Asian Journal of Information Technology 18 (2): 57-66, 2019
ISSN: 1682-3915

Dimensionality Reduction of Remotely Sensed Hyperspectral Image for Classification using PCA with Autoencoder Technique

B.R. Shivakumar and J. Prakash
[1]Department of Information Science and Engineering, Bangalore Institute of Technology, VTU, K.R. Road, V.V. Puram, Bengaluru, Belagavi, India
[2]Department of Information Science and Engineering, Bangalore Institute of Technology, K.R. Road, V.V. Puram, Bengaluru, India
shivakumar_br@rediffmail.com

Abstract: Hyperspectral Imagery (HSI) is widely used in the application domains such as agriculture, environment, forestry and geology for the identification and observations which demands the efficient classification accuracy. The supervised classification is a challenging task due to limited number of available training samples compared to large number of spectral bands. This phenomena reduces the classification accuracy. To overcome this problem, the dimensionality reduction preprocessing step is adopted. This process reduces the number of spectral bands which leads to decrease in computational complexity and enhancement in classification accuracy. In this study, AEPCA (Auto Encoder and Principle Component Analysis) method is proposed for dimensionality reduction of HSI. The performance of AEPCA is evaluated against AE (Autoencoder) and PCA (Principle Component Analysis) method. The dimensionally reduced components are classified using CNN (Convolutional Neural Network) based classifier. The proposed model of dimensionality reduction demonstrates superior classification accuracy due to effective combination of characteristics of AE and PCA. The noisy or corrupted pixels are recovered by AE Model and high dimensional image is represented by efficient fewer number of principle components by PCA is the potential advantage of AEPCA Model.

Key words: Auto encoder and principle component analysis, classification, convolutional neural network, deep neural network, dimensionality reduction, hyperspectral image

INTRODUCTION

The rich spectral information available in remotely sensed HSI allows for the possibility to distinguish between spectrally similar materials. However, supervised classification of hyperspectral images is a very challenging task due to the generally, unfavourable ratio between the large number of spectral bands and the limited number of training samples available a priori which results in the Hughes phenomenon (Hughes, 1968). When the number of features considered for classification is larger than a threshold, the classification accuracy starts to decrease. On the contrary, the collection of reliable training samples is very expensive. To address this issue, a dimensionality reduction step is often performed prior to the classification process which allows separating the classes by discarding information that is ineffective for classification purposes (Zhou *et al*., 2015).

Dimensionality reduction methods can be divided into two categories (Sellami and Farah, 2018) feature selection and feature extraction. Feature selection selects the best band combinations whereas feature extraction preserves most important spectral features through mathematical transformations. Feature extraction includes linear methods and nonlinear manifold learning methods. The linear methods such as PCA (Jolliffe, 2002) and MNF (Zhao *et al*., 2016; Amato *et al*., 2009) perform well in HSI data and have low computational complexity. Nonlinear manifold learning methods extract features by reconstructing the underlying manifold from which the HSI data was sampled and require high computational complexity. Manifold learning methods are more suitable for HSI data because of the nonlinear structure of HSI data that originates from multi-scattering and the heterogeneity of pixels (Sun *et al*., 2014).

Ongoing with the data acquisition process, the covariance matrix of HSI hypercube can be computed in real time. This offers great potential for HSI embedded devices to provide not only conventional HSI data but also pre-processed information (Zabalza *et al*., 2015).

Corresponding Author: B.R. Shivakumar, Department of Information Science and Engineering, Bangalore Institute of Technology, VTU, K.R. Road, V.V. Puram, Bengaluru, Belagavi, India, shivakumar_br@rediffmail.com

Regarding the unsupervised band selection domain, some approaches were proposed. Cluster-based methods use clustering techniques in which the first step is to compute a distance measure for each pair of bands. With these metrics, the bands are grouped in disjoint clusters such that bands in a given cluster tend to be similar to each other, according to these metrics and bands in different clusters tend to be dissimilar. After the grouping a representative band from each cluster is chosen. Search-based methods aim at finding a good set of band by evaluating sub-sets features. Using exhaustive search strategies to find the best sub-set is normally unmanageable for this kind of data, however, several sub-optimal search strategies like sequential backward selection or evolutionary techniques are used in this kind of problem. Rank-based methods use metrics such as entropy, mutual information and correlation criterion to sort the bands by their importance in the selection (Lu and Weng, 2007).

Spectral Rhythm (SR) representation is a suitable way to compute bands dissimilarities the proposed SR-based band selection strategy is as effective as state-of-the-art hyperspectral dimensionality reduction techniques the proposed random pixel sampling is able to speed up the band selection process keeping high accuracy rates on classification tasks (Santos *et al*., 2015). Unlike other nonlinear dimensionality reduction methods the invertibility, volume-preservationand straightforward out-of-sample extension, makes Dimensionality Reduction via. Regression (DRR) interpretable and easy to apply.

The properties of DRR enable learning a broader class of data manifolds than the recently proposed Non-Linear Principal Components Analysis (NLPCA) and Principal Polynomial Analysis (PPA) (Laparra *et al*., 2015). At present, deep learning is one of the foremost technique in machine learning and computer vision. This technique works based on deep neural networks which has larger capacity of deep feature learning from the samples. Hinton and Salakhutdino (2006) proposed dimensionality reduction of the data with deep neural network (Hinton and Salakhutdinov, 2006). The high dimensional data is transformed into low dimensional data using multilayer neural network. This is achieved by latent space layer which is smaller than the input and output layer to reconstruct the high dimensional input vectors. The Deep Belief Network (DBN) is experimented for HSI dimensionality reduction. The depth of the network plays an important role in the effectiveness of the dimension reduction results (Arsa *et al*., 2016). In this study, a new method is proposed to reduce dimensions of HSI and spectral spatial classification (Jia *et al*., 2016; Zhao *et al*., 2016). The proposed framework is constructed based on AE for dimensionality reduction and Convolutional Neural Network (CNN) for classification. Here, two different HSI data setsare used to evaluate performance of the proposed model. The overall accuracy at different number of components, the loss functions are calculated.

MATERIALS AND METHODS

This study mainly introduces to the architecture and algorithm of the proposed methodology for dimensionality reduction and classification of HSI. The architecture which contains the autoencoder and PCA Models for dimensionality reduction and CNN Model for classification purpose as shown in Fig. 1. In first phase, the HSI with L spectral bands are reduced to the desired number of components PCA or AE or AEPCA algorithms. In second phase, the spectral spatial features are extracted from the resultant components and fed to CNN Model. The classification model is trained with the input image feature to produce the predicted classified map.

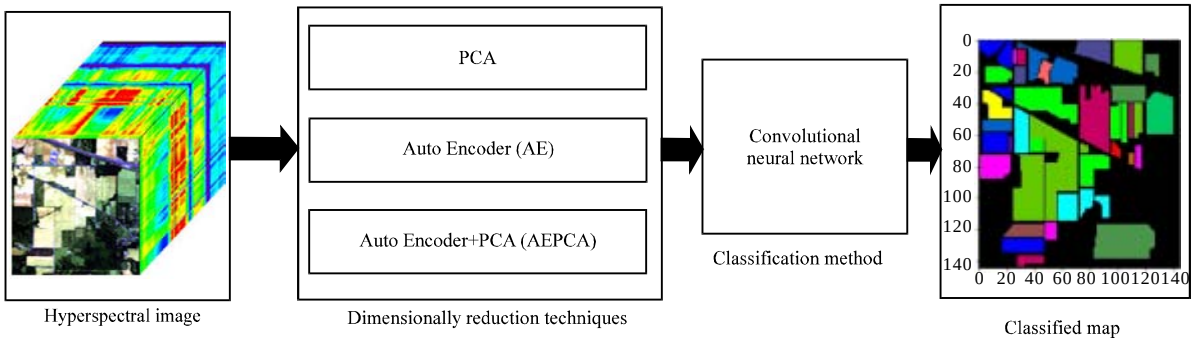

Fig. 1: Architecture for proposed HSI dimensionality reduction and classification model

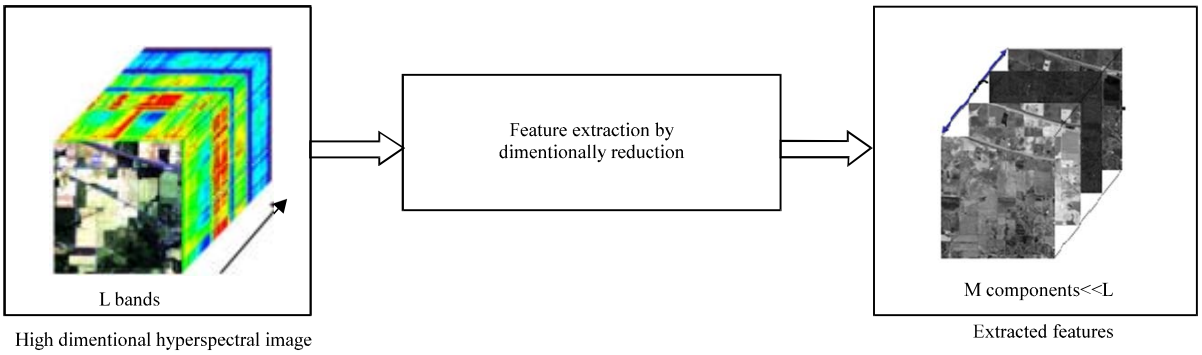

Fig. 2: Feature extraction using dimensionality reduction method for HSI

Dimensionality reduction: The primary scheme of dimensionality reduction for HSI is given as shown in Fig. 2. The input L spectral bands are reduced to M components which are small numbers compared to L. HSI is represented by 3D image with the dimension of R^{RxCxL}, each pixel P_i is represented as $P_i = [P_{i1}, P_{i2}, ..., P_{iL}]$ where, L is the number of spectral channels and i = [1, ..., M] where, M represents spatial features in 2D. Here, M = RxC, in which R is the number of rows and C is the number of columns. The objective of reduction of dimensions of HSI is to get a reduced image I_R from I_{in} with the lesser dimensions M where M<<L. The dimensionality reduction can be expressed as:

$$I_R = f(I_{In}, L) \quad (1)$$

Where:
f (.) = The reduction function
L = The No. of spectral bands of input image I_{in}

In this study, f (.) is considered for experimentation are unsupervised and linear reduction method PCA, unsupervised and non linear reduction method AE and AEPCA method which is combination of both AE and PCA method.

Dimensionality reduction using autoencoder: The autoencoder model can be used to both denoising and reduction of bands in HSI (Vincent *et al*., 2010). The AE Model primarily constructed using multilayer back propagation neural network. This model consists of 1 input layer, 1 output layer and many hidden layers. In this architecture, the size of input and output layer nodes are equal to L spectral bands. All hidden layers has bias nodes. The main function of AE is that it takes the pixel vector P as an input to the network and reconstruct the input pixel P at the output O. This is done by training the autoencoder model in unsupervised manner. Once, the AE Model is trained up to the desired accuracy, pixel P represented with L number of spectral bands is effectively represented at the latent space with M number of nodes where M<<L. By utilizing this, L dimensions are reduced to M dimensions and also this model denoise the noisy bands if any. The computation is carried out as in the procedure from Eq. 2-8 by using the notation which are given in Fig. 3:

$$h_{Inet} = f(W_{hl}, b_{hl}) \quad (2)$$

Where:

$$h_{Inet} = b_{hl} + \sum_j^L W_j P_j \quad (3)$$

For all K nodes and b_{h1} is the bias at h_1 layer. Similarly for consecutive layers it is shown that:

$$h_2 = f(W_{h2}, b_{h2}) \quad (4)$$

$$h_3 = f(W_{h3}, b_{h3}) \quad (5)$$

$$o = f(W_o, b_o) \quad (6)$$

The main intention of training of the model is to reduce the reconstruction loss between input p and output pixel o where, C is a loss function which is given as in Eq. 7. The mean square error is used as loss function which is expressed in Eq. 8:

$$p, o = \arg\min\left[C\left(p^{(i)}, o^{(i)}\right)\right] \quad (7)$$

$$C(w, b) = \frac{1}{2n}\sum_P \|p\text{-}o\|^2 \quad (8)$$

Where:
w = The set of all weights
b = The biases
n = The total number of training samples
p = The set of input
o = The actual output for the input p

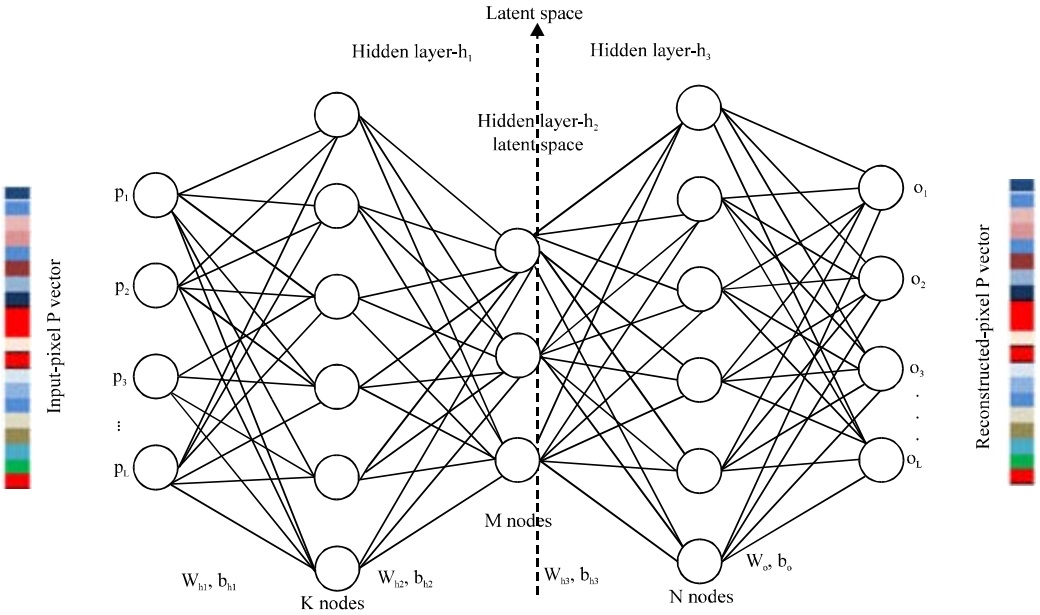

Fig. 3: A multi layer denoising autoencoder for HSI

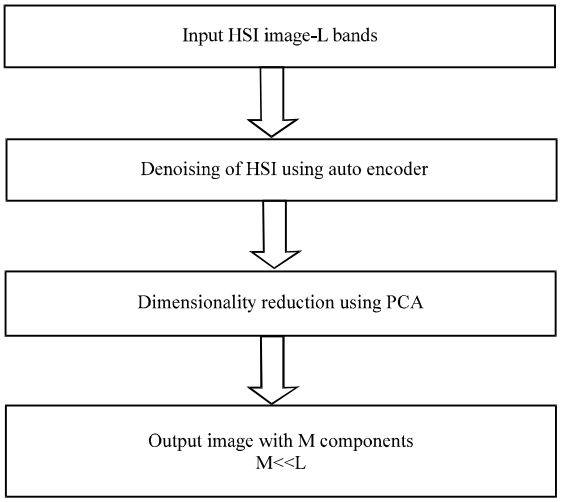

Fig. 4: Dimensionality reduction using AE PCA method

Dimensionality reduction using AEPCA: In this method, the HSI is denoised using AE method as explained above. But the reduced bands are not extracted from latent space as given in that study. Rather, the reconstructed image at the output layer is considered for the application of PCA in the next stage for dimensionality reduction. The procedure for application on AEPCA method on HSI is depicted in Fig. 4. PCA is a statistical technique used to transform, inter correlated variables into a set of new potentially uncorrelated variables. The linear transformation of the HSI using PCA results in principle components these may be more interpretable than the original spectral bands. Few number of M components with maximum amount of variance can represent large number of L spectral bands. This ability is an important economic condition which saves space and computation time in advanced stages of HSI processing. It is required compute covariance, correlation, eigen values and eigen vectors while applying PCA on HSI. The eigen values are computed using the relation as in Eq. 9. The Eigen values are represented as $E = [\bullet_1, \bullet_2, \bullet_3, ..., \bullet_L]$ and Eigen Vectors $EV = [a_{kp}, ...,$ for k = 1 to L bands and p = 1 to M components] of the covariance matrix computed:

$$[A].[X] = \lambda[X] \quad (9)$$

Where:
[A] = Covariance matrix
[X] = Eigen vector
• = Eigen values

The resulted eigen vectors are sorted in descending order such that first principle component must have more details. Then top M components are chosen which are able to represent more than 99% of the input details.

Algorithm for dimensionality reduction of HSI:

Step 1; Denoising by autoencoder model:

i. Build the autoencoder model for a given data set. Initialize the appropriate parameters
ii. Preparation of the dataset-normalize image pixels to [0-1] value. Separate the pixels training and testing samples
iii. Train the model with the training dataset until the desired accuracy is achieved for reconstruction or permissible threshold error
iv. Test the model with testing samples. If the testing accuracy is less than the desirable, then fine tune the hyper parameters of the model and repeat the process

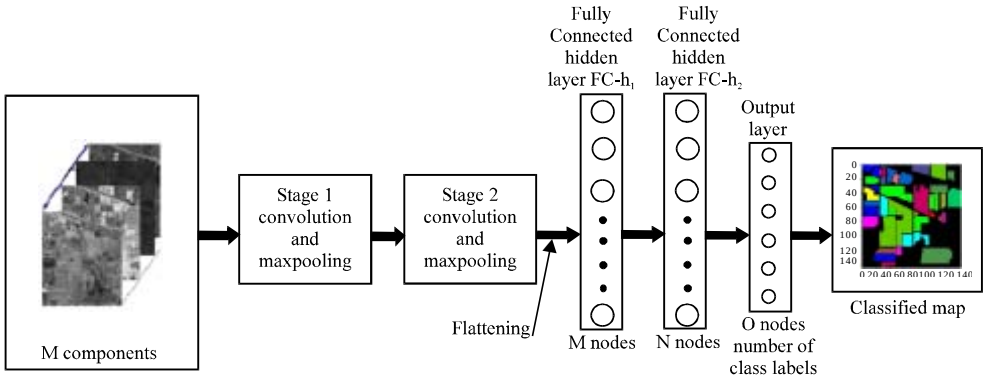

Fig. 5: HSI classification CNN Model for dimension reduced components

Step 2; Dimensionality reduction using PCA:

i. Set the number of principle components M
ii. Apply PCA algorithm to the HSI with N spectral bands
iii. Sort the principle components in descending order with respect to their variances
iv. Select the first M components which is nothing but dimension reduced HSI

Classification using CNN: In this study, classification of reduced channels using CNN Model is proposed. It is a potential feature learning model and consist of convolution, pooling and FCN layers (Santara *et al*., 2017; Hu *et al*., 2015). It is capable of learning both macro and micro level features which ensures the better classification compared to traditional models. This model is adaptive to the features of the specified input images and trains both feature extractor and classifier in supervised mode. This model is based on the principle of deep neural network architecture which learns spectral-spatial features (Lin *et al*., 2013). The architecture diagram which presented in Fig. 5 gives details of steps involved in the classification process. The steps are patch extraction convolution and maxpooling (Stage 1 and 2) Fully Connected layer (FCN). The output of the model is predicted class labels for each corresponding pixel of HSI. The whole model is trained by back propagating the error which is the difference between predicted and ground truth pixel.

The first step is to extract the patch of the M components which represents both spatial and spectral information. In this step, a pixel P_i represented by a patch of the input HSI with its neighbourhood of the size R_pxC_p (Spatial adjacent features), along M_p components (Spectral features) is extracted. The successive pixel representing patches, contains the overlapping window of the neighbourhood patches. Then, every patch of the size R_pxC_pxM_p is fed to stage i which is consists of a convolution layer and a maxpooling layer. The entire patch is convolved by the weights of kernels of the size typically 3×3 or 5×5 which is smaller than the patch size. The kernel weights extracts features local to the window as well as across the whole image. In the same way, F_1 number of feature maps are produced, each representing specific feature. Since, the M components are input to this layer, the number of F_1 varies depends on the M components. Hence, computation complexity of CNN will drastically reduces. The convolution function is computed as given Eq. 10. The convolved output is fed to Rectified Linear Unit (ReLU) activation function as in Eq. 11:

$$A_i^l = \sum_j f\left(w_{ij}^l * A_i^{l-1} + b_i^l\right) \quad (10)$$

Where:

A_i^l = The activation of the ith feature map in lth layer
A_i^{l-1} = The activation in ith feature map in l-1th layer that is previous layer
$w^l_{i,j}$ = The convolution kernel (Filter kernel) for A_i^{l-1}, b^l is bias
* = Represents convolution operator

$$F_i = f\left(A_j^i\right) = \max\left(0, A_i^l\right) \quad (11)$$

where, F_i is feature map of ith layer. The output of activation function is fed to maxpooling layer where in the maximum value is chosen from the local pool window of the feature map. This operation makes the features to invariant to location and distortion (Meher, 2015; Hu *et al*., 2015) by reducing the spatial size of the feature maps. After maxpooling, successive stage will get half of the size of the data of its predecessor stage.

The stage 2 layers helps in detecting higher level features such as curves and textures. Finally, the feature maps may have the patterns present in the scene as total visual concept. By adding the stages it increases the visual hierarchy of the concepts with smaller

representation. Therefore, similar operation of stage 1 is repeated in stage 2 with same convolution and maxpooling operations. The stage 2 output feature maps of the size $(R_p/4)x(C_p/4)xF_4$ are converted to 1-D vector to feed to FCN. At the output layer, the input pixel is classified using softmax activation function. The prediction of the label for a pixel is calculated based on the probabilities of O number of output classes as in Eq. 12:

$$P_i = \frac{e^{y_i}}{\sum_{i=1}^{O} e^{y_i}} \quad (12)$$

Where:

$y_i = [y_1, y_2, y_3, ..., y_O]$ = The input to softmax activation function at the output layer

P_i = The conditional probability of the ith class

Algorithm for classification:

i. Build the classification Model using CNN. Set the appropriate model parameters
ii. Prepare of the Dataset-Normalize image pixels to [0-1] value. Separate the pixels training and testing samples
iii. Train the model with the training dataset until the desired accuracy or permissible threshold error
iv. Test the model with testing samples. If the testing accuracy is less than the desirable, then fine tune the hyper parameters of the model and repeat the process

RESULTS AND DISCUSSION

In this study, the datasets used, experimental setup, obtained results at various conditions and result analysis are discussed.

Experiment set up and datasets: To study the effect of dimensionality reduction of HSI different images with diverse specification are considered in this experiment. Therefore, the images captured using different sensors and having significant environment variations in land cover, spectral and spatial resolutions are chosen. The Indian pines (Baumgardner *et al*., 2015) image is acquired from AVIRIS (Airborne Visible Infrared Imaging Spectrometer) sensor and Pavia University image (Baumgardner *et al*., 2015) of Italy is captured by ROSIS (Reflective Optics System Imaging Spectrometer) sensors. The implementation has been done using python with the help of tensor flow which is an open-source software library. The autoencoder deep network is built for both denoising and dimensionality reduction with 4 hidden layers. The input and output layers has the number of nodes equal to the number of bands of HSI. In case of Indian pines images the number of bands are 220 and in Pavia University scene the size is 103. About 60% of samples are used for training, 20% for validation testing and rest of the samples are used for testing purpose. In this experiment the loss function used is mean squared error and L_1 regularizer used for penalize the large magnitude weights. The learning rate is tuned between 10^{-2}-10^{-6}.

Table 1: Comparison of classification accuracy obtained after dimensionality reduction using PCA, AE and AEPCA for classification of Indian pines images

No. of components	PCA	AE	AE-PCA
10	0.8881	0.8069	0.9012
20	0.9363	0.8225	0.9431
30	0.9469	0.8300	0.9669
50	0.9587	0.8912	0.9735
100	0.9606	0.8875	0.9681
150	0.9712	0.9225	0.9790

The CNN Model used in this experiment for HSI classification is built with the convolution layers, maxpooling layers and fully connected network, respectively. In this phase, the patch of an image of the size $R_pxC_pxN_b$ is extracted from the principle component and fed to CNN Model. The filter size is $3x3xF_n$ for all convolution layers in which F_n is the size of feature maps. The number of features maps chosen depend upon the number of components which are input to the CNN Model. To reduce the spatial size of the feature maps to the half, maxpooling operation is carried out followed by convolution step. The resultant feature of stage 2 is flattened as single dimension vector and fed to FCN. In this experiment, the size of the hidden layer nodes are dependent on the size of 1-D vector. The activation functions selected for hidden layers and output layers are ReLU and sigmoidal functions, respectively. The performance of the proposed model is assessed using overall accuracy:

$$\text{Overall accuracy} = \frac{\text{No. of successful prediction}}{\text{Total No. of prediction}} \quad (13)$$

Dimensionality reduction and classification of Indian pines image: To assess performance of the proposed model, different datasets which represents the diverse environmental setting are considered (Table 1). In this case, Indian pines image covers the area of 2×2 miles by 145×145 pixels with the spectral resolution of 224 bands in the range of 0.4-2.5 μm which covers both visible and infrared region. This dataset contains one quarter area of building, railway tracks and highways and remaining 3 quarter area is covered by forest vegetation and agricultural land. The available ground truth for this scene has sixteen different classes such as steel structures, road, railway track, wheat fields, soybean, oats, etc. The overall accuracy as in Eq. 13 is used as performance measuring parameter for classification after dimensionality reduction. In this image scene, some classes have less

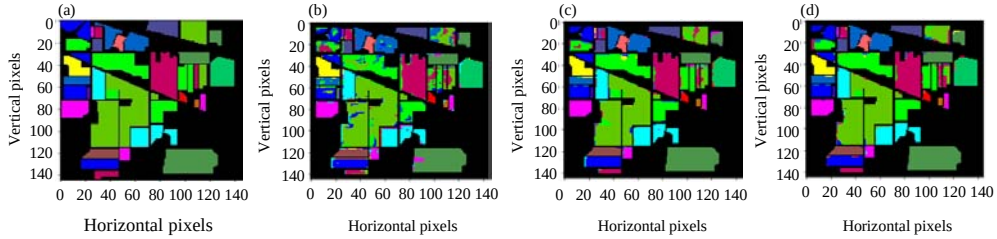

Fig. 6: Classification map of Indian pines image for 30 components using: a) Ground truth; b) AE -30 bands; c) PCA-30 bands and d) AEPCA-30 bands

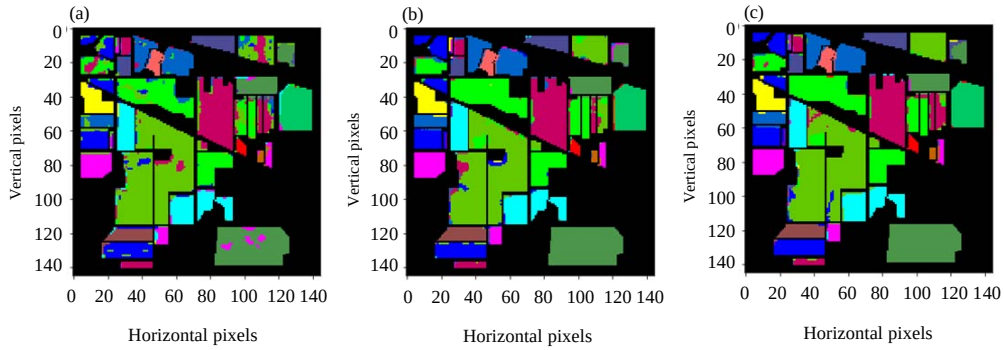

Fig. 7: Classification map of Indian pines image for 50 components using: a) AE-50 bands; b) PCA-50 bands and c) AEPCA-50 bands

samples such as 45 samples for alfalfa and very large number such as 2456 for soya bean. The obtained results from the experiment are presented in Table 1 which lists testing accuracy at different number of components from 10-150 for AE, PCA and AEPCA methods. By the observation from Table 1, the less number of components yielding less accuracy but after 50 input components to classifier, the accuracy is not significantly increasing. The visual inspection of ground truth image and the predicted image has been carried out at the component numbers 10, 20, 30, 50, 100 and 150.

For the perception quality observation of the dimensionality reduction models, the ground truth andpredicted images are presented in Fig. 6 a-d for 30 components. Similarly, for 50 components. The predicted images are given in Fig. 7a-c. It is clearly demonstrates that the predicted image obtained by AEPCA method has better classification accuracy of 97% compared to AE and PCA methods. The loss or the error value across different number of components are observed as shown in Fig. 8 and 9. From this graph it is evident that the loss value after 30 components reached 0.02 for AEPCA method. The AEPCA Model for dimensionality reduction is outperforming with classification accuracy of 97% by 30 components which is equal to 14% of total number of spectral bands.

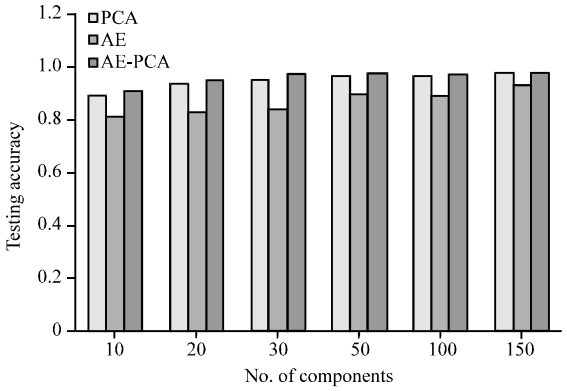

Fig.8: Classification accuracy after dimensionality reduction for Indian pines image

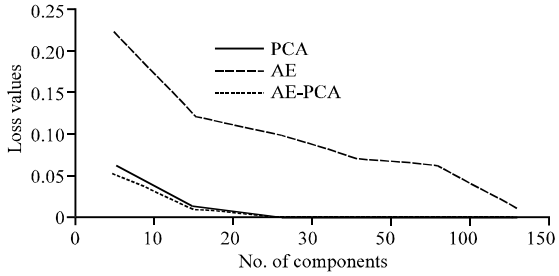

Fig. 9: Classification loss after dimensionality reduction for Indian pines image

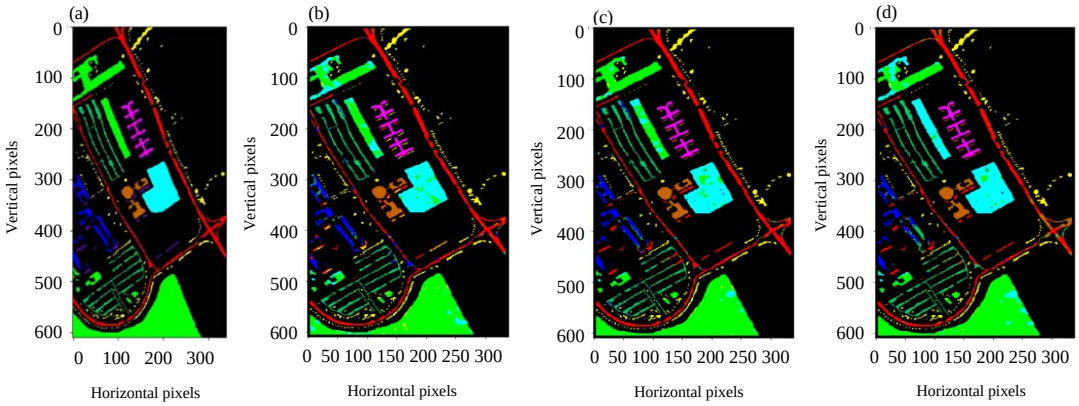

Fig. 10: Classification map of Pavia University image for 30 components using: a) Ground truth; b) AE-30 bands; c) PCA-30 bands and d) AEPCA-30 bands

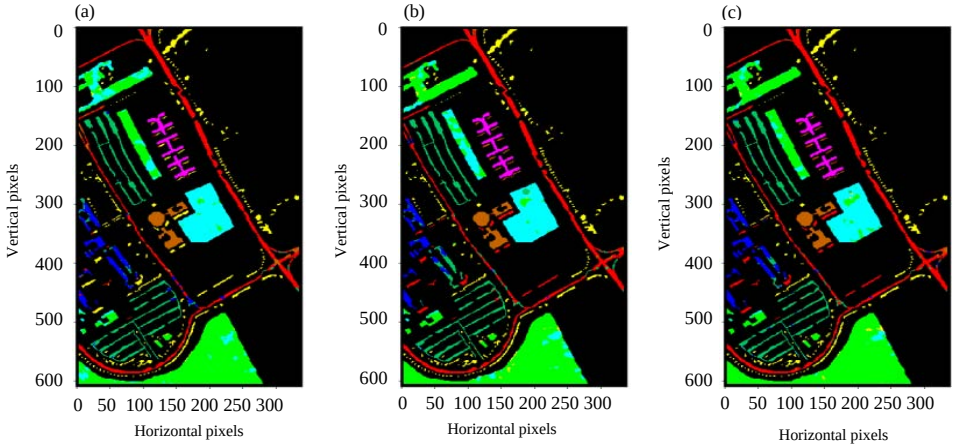

Fig. 11: Classification map of Pavia University image for 50 components using: a) AE-50 bands; b) PCA-50 bands and c) AEPCA-50 bands

Dimensionality reduction and classification of Pavia University image: To experiment the model with different set of characteristics, Pavia University image has been selected. In contrast to Indian pines image it has the better spatial resolution of 1.2 m^2 and has spectral resolution of 103 bands in the range of 430-860 nm. The image size is 610×340 pixels. The dataset has the signature of the objects like building, roads covered with different materials like bitumen, asphalt, gravel, land cover by meadows, bare soil, metal sheets structures. The ground truth image has only 9 classes. This dataset is applied to the proposed model to study the performance at the diverse conditions.

From the results from Table 2 and Fig. 12, it can be examined that the classification accuracy listed in the table for corresponding number of dimensionally reduced components, more number components producing the more accuracy. But the accuracy for AEPCA is better compared to other techniques. The balance between the number of components and the accuracy can be managed by compromising with very small loss in accuracy. For 30 components, the accuracy is 92.63% while for 50 components it is 97.55% for AEPCA method with the loss of 5% accuracy. It means, there is a reduction of 5 times in space and time complexity with the loss of 5% accuracy. The predicted output of the AE, PCA and AEPCA methods along with ground truth images are shown in Fig. 10a-d by considering 30 components as input to classifier. Similarly by considering 50 components as

Table 2: Comparison of classification accuracy obtained after dimensionality reduction using PCA, AE and AEPCA for classification of Pavia University images

No. of components	PCA	AE	AE-PCA
10	0.9433	0.8928	0.9263
20	0.9113	0.8888	0.9355
30	0.9042	0.9225	0.9413
50	0.9615	0.9640	0.9755

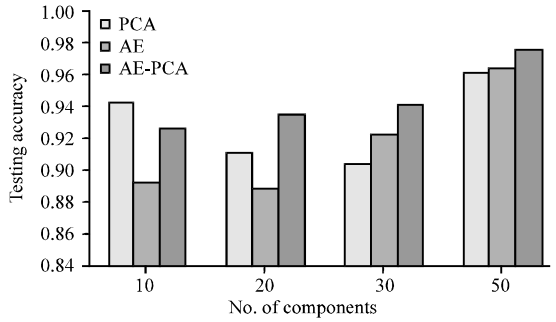

Fig. 12: Classification accuracy after dimensionality reduction for Pavia image

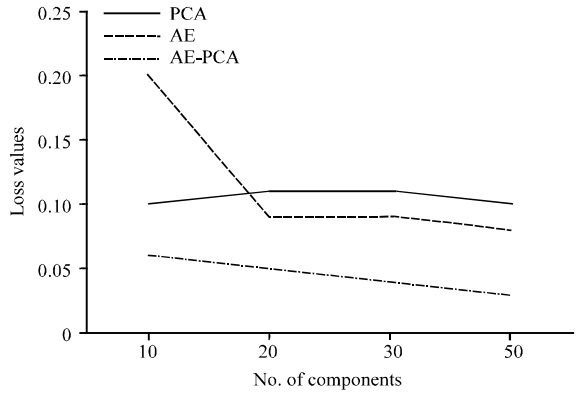

Fig. 13: Classification loss after dimensionality reduction for Pavia image

input to the classifier, the resultant images are shown in Fig. 11a-c and 12. From the loss function graph which is specified in Fig. 13, shows the loss value obtained by AEPCA is moderately varying from 10-50 components and has lowest loss compared to other methods.

CONCLUSION

In this study, a new HSI dimensionality reduction model which is combination of AE and PCA is proposed. To verify the performance of the model, CNN based classification is applied on the reduced number of components. This model is evaluated with individual AE and PCA techniques and found that outperforming against those methods. The same model is applied to other image with diversified characteristics, then also AEPCA is performing better. The model is used to produce 10, 20, 30, 50, 100 number of components and evaluated for its classification accuracy. It is found that more than 90% of accuracy obtained by 10 components reduced out of 220 bands. From this it can be inferred that AEPCA Model is out performing with the classification accuracy of 97% with the quantity of <15% of original spectral bands.

ACKNOWLEDGEMENT

The researchers are thankful to the R&D Centre, Information Science and Engineering, Bangalore Institute of Technology for providing us amenities to carry out our research work. The researchers also thank management of BIT, Visvesvaraya Technological University (VTU), Belagavi for their timely kind support.

REFERENCES

Amato, U., R.M. Cavalli, A. Palombo, S. Pignatti and F. Santini, 2009. Experimental approach to the selection of the components in the minimum noise fraction. IEEE. Trans. Geosci. Remote Sens., 47: 153-160.

Arsa, D.M.S., G. Jati, A.J. Mantau and I. Wasito, 2016. Dimensionality reduction using deep belief network in big data case study: Hyperspectral image classification. Proceedings of the 2016 International Workshop on Big Data and Information Security (IWBIS), October 18-19, 2016, IEEE, Jakarta, Indonesia, ISBN:978-1-5090-3478-9, pp: 71-76.

Baumgardner, M.F., L.L., Biehl and D.A. Landgrebe, 2015. 220 band Aviris hyperspectral image data set: June 12, 1992 Indian pine test Site 3. Purdue University Research Repository,ý West Lafayette, Indianaý, USA. https://purr.purdue.edu/publications/1947/1

Hinton, G.E. and R.R. Salakhutdinov, 2006. Reducing the dimensionality of data with neural networks. Science, 313: 504-507.

Hu, W., Y. Huang, L. Wei, F. Zhang and H. Li, 2015. Deep convolutional neural networks for hyperspectral image classification. J. Sens., 2015: 1-12.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE. Trans. Inf. Theor., 14:55-63.

Jia, P., M. Zhang, W. Yu, F. Shen and Y. Shen, 2016. Convolutional neural network based classification for hyperspectral data. Proceedings of the 2016 IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), July 10-15, 2016, IEEE, Beijing, China, ISBN:978-1-5090-3333-1, pp: 5075-5078.

Jolliffe, I.T., 2002. Principal Component Analysis. 2nd Edn., Springer, New York, USA., ISBN:978-0-387-22440-4, Pages: 488.

Laparra, V., J. Malo and G. Camps-Valls, 2015. Dimensionality reduction via regression in hyperspectral imagery. IEEE. J. Sel. Top. Signal Process., 9: 1026-1036.

Lin, Z., Y. Chen, X. Zhao and G. Wang, 2013. Spectral-spatial classification of hyperspectral image using autoencoders. Proceedings of the 9th International Conference on Information, Communications and Signal Processing, December 10-13, 2013, IEEE, Tainan, Taiwan, ISBN:978-1- 4799-0433-4, pp: 1-5.

Lu, D. and Q. Weng, 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens., 28: 823-870.

Meher, S.K., 2015. Knowledge-encoded granular neural networks for hyperspectral remote sensing image classification. IEEE. J. Sel. Top. Appl. Earth Obs. Remote Sens., 8: 2439-2446.

Santara, A., K. Mani, P. Hatwar, A. Singh and A. Garg *et al*., 2017. Bass net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification. IEEE. Trans. Geosci. Remote Sens., 55: 5293-5301.

Santos, L.C.B.D., S.J.F. Guimaraes and J.A.D. Santos, 2015. Efficient unsupervised band selection through spectral rhythms. IEEE. J. Sel. Top. Signal Process., 9: 1016-1025.

Sellami, A. and M. Farah, 2018. Comparative study of dimensionality reduction methods for remote sensing images interpretation. Proceedings of the 2018 4th International Conference on Advanced Technologies for v Signal and Image Processing (ATSIP), March 21-24, 2018, IEEE, Sousse, Tunisia, ISBN:978-1-5386-5240-4, pp: 1-6.

Sun, W., A. Halevy, J.J. Benedetto, W. Czaja and W. Li *et al*., 2014. Nonlinear dimensionality reduction via. the ENH-LTSA method for hyperspectral image classification. IEEE. J. Sel. Top. Appl. Earth Obs. Remote Sens., 7: 375-388.

Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio and P.A. Manzagol, 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res., 11: 3371-3408.

Zabalza, J., J. Ren and S. Marshall, 2015. On the fly dimensionality reduction for hyperspectral image acquisition. Proceedings of the 2015 23rd European Conference on Signal Processing (EUSIPCO), August 31-September 4, 2015, IEEE, Nice, France, ISBN:978-0-9928-6263-3, pp: 749-753.

Zhao, B., L. Gao and B. Zhang, 2016. An optimized method of kernel minimum noise fraction for dimensionality reduction of hyperspectral imagery. Proceedings of the 2016 IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), July 10-15, 2016, IEEE, Beijing, China, ISBN:978-1-5090-3333-1, pp: 48-51.

Zhou, Y., J. Peng and C.P. Chen, 2015. Extreme learning machine with composite kernels for hyperspectral image classification. IEEE. J. Sel. Top. Appl. Earth Obs. Remote Sens., 8: 2351-2360.