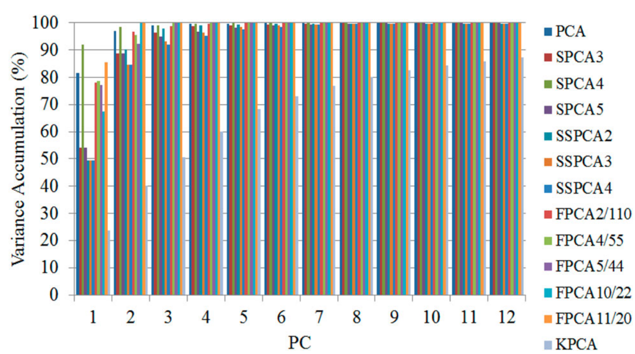


Figure 9: PCs versus cumulative variance (%). PCA converges very fast to 100% cumulative variance and while KPCA converges comparatively very slow to 100% cumulative variance

for FPCA. Moreover, it is already cited that the major steps involved in the feature extraction methods are data and covariance matrices obtainment where the Eigendecomposition operation represents the core of the

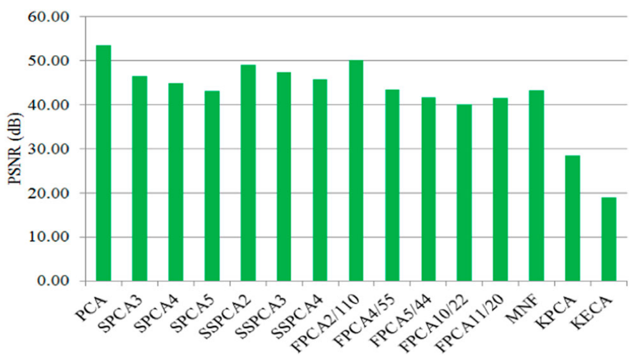


Figure 10: PSNR values using the 12 features extracted from the investigated methods. Linear methods can produce better PSNR than the nonlinear methods

Table 7: Memory requirement comparison

Technique	Stage		
	Size of covariance matrix	Size of data matrix	Size of projection matrix
Conventional PCA	$HW \times HW$	$S \times HW$	$HW \times q$
Segmented-PCA	$W \times W$	$S \times W$	$W \times q/H$
Spectrally SPCA	$W \times W$	$S \times W$	$W \times q/H$
Folded-PCA	$\boldsymbol{W \times W}$	$\boldsymbol{H \times W}$	$\boldsymbol{W \times q/H}$
MNF	$HW \times HW$	$S \times HW$	$HW \times q$
KPCA	$S \times S$	$S \times HW$	$S \times q$
KECA	$S \times S$	$S \times HW$	$S \times q$

PCA-based feature extraction methods. Therefore, the processing time for these operations is also crucial to analyze. Based on the size of the associated matrices for the feature extraction methods the computation time of the methods can tentatively be estimated as in Table 8. Consequently, FPCA requires the least computation time than the other investigated methods whereas the nonlinear methods, KPCA and KECA, need the highest computation time. Moreover, the MNF may require the maximum computation time among the studied linear methods as it performs two cascaded PCA while SPCA and SSPCA also need less time than that of the conventional PCA.

3.3 Classification Accuracy and Evaluation

The OA is manipulated when RBF kernel based SVM is used to assess the performance of the investigated feature extraction approaches. MATLAB LibSVM package is used [90] to implement of SVM. Grid search strategy and 10-fold cross validation scheme and the have been employed to select optimal C and *gamma* [91] of the SVM. The main motives to pick out the SVM are not only that it can exploit the margin-based criteria and is much robust to Hughes phenomenon but also it is generally adopted by various scholars in this context [2,7,11,16,47,76–78,88]. However, in the grid search strategy along 10-fold cross validation scheme, the range for determining C value is 1–10 and its step size is 1. For gamma value determination, the standard range is employed which is 0.1 o 3.0 with step size 0.1.

In this experiment, there are 1121 pixels of 14 different classes as training set and 1006 pixels of those 14 different classes as testing set for the Indian Pine HSI as illustrated in Table 9. In this experiment, more samples are used for training in hope to the efficient training as several classes of the Indian Pine HSI show very similar characteristics. The reasons of having similar characteristics of several classes are illustrated as follows. Since the HSI were collected in the early part of the growing seasons, green soybean and corn canopies covered about 5% only of the ground [44,92] and the residue from the harvest of previous year and soil dominated many samples. Besides,

Table 9: Ratio of training and testing pixels for Indian Pine

Name of the class	No. of training samples	No. of testing samples
Alfalfa	20	15
Corn	21	10
Soybean-clean	15	15
Corn-notill	25	20
Grass-pasture	15	14
Buildings-Grass-Trees-Drives	48	44
Stone-Steel-Towers	48	40
Corn-min	108	72
Soybean-min	130	165
Grass-trees	96	80
Wheat	42	63
Woods	279	248
Soybean-notill	109	85
Hay-windrowed	165	135

Table 8: Computational cost estimation

Technique	Stage		
	Eigendecomposition	Covariance	Data projection
Conventional PCA	$O(H^3W^3)$	$O(SH^2W^2)$	$O(SHWq)$
Segmented-PCA	$O(HW^3)$	$O(SHW^2)$	$O(SWq)$
Spectrally SPCA	$O(HW^3)$	$O(SHW^2)$	$O(SWq)$
Folded-PCA	$\boldsymbol{O(W^3)}$	$\boldsymbol{O(SHW^2)}$	$\boldsymbol{O(SWq)}$
MNF	$O(H^3W^3)$ for the noise	$O(SH^2W^2)$ for the noise	$O(SH^2W^2)$ for the noise
	$O(H^3W^3)$ for the whitened data	$O(SH^2W^2)$ for the whitened data	$O(SHWq)$ for the whitened data
KPCA	$O(S^3)$	$O(S^3)$	$O(S^2q)$
KECA	$O(S^3)$	$O(S^3)$	$O(S^2q)$

Table 10: Ratio of the training and testing pixels for Washington DC

Name of the class	No. of training samples	No. of testing samples
Shadow	20	17
Roof	96	156
Roads	91	128
Trees	262	380
Grass	510	1007
Water	255	374

many scholars in this field have used approximately half of the referenced samples of the Indian Pine hyperspectral data for the adequate training [17,44,88,93,94] because of the aforementioned alike characteristics of classes of the dataset. However, although the reference samples are seemed trivial in number but they are taken from almost all their scattering regions of the site. Also, almost the same number of reference samples is used by various scholars in this context for the faster and convenience learning purpose [44,88].

On the other hand, the training set comprises 1234 pixels (35%) and 2062 pixels (65%) of 6 different classes are used for testing for the Washington DC HSI as illustrated in Table 10. The "Paths" class was not used in this experiment as this class contains inadequate training examples and does not essentially offer the overall representative results [88]. As like Indian Pine HSI, the reference samples are taken from almost all their scattering regions of the test site though the total number of samples is seemed trivial.

However, in context of classification task, the feature extraction approaches commonly undergo for searching the optimal number of PCs or transformed features [60]. To fix this issue in the experiment, the first PC or transformed feature calculated using the biggest eigenvalue or SNR or RQE is used as the input of SVM to perform the classification task and the OA for testing set is noticed. Then, the PCs or transformed features are increased one by one corresponding to the largest eigenvalue or SNR or RQE at each step [60] and similarly the OA for testing

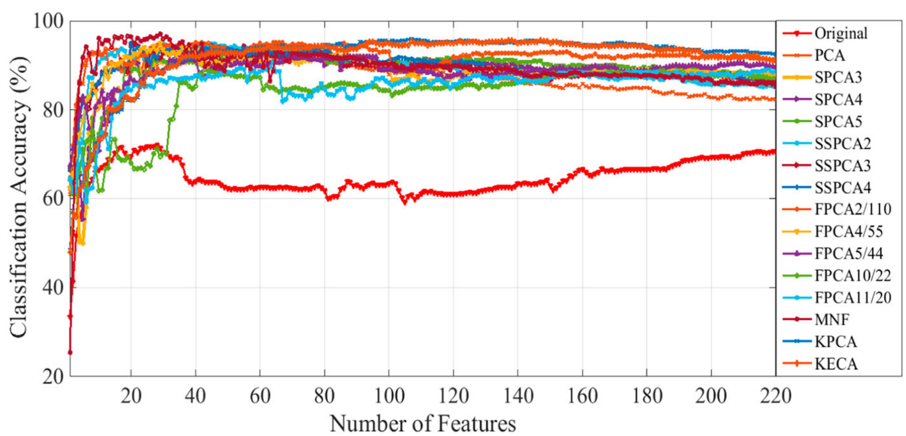


Figure 11: Classification result versus features of the investigated methods. MNF offers the best OA with comparatively less features for the experimented Indian Pine HSI

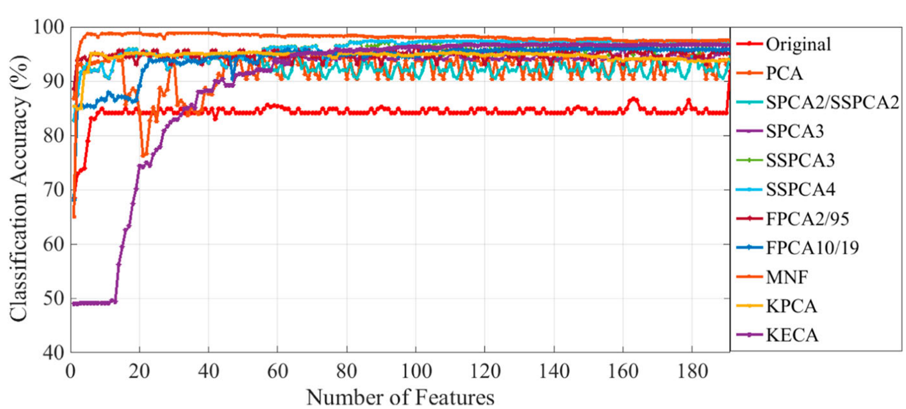


Figure 12: Classification accuracy versus features of the investigated methods. MNF offers the best OA with comparatively less features for the experimented Washington DC HSI

Table 11: Optimal classification result for the Indian Pine dataset

Method	Optimal number of PCs or transformed features	*C*, *gamma*	Overall Accuracy (%)	*Kappa*
Original dataset	All bands	10, 0.4	70.2783	0.6561
PCA	46	6, 0.3	93.7376	0.9274
SPCA3	79	10, 0.1	93.8370	0.9288
SPCA4	50	10, 0.1	93.7376	0.9276
SPCA5	51	10, 0.1	93.9364	0.9299
SSPCA2	42	10, 0.2	95.0298	0.9424
SSPCA3	20	10, 1.0	94.5328	0.9367
SSPCA4	**20**	**10, 0.8**	**95**.1292	**0**.9437
FPCA2/110	**40**	**10, 0.4**	**95**.1292	**0**.9437
FPCA4/55	29	10, 1.4	95.0298	0.9425
FPCA5/44	40	3, 1.1	91.8489	0.9054
FPCA10/22	43	5, 0.7	88.9662	0.8714
FPCA11/20	64	10, 1.3	90.7555	0.8926
MNF	**29**	**10, 0.5**	**97**.0179	**0**.9656
KPCA	**107**	**10, 0.1**	**95**.9245	**0**.9529
KECA	**147**	**10, 0.1**	**95**.6262	**0**.9493

Table 12: Optimal classification result for the Washington DC dataset

Method	Optimal number of PCs or transformed features	*C*, *gamma*	Overall Accuracy (%)	*Kappa*
Original dataset	Allbands	10, 1.2	91.9011	0.9054
PCA	45	10, 0.1	95.0048	0.9412
SPCA2	16	10, 0.2	95.4898	0.9425
SPCA3	58	10, 0.2	96.1688	0.9588
SSPCA2	16	10, 0.2	95.4898	0.9478
SSPCA3	**81**	**10, 0.2**	**97**.2357	**0**.9680
SSPCA4	**93**	**10, 0.2**	**97**.4297	**0**.9694
FPCA2/95	100	10, 0.2	96.3628	0.9598
FPCA10/19	102	10, 0.1	96.0718	0.9577
MNF	**17**	**10, 0.1**	**98**.9331	**0**.9712
KPCA	111	10, 0.1	95.3443	0.9412
KECA	120	3, 0.1	96.7507	0.9602

set is noticed. The graphs in Figures 11 and 12 show the OA for testing set using the feature extraction methods with every feature for the Indian Pine and Washington DC dataset respectively. Also, the optimal number of PCs or transformed features that produces the best OA and respective *Kappa* [95] using various feature extraction methods for the Indian Pine HSI is illustrated in Table 11. Due to the classes of the datasets are sometimes imbalanced as mentioned earlier, *Kappa* is calculated as a measure of how closely the instances classified by the SVM classifier matched the ground truth for each class. Similarly, the optimal number of PCs or transformed features producing the best OA using the feature extraction methods for the Washington DC dataset is shown in Table 12. However, along with the OA and *Kappa* the confusion matrix has been calculated for better illustrating the effectiveness of the feature extraction methods in case of discrepancy among the classes. Tables 13 and 14 show the confusion matrices generated for the testing set of Indian Pine HSI using the entire original and MNF features respectively. In another words, the confusion matrix for the classification model using original entire dataset and the MNF features illustrate the overall performance associated with each of the 14 classes.

From Figure 11 and Table 11, it can be said that any feature extraction technique offers improved result than using the entire original Indian Pine HSI without employing any feature reduction, which is 70.2783% ($C = 10$, $gamma = 0.4$ and $Kappa = 0.6561$) clarifying the need of feature reduction to alleviate the curse of dimensionality problem or Hughes phenomena for efficient classification. The MNF transformation produces the highest classification accuracy among all the investigated feature extraction methods. The reason behind this is that the MNF transformation takes the noise into account to be whitened for better treatment to the effective feature extraction from the experimented agricultural Indian Pine HSI. Also, as the noise has been estimated through the difference between the two consecutive spectral signatures, the subtle local and most useful characteristics have been extracted by MNF. Furthermore, KPCA with RBF kernel in which $\sigma = 0.019$ produces the second highest OA with comparatively more input features. KPCA addresses the nonlinearity (as shown in Figure 13 of scatter plots) in the dataset for obtaining improved classification result. Then again,

Table 13: Confusion matrix for the Indian Pine HSI using the entire original dataset

124	2	0	0	0	0	0	0	0	0	0	9	0	0
0	2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	247	0	17	0	43	0	0	0	0	0	0	7
0	0	0	47	0	0	0	0	0	0	0	0	0	0
0	0	0	16	59	0	0	0	0	0	0	0	0	0
0	39	0	0	0	115	0	0	1	0	0	0	0	0
0	0	0	0	4	0	29	0	0	0	0	0	0	1
0	0	0	0	0	0	0	27	0	0	1	0	0	0
0	0	0	0	0	0	0	0	11	1	1	0	0	0
0	6	0	0	0	2	0	0	0	6	0	0	0	0
0	0	0	0	0	0	0	0	0	0	17	0	0	0
11	0	0	0	0	0	0	0	0	0	0	6	0	0
0	36	0	0	0	48	0	13	32	7	1	0	10	0
0	0	1	0	0	0	0	0	0	0	0	0	0	7

Table 14: Confusion matrix for the Indian Pine HSI using first top MNF features

134	0	0	0	0	0	0	0	0	0	0	3	0	0
0	81	0	0	0	7	0	6	0	0	0	0	0	0
0	0	248	0	1	0	0	0	0	0	0	0	0	1
0	0	0	63	0	0	0	0	0	0	0	0	0	0
0	0	0	0	79	0	0	0	0	0	0	0	0	0
0	4	0	0	0	158	0	1	0	0	0	0	0	0
0	0	0	0	0	0	72	0	0	0	0	0	0	0
0	0	0	0	0	0	0	30	0	0	0	0	0	0
0	0	0	0	0	0	0	0	43	0	1	0	0	0
0	0	0	0	0	0	0	0	0	14	0	0	0	0
0	0	0	0	0	0	0	0	0	0	18	0	0	0
1	0	0	0	0	0	0	0	0	0	0	12	0	0
0	0	0	0	0	0	0	3	1	0	1	0	10	0
0	0	0	0	0	0	0	0	0	0	0	0	0	14

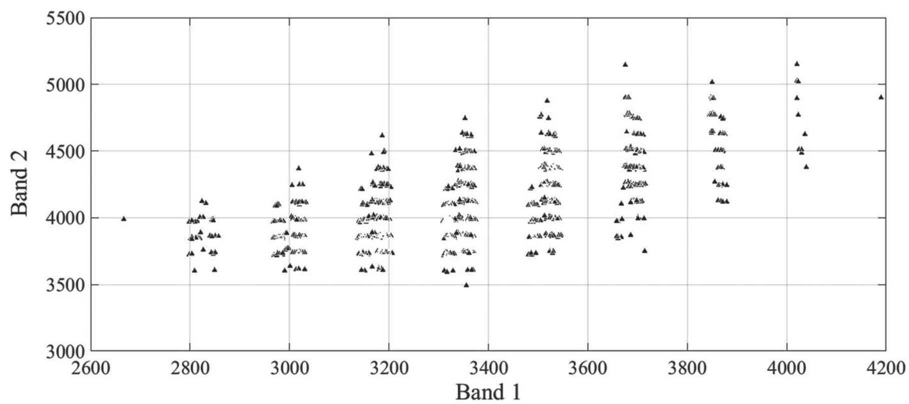


Figure 13: Scatter plot of bands 1 and 2 showing nonlinear relationship

KECA with RBF kernel ($\sigma = 0.011$) attempts to give dealings for the nonlinearity in the dataset to offer better OA. The reason is that although KECA organizes the PCs in terms Renyi-quadratic entropy, they may fail to reveal the distinct contribution of features in comparison to KPCA. In addition, although SSPCA4 and FPCA2/110 are the linear techniques, they have produced identical OA with comparatively fewer features than the nonlinear methods KPCA and KECA. Their classification accuracies are very close to that of KECA, KPCA and MNF. The reason for SSPCA4 is that it has successfully extracted the local subtle and useful structures from the 4 different spectral regions of the agricultural HSI individually. In another words, SSPCA can extract the helpful and local subtle characteristics from each of the four spectral regions (VIS, NIR, SWIR-1, and SWIR-2) independently than that of 2 or 3 regions. The reasons for FPCA2/110 are that several bands of the HSI are linear as shown in Figure 14 and both the global and local characteristics of the dataset are finely extracted from the bands of each of the 2 folds with 110 bands for each. Furthermore, SSPCA based on the spectral regions of the HSI performs

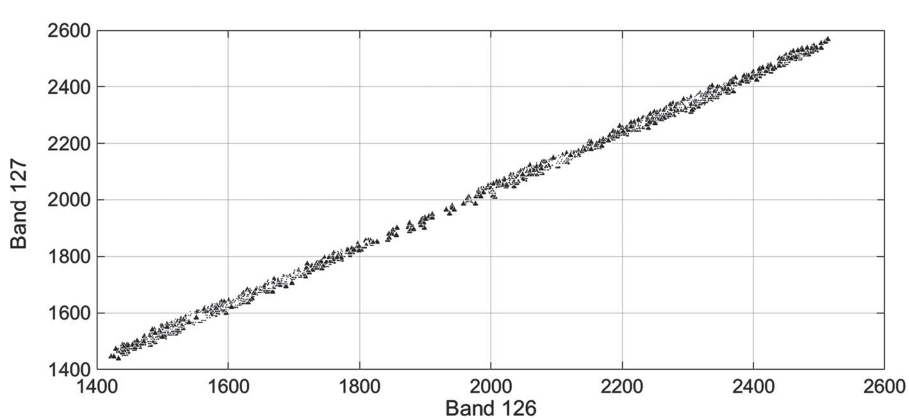


Figure 14: Scatter plot of bands 126 and 127 showing linear relationship

better than the correlation-based SPCA as it is obvious that the subtle intrinsic characteristics of plant covers can be extracted more tremendously by SSPCA from different spectral regions of the dataset separately. To this end, it is observed that FPCA attempts to perform better over SPCA and PCA, and SPCA over PCA as these two features extraction methods also deal with the extraction of local characteristics of the HSI dataset in different ways.

On the other hand, for the Washington DC Mall dataset, it can also be seen from Figure 12 and Table 12 that any feature extraction method offers superior OA than using the entire original dataset without adopting any feature reduction, which is 91.9011% ($C = 10$ and $gamma = 1.2$). The MNF transformation also produces the highest OA among all the studied PCA-based feature extraction methods since the MNF transformation whitens the data for the operative treatment to the effective feature extraction. As the noise in the dataset has been projected through taking the difference between the two consecutive spectral patterns, the local subtle and most useful characteristics have been extracted by MNF. Another linear method SSPCA offers the second highest OA requiring more input features as it is already evidenced that it can tremendously extract the local subtle structures of the dataset from the different spectral regions of the HSI. However, the RQE based KECA outperforms variance based KPCA as KECA can fruitfully keep the distinct featured-contribution than KPCA. The other linear methods also produce satisfactory classification accuracy with comparatively less PCs while the nonlinear methods perform better demanding more input features to the classifier.

From the experiment it can be seen that 12 PCs are enough to represent approximately 100% cumulative variance whereas more PCs are considered for demonstrating classification accuracies. As variance is not always proportional to distinct information for classification [67–69], the PCs with small variance can contribute significantly to produce the better classification accuracies. Thus, more PCs are used to search the optimal (maximum) classification accuracy using the feature extraction methods. However, the classification accuracy as shown in the graph plots of accuracies up to 12 PCs is also satisfactory although it is not the optimal. To this end, it can definitely be seen that MNF produces the highest classification accuracy. The reason is that MNF chooses those transformed features whose SNR is high rather than the features with high variance. This strategy attempts to supply some additional information that might be helpful to the classifier for achieving the maximum classification accuracy. However, if the original dataset is reconstructed using the features having high SNR, it may necessarily not contribute to better reconstruction of the original dataset. In another words, although the additional helpful information of high SNR-based features may not significantly produce more PSNR, it can be very useful for obtaining better classification result.

4. CONCLUSION AND FUTURE RESEARCH

Effective feature extraction plays a significant contribution to the proper classification of the remote sensing data such as HSI. In this paper, the PCA-based feature reduction approaches namely PCA, SPCA, SSPCA, FPCA, MNF, KPCA and KECA are deliberated to extract the intrinsic features from the HSI data. The methods are explained through working stages, accumulation of variance, PSNR, memory and computational cost. Then, the methods are compared on the basis of OA with the mostly used feature pick-up strategy, *i.e.* variance for PCA, SPCA, SSPCA, FPCA and KPCA, SNR for MNF and RQE for KECA. For both experimented agricultural and urban HSIs, the MNF can offer the best classification result with the fewer input features through adjusting the noises calculated using the adopted method by subtracting the consecutive spectral signatures in the datasets. Also, the nonlinear approaches KPCA and KECA obtain improved OA for the Indian Pine dataset but at a greater computation and memory cost, and more input features. For both HSIs, the SSPCA extracts the subtle features from the various spectral regions individually to perform better than the correlation-based SPCA and conventional PCA. The SSPCA produces identical OA to FPCA for the agricultural dataset and second highest OA for the urban dataset. On the other hand, FPCA is the most space and computational-cost saving feature extraction method that offers very close OA to KPCA and KECA for the Indian Pine dataset with $H = 2$ and $W = 110$ as it is able to mine the local structures of the dataset more finely. To this end, the experimental outcome clarifies the necessity of feature reduction for efficient HSI classification where MNF produces the highest classification result but comparatively more computational cost. The FPCA possesses the least memory and time complexity with satisfactory classification performance while the nonlinear methods also perform better requiring more input features, memory and computational time. However, if the set-up, *e.g.* segmentation principles for SPCA, SSPCA and FPCA, the noise calculation strategy for MNF, the kernel trick for KPCA and KECA, etc. are changed, the OA may oscillate. Additional enhancement on OA may be promising through supplying further training pixels. In future, based on the assessment of these PCA-based feature extraction methods, some probable hybrid feature extraction methods can be proposed, *e.g.* combining MNF with SPCA and SSPCA in such a way that will perform conventional MNF instead of PCA in SPCA and SSPCA. Besides, other feature mining approaches such as deep feature extraction methods and spectral-spatial feature extraction methods can be exploited with PCA-based feature extraction methods for assessment.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

ORCID

Md. Palash Uddin http://orcid.org/0000-0002-4429-6590
Md. Al Mamun http://orcid.org/0000-0002-7757-8551

REFERENCES

1. J. A. Richards, and X. Jia. *Remote Sensing Digital Image Analysis*. 4th ed. Berlin: Springer-Verlag, 2006.
2. J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS. J. Photogramm. Remote. Sens.*, Vol. 93, pp. 112–122, 2014.
3. B. K. Mohan, and A. Porwal, "Hyperspectral image processing and analysis," *Current Sci*, Vol. 108, no. 5, pp. 833–841, 2015.
4. M. Teke, H. S. Deveci, O. Haliloğlu, S. Z. Gürbüz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in Proceedings of the 6th IEEE International Conference on Recent Advances in space Tech. (RAST), 2013.
5. D. M. Varade, A. K. Maurya, and O. Dikshit, "Development of spectral indexes in hyperspectral imagery for land cover assessment," *IETE Tech. Rev.*, Vol. 36, no. 5, pp. 475–483, 2019.
6. M. F. Baumgartner, L. Labial, and D. A. Landgrebe. "220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3," *Purdue University Research Repository*, 2015.
7. M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, "A review of remote sensing image classification techniques: the role of spatio-contextual information," *Eur. J. Remote Sens.*, Vol. 47, pp. 389–411, 2014.
8. F. van der Meer, "Analysis of spectral absorption features in hyperspectral imagery," *Int. J. Appl. Earth Observ. Geoinf.*, Vol. 5, no. 1, pp. 55–68, Jan. 2004.
9. O. A. de Carvalho, Jr, and P. R. Meneses, "Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM)," in AVIRIS Workshop Proceedings, 2000.
10. J. D. Paola, and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. on Geosci. and Remote Sens*, Vol. 33, no. 4, pp. 981–996, 1995.
11. J. A. Gualtieri, and S. Chettri, "Support vector machines for classification of hyperspectral data," in Proceedings of the IEEE Geoscience and remote Sensing Symposium, 2000.
12. Y. Xu, P. Yu, B. Guo, X. Gao, and Y. Guo, "On hyperspectral remotely sensed image classification based on MNF and AdaBoosting," in Proceedings of the IEEE 2012 International Conference on Audio, Language and image process. (ICALIP), China, 2012, pp. 605–609.
13. F. Tsai, E.-K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *Int. J. of Remote Sens.*, Vol. 28, no. 5, pp. 1023–1039, 2007.
14. B. Guo, S. R. Gunn, R. I. Damperand, and J. D. B. Nelson, "Band selection for hyper spectral image classification using mutual information," *IEEE Geosci. and Remote Sens. Lett*, Vol. 3, no. 4, pp. 522–526, 2006.
15. J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: from basics to recent developments," in *Multiple Classifier Systems*, vol. 4472, M. Haindl, J. Kittler and F. Roli, Eds. Berlin: Springer-Verlag, 2007, pp. 501–512, ser. Lecture Notes in Computer Science.
16. B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, Vol. 48, no. 7, pp. 2880–2889, Jul. 2010.
17. J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens*, Vol. 48, no. 7, pp. 2840–2853, Jul. 2010.
18. G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theor.*, Vol. 14, no. 1, pp. 55–63, 1968.
19. J. C. Harsanyi, and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, Vol. 32, no. 4, pp. 779–785, Jul. 1994.
20. W. Xiong, C. I. Chang, C. C. Wu, K. Kalpakis, and H.-M. Chen, "Fast algorithms to implement N-FINDR for hyperspectral endmember extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 4, no. 3, pp. 545–564, Sept. 2011.

21. L. Chang, Y. L. Chang, Z. S. Tang, and B. Huang, "Group and region based parallel compression method using signal subspace projection and band clustering for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 4, no. 3, pp. 565–578, Sept. 2011.

22. A. Dowler, and M. Andrews, "On the convergence of N-FINDR and related algorithms: to iterate or not to iterate?" *IEEE Geosci. Remote Sens. Lett*, Vol. 8, no. 1, pp. 4–8, Jan. 2011.

23. I. Dópido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 5, no. 2, pp. 421–435, Apr. 2012.

24. H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 5, no. 2, pp. 544–554, Apr. 2012.

25. J. Wen, Z. Tian, X. Liu, and W. Lin, "Neighborhood preserving orthogonal PNMF feature extraction for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 6, no. 2, pp. 759–768, Apr. 2013.

26. W. Li, and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 7, no. 4, pp. 1012–1022, Apr. 2014.

27. X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyper-spectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens*, Vol. 52, no. 6, pp. 3742–3752, Jun. 2014.

28. X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image de-composition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens*, Vol. 53, no. 4, pp. 2241–2253, Apr. 2015.

29. G. Y. Chen, and S. E. Qian, "Dimensionality reduction of hyperspectral imagery using improved locally linear embedding," *J. Appl. Remote Sens.*, Vol. 1, pp. 1–10, 2007.

30. M. Zaouali, S. Bouzidi, and E. Zagrouba, "3-D shearlet transform based feature extraction for improved joint sparse representation HSI classification," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens*, Vol. 11, no. 4, pp. 1306–1314, 2018.

31. Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 7, no. 6, pp. 2094–2107, 2014.

32. Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, no. 10, pp. 6232–6251, 2016.

33. W. Zhao, and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, no. 8, pp. 4544–4554, 2016.

34. M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Improved feature extraction using segmented FPCA for hyperspectral image classification," in IEEE 2nd international Conference on Electrical & Electronic Engineering (ICEEE), Bangladesh, Dec. 27–29, 2017.

35. M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Segmented FPCA for hyperspectral image classification," in IEEE 3rd International Conference on Electrical Information and Communication Technology (EICT), Bangladesh, Dec. 7–9, 2017.

36. M. R. Islam, M. A. Hossain, and B. Ahmed, "Improved subspace detection based on minimum noise fraction and mutual information for hyperspectral image classification," in International Joint Conference on computational intelligence, Bangladesh, December 2018.

37. M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 40, no. 18, pp. 7190–7220, 2019.

38. M. R. Islam, B. Ahmed, and M. A. Hossain, "Feature reduction based on segmented principal component analysis for hyperspectral images classification," in IEEE international Conference on Electrical, Computer and Communication Engineering (ECCE), Bangladesh, Feb. 2019.

39. N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 57, no. 2, pp. 755–769, 2019.

40. Z. Chen, J. Jiang, C. Zhou, S. Fu, and Z. Cai, "SuperBF: superpixel-based bilateral filtering algorithm and its application in feature extraction of hyperspectral images," *IEEE. Access.*, Vol. 7, pp. 147796–147807, 2019.

41. M. Imani, and H. Ghassemian, "Morphology-based structure-preserving projection for spectral–spatial feature extraction and classification of hyperspectral data," *IET Image Proc.*, Vol. 13, no. 2, pp. 270–279, 2019.

42. W. Tian, L. Xu, Z. Chen and A. Shi, "Multiple feature learning based on edge-preserving features for hyperspectral image classification," *IEEE. Access.*, vol. 7, pp. 106861–106872, 2019.

43. H. Li, H. Li, and L. Zhang, "Quaternion-based multiscale analysis for feature extraction of hyperspectral images," *IEEE Trans. Signal Process*, Vol. 67, no. 6, pp. 1418–1430, 2019.

44. X. Jia, B.-C. Kua, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. of the IEEE*, Vol. 101, no. 3, pp. 676–697, 2013.

45. C. Chang, D. Qian, T. L. Sun, and M. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 37, no. 6, pp. 2631–2641, 1999.

46. A. Sellami, M. Farah, I. R. Farah, and B. Solaiman, "Hyperspectral imagery semantic interpretation based on adaptive constrained band selection and knowledge extraction techniques," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 11, no. 04, pp. 1337–1347, Apr. 2018.

47. F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, Vol. 42, no. 8, pp. 1778–1790, 2004.

48. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relecvance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, no. 8, pp. 1226–1238, 2005.

49. X. Kang, X. Xiang, S. Li, and J. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 55, no. 12, pp. 7140–7151, Dec. 2017.

50. M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectral image analysis," in Proceedings of the IEEE 7th Nordic Signal Processing Symposium (NORSIG), Iceland, 2006, pp. 238–241.

51. L. Chen, R. Huang, and W. Huang, "Graph-based semi-supervised Weighted band selection for classification of hyperspectral data," in Proceedings of the IEEE ICALIP, 2010.

52. C. Yang, S. Liu, L. Bruzzone, R. Guan, and P. Du, "A semi-supervised feature Metric based band selection method for hyperspectral image classification," in Proceedings of the IEEE 4th Workshop on hyperspectral image and signal processing (WHISPERS), 2012.

53. Z. Guo, H. Yang, X. Bai, Z. Zhang, and J. Zhou, "Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models," in Proceedings of the IEEE IGARSS, 2013.

54. C. Chang, K. Chen, B. Kuo, M. Wang, and C. Li, "Semi-supervised local discriminant analysis with nearest neighbors for hyperspectral image classification," in Proceedings of the IEEE IGARSS, 2014.

55. X. Chen, L. Song, Y. Hou, and G. Shao, "Efficient semi-supervised feature selection for VHR remote sensing images," in Proceedings of the IEEE IGARSS, 2016.

56. A. Cheriyadat, and L. M. Bruce, "Why principal Component analysis is not an Appropriate feature extraction method for hyperspectral data," in Proceedings of the IEEE Geosci. and remote Sens. Symp., 2003, pp. 3420–3422.

57. X. Jia, and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 37, no. 1, pp. 538–542, 1999.

58. P. Deepa, and K. Thilagavathi, "Data reduction techniques of hyperspectral images: a comparative study," in Proceedings of the IEEE 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), 2015.

59. L. Xiang, Z. Bing, G. LianRu, and C. DongMei, "A maximum noise fraction transform with improved noise estimation for hyperspectral images," *Springer Sci. China Series F-Information Sci.*, Vol. 52, no. 9, pp. 1578–1587, 2009.

60. L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, and Q. M. Gu, "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine," *Neurocomputing*, Vol. 55, no. 1–2, pp. 321–336, 2003.

61. M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Feature extraction for hyperspectral image classification," in Proceedings of the IEEE 5th region 10 humanitarian Technology Conference (R10HTC), 2017.

62. D. A. Landgrebe. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html.

63. C. Rodarmel, and J. Shan, "Principal component analysis for hyper-spectral image classification," *ACM Surv. and Land Info. Sci*, Vol. 62, no. 2, pp. 115–122, 2002.

64. A. Shabna, and R. Ganesan, "HSEG and PCA for Hyperspectral image classification," in Proceedings of the IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014.

65. I. E. Bell, and G. V. G. Baranoski, "Reducing the dimensionality of plant spectral databases," *IEEE Trans. Geosci. Remote Sens.*, Vol. 42, no. 3, pp. 570–576, Mar. 2004.

66. J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Analy. Machine Intell.*, Vol. 26, no. 1, pp. 131–137, Jan. 2004.

67. C. Chang, and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.*, Vol. 37, no. 5, pp. 2387–2396, 1999.

68. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of

image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, Vol. 26, pp. 65–74, 1988.

69. J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sens.*, Vol. 28, pp. 295–304, 1990.

70. J. W. Boardman, and F. A. Kruse, "A geological example using AVIRIS data, Northern Grapevine mountains, Nevada," in Proceedings of the 10th thematic Conference Geologic remote Sensing, vol. I, Michigan, 1994, pp. 407–418.

71. R. Roger, "A faster way to compute the noise-adjusted principal components transform matrix," *IEEE Trans. Geosci. Remote Sens.*, Vol. 32, no. 6, pp. 1194–1196, 1994.

72. B. Xu and P. Gong, "Noise estimation in a noise-adjusted principal component transformation and hyperspectral image restoration," in Proceedings of the IEEE Geoscience and remote Sensing Symposium, 2003, pp. 271–286.

73. Q. Wang. "Kernel Principal Component Analysis and its Application in Face Recognition and Active Shape Models," *ARXIV1207*.3538, 2012.

74. R. Jenssen, "Kernel entropy Component analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 32, no. 5, pp. 847–860, 2010.

75. L. Gómez-Chova, R. Jenssen, and G. Camps-Valls, "Kernel entropy component analysis for remote sensing image clustering," *IEEE Geosci. Remote Sens. Lett.*, Vol. 9, no. 2, pp. 312–316, 2012.

76. M. Pal, and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, Vol. 48, no. 5, pp. 2297–2307, May 2010.

77. H. Huang, J. Liu, and Y. Pan, "Semi-supervised marginal fisher analysis for hyperspectral image classification," in ISPRS Annals of the Photogrammetry, remote Sens. and spatial Infor. Sci., vol. I-3, XXII ISPRS Congress, 2012.

78. M. Rojas, I. Dópido, A. Plaza, and P. Gamba, "Comparison of support vector machine-based processing chains for hyperspectral image classification," in Proceedings of the SPIE, vol. 7810, 2010.

79. G. Camps-Valls, and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

80. A. Datta, S. Ghosh, and A. Ghosh, "Supervised feature extraction of hyperspectral images using partitioned maximum margin criterion," *IEEE Geosci. Remote Sens. Lett.*, Vol. 14, no. 1, pp. 82–86, Jan. 2017.

81. Z. Hao, M. Berman, Y. Guo, G. Stone, and I. Johnstone, "Semi-realistic simulations of natural hyperspectral scenes," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 9, no. 9, pp. 4407–4419, Sept. 2016.

82. M. J. Marin-McGee, and M. Velez-Reyes, "A spectrally weighted structure tensor for hyperspectral imagery," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 9, no. 9, pp. 4442–4449, Sept. 2016.

83. P. Ghamisi, A.-Rahman Ali, M. S. Couceiro, and J. A. Benediktsson, "A novel evolutionary swarm fuzzy clustering approach for hyperspectral imagery," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 8, no. 6, pp. 2447–2456, Jun. 2015.

84. P. Ghamisi, and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, Vol. 12, no. 2, pp. 309–313, Feb. 2015.

85. A. Soltani-Farani, and H. R. Rabiee, "When pixels team up: spatially weighted sparse coding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, Vol. 12, no. 1, pp. 107–111, Jan. 2015.

86. I. Dópido, J. Li, P. Gamba, and A. Plaza, "A new hybrid strategy combining semisupervised classification and unmixing of hyperspectral data," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 7, no. 8, pp. 3619–3629, Aug. 2014.

87. A. Yang, H. Jin, M. Wang, Y. Ren, and L. Jiao, "Data-driven compressive sampling and learning sparse coding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, Vol. 11, no. 2, pp. 479–483, Feb. 2014.

88. M. A. Hossain, X. Jia, and M. Pickering, "Subspace detection using a mutual information measure for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, Vol. 11, no. 2, pp. 424–428, Feb. 2014.

89. D. Hundley, M. Kirby, and M. Anderle, "A solution procedure for blind signal separation using the maximum noise fraction approach: algorithms and examples," in Proceedings of the 3rd International Conference on independent Comp. Anal. and Blind Sig. Sep., California, 2001, pp. 337–342.

90. C. C. Chang, and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Sys. Techn.*, Vol. 2, no. 3, pp. 27:1–27:27, 2011.

91. Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and testing low-degree polynomial data mappings via linear SVM," *J. Mach. Learn. Res.*, Vol. 11, pp. 1471–1490, 2010.

92. D. A. Landgrebe, "Information Extraction Principles and Methods for Multispectral and Hyperspectral Image

Data," C. H. Chen, Ed. Singapore: World Scientific; 1999.

93. M. A. Hossain, X. Jia, and J. A. Benediktsson, "One-class oriented feature selection and classification of heterogeneous remote sensing images," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 9, no. 4, pp. 1606–1612, Apr. 2016.

94. Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, Vol. 8, no. 6, pp. 2381–2392, June 2015.

95. J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, Vol. 20, no. 1, pp. 37–46, 1960.

Authors

Md. Palash Uddin (palash_cse@hstu.ac.bd), a member of IEEE, is presently serving as an Assistant Professor in department of Computer Science and Engineering (CSE) in HSTU, Dinajpur, Bangladesh. Previously, he was a lecturer in the same university and in Central Women's University, Dhaka, Bangladesh. He has completed his M. Sc. degree from department of CSE, RUET, Rajshahi, Bangladesh in 2018 and B. Sc. degree in CSE from HSTU, Dinajpur, Bangladesh in 2011. He is currently pursuing his PhD degree in Deakin University, Geelong VIC 3220, Australia. His focal research interest is based on the remote sensing image analysis, artificial intelligence-based application development and machine learning algorithms for data mining. He has several national and international peer-reviewed journal and conference publications in various fields of Computer Science and Technology.

Corresponding author. Email: palash_cse@hstu.ac.bd

Md. Al Mamun, is working as a Professor in the Department of Computer Science and Engineering, RUET, Bangladesh. His teaching interests are inseparably conjoined with his research ambitions. His research objective was to gain a deep and clear understanding and to establish a successful career record as a researcher in image/video analysis, data mining, and computer vision. He completed BSc in Computer Science and Engineering in 2005 and PhD in Computer Science in 2011. In terms of professional experiences, he worked in reputed national and multinational organizations with outstanding track records. In every position, he showed outstanding analytical and problem solving skills, abilities to work as a part of team with minimal supervision. His area of research interest includes Satellite Image Mining: Image Compression, Change Detection, Prediction and Forecasting, Adaptive Linear and Non-linear Modeling, Remote Sensed Image Interpretation and Symbolic Representation of Image Contents, Visualization and Model Generation for Handling Complex Data Set and Computer Vision: Pattern Recognition and Image Classification, Objects Recognition and Feature Extraction and Nonlinear Image Classification. He is serving as a reviewer for various conferences and journal such as IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing etc.

Md. Ali Hossain, presently serving as an Associate Professor in the Department of Computer Science and Engineering, RUET. He completed his Doctor of Philosophy (PhD) from the University of New South wales (UNSW) Canberra with **most outstanding research contribution** in the area of Computer Science. In tenure at the UNSW, he worked as a research fellow in the School of Engineering and Information Technology from 2014 to 2015. Details research contribution can be reached through https://sites.google.com/site/alicse03/home.