Enhancing Vegetation Classification through Hyperspectral Remote Sensing: Objectives and Methodological Approach

Debsunder Datta, Amgoth Pavan Kumar
CEE 277: Remote Sensing in Ecohydrology
Department of Civil Engineering
Indian Institute of Science, Bangalore
ddutta@iisc.ac.in, amgothpavan@iisc.ac.in

Abstract

Recent developments in remote sensing technology have led to a significant rise in the demand for precise and reliable algorithms for analyzing hyperspectral remote sensing images. Hyperspectral remote sensing offers a powerful tool for monitoring and classification of vegetation types with unprecedented spectral resolution. This work involves classification of hyperspectral images using different methodological approaches. This involves 3 steps. Firstly, classification of hyperspectral images using Support vector machine (SVM) and Random Forest (RF). Secondly, principal component analysis (PCA) is utilized for dimension reduction of the hyperspectral image and then reduced features are trained by SVM and RF. The PCA + SVM and PCA + RF methods are compared with the SVM and RF and it is observed that the PCA + SVM method gives better results in terms of classification accuracy.

Keywords: *hyperspectral image (HSI); support vector machine (SVM); random forest (RF; principal component analysis (PCA)*

I. INTRODUCTION

Hyperspectral image contains information about objects across the electromagnetic spectrum. The information in the hyperspectral datasets is crucial for tracking changes in the environment over time. The fact that HSI contains both spectral and spatial data is a key characteristic. The HSI classification process identifies the class of each pixel in the hyper image taken using (hyper) spectral sensors. Unlike color image stores the intensity value of three bands (red, green, and blue), HSI keeps intensity information of more than 100 bands which are in discrete intervals of 5-10 nm across the visible to the infrared region. These bands give information about each has a distinct range of value for a specific band which is also called a spectral signature. Thus, it makes it easy to identify and distinguish objects from the surface of the earth. This characteristic HSI suitable for land cover classification (*Ranjan et al 2017, Arun Solomon et al 2023*).

The goal of HSI is to obtain the spectrum for each pixel in the image of a scene, with the purpose of finding objects, identifying materials, and detecting processes. Spatial and spectral resolutions have inherently inverse relation between them due to manufacturing tradeoffs (i.e., capturing 3-D signals with a 2D sensor). Hence, it is hard to achieve both at the same time. It is worth mentioning that spatial resolution is also a function of the distance between the sensor and the target (*Aburaed et al, 2023*).

The intrinsic information content of HSI data, often expressed through its spectral dimensionality, determines the number of degrees of freedom required to fully represent the variations in reflected solar energy (*Jie Dai et al., 2022*). Machine learning classifiers have become a major focus for the remote sensing community since such algorithms deal with high dimensional feature space and are able to model complex class signatures (*Maxwell et al., 2018).* Nowadays the non-parametric machine learning algorithms SVM, RF and Neural Networks (NN) are state-of-the-art for remote sensing image classification (*Khatami et al., 2017*), producing higher levels of accuracy than parametric algorithms like Maximum Likelihood (*Yu et al., 2014)*.

In hyperspectral image the data points are distributed nonlinearly, the linear classifiers are not able to classify the data correctly. The solution to this problem is to train a nonlinear classifier, which can map the training data into a higher dimension (*Ranjan et al 2017*). SVM and RF can be utilized for this reason , however, studies have shown that the time complexity of training a SVM and a RF is quadratic with respect to the number of training data (*L.E. Christovam et al., 2019)*. Hence, there is a need to classify nonlinear data with less complexity. In this work, the hyperspectral images are classified using SVM and RF by training the data obtained from PCA. Simulations on the dataset show that PCA + SVM outperforms the method SVM, RF and PCA + RF with respect to accuracy.

II. RELATED WORK

Ranjan et al., 2017 presented an approach for HSI classification using PCA, k-means and M-SVM. The approach is compared with an existing technique that uses PCA and M-SVM. Validated the proposed scheme by simulations on three benchmark images and showed that the proposed method is better in terms of both classification accuracy and time complexity. Similarly, *Mounika. K et al*., 2021 described a brief about on image classification models using SVM with PCA carried upon one common hyperspectral dataset i.e., Indian Pines.

Guang Yi Chen, 2021 developed a novel method for hyperspectral image (HSI) classification based on principal component analysis (PCA) and SVM. This method benefits from both spatial and spectral resolution at the same time. *Christovam et al., 2019* goal was to provide a baseline for future research in LULC classification with the HyRANK dataset. They adopted traditional Spectral Angle Mapper (SAM) and two-of-the-art machine learning algorithms, SVM and RF.

Craig Rodarmel and Jie Shan studied the use of principal component analysis as a preprocessing technique for the classification of hyperspectral images. They used HYDICE and AVIRIS hyperspectral datasets for the study which revealed that the use of first few principal component images can yield about 70 percent correct classification rate and suggested the benefit and efficiency of using PCA technique as a preprocessing step for the classification of hyperspectral images.

In addition, *Kang et al., 2014* have presented a method based on image fusion and recursive filtering (IFRF) to extract significant features from different hyperspectral images and thereafter used SVMs for classification. They have achieved better results than PCA and ICA based schemes. In *Chen et al., 2016* work, they have applied a regularized deep feature extraction (FE) scheme for hyperspectral image (HSI) classification with the help of a convolutional neural network (CNN) and got significant results.

From the literature, it is observed that many techniques like PCA, SAM, SVM, RF, etc., have been used for HSI classification. In this work, an approach is used which involves PCA, SVM and RF and term it as PCA + SVM and PCA + RF.

III. METHODOLOGY

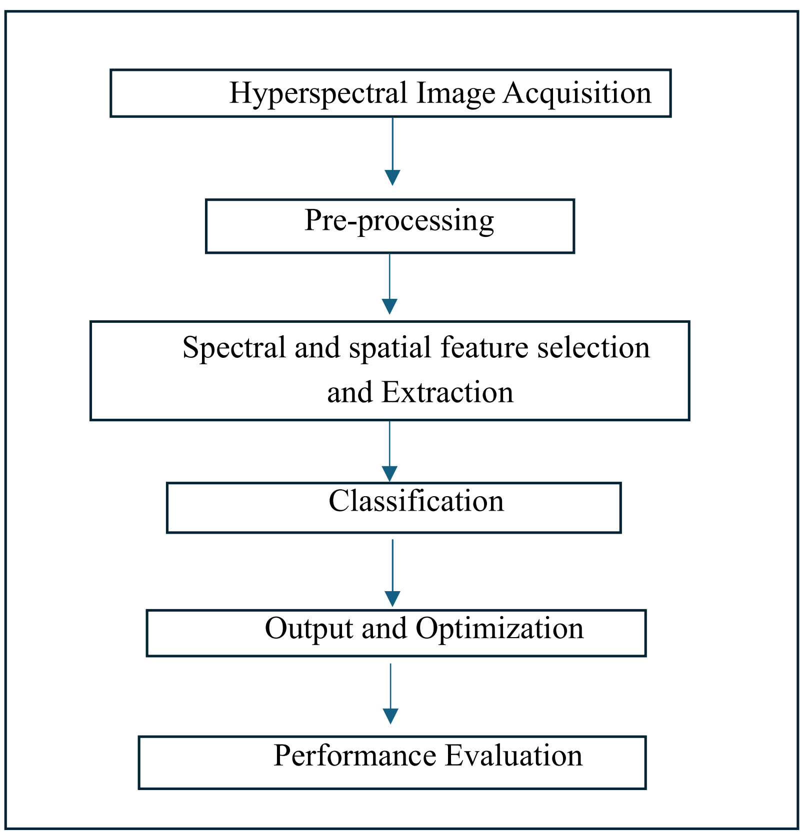

Fig. 1 Flow chart

The systematic implementation of HSI technique is represented in a sequence chart using the Principal Component Analysis with Support Vector Machine and Random Forest as shown in Fig 1. For HSI implementation using Corrected Indian Pines data set, with 200 spectral bands of unequal wavelengths are used.

Principal Component Analysis (PCA)

The principal component analysis is based on the fact that neighboring bands of hyperspectral images are highly correlated and often convey almost the same information about the object. The analysis is used to transform the original data so to remove the correlation among the bands. In the process, the optimum linear combination of original bands accounting for the variation of pixel values in an image is identified.

The PCA employs the statistical properties of hyperspectral bands to examine band dependency or correlation. Though, one may find many synonyms for PCA, such as Hotelling transformation or Karhunen-Loeve transformation, all these transformations are based on the same mathematical principle known as eigen value decomposition of the covariance matrix of the hyperspectral image bands to be analyzed (*Craig Rodarmel and Jie Shan*). Detailed discussion may be found in *Gonzalez and Woods* (1993) and *Schowengerdt* (1997).

Support Vector Machine (SVM)

SVM finds the optimal hyper plane that separates the classes in a multi-dimensional feature space. The best decision boundary is that which minimizes the errors and maximizes the distance between the training samples. SVM is especially useful for small training datasets since it relies only on observations located on the decision boundaries (support vectors). This advantage makes SVM more relevant for remote sensing applications, due to the problem of getting training samples which normally require field work and consequently high cost (*Cortes and Vapnik, 1995; Christovam et al., 2019).*

Despite the advantages, SVM has a drawback in the higher number of parameters to be tuned, a high computational cost and the need to choose a kernel function. The kernel function is used to transform the n-dimensional feature space into a larger dimension space where the classes are linear, polynomial, sigmoid, and the radial-base function (RBF). RBF kernel is used for this classification as it is one of the most suitable for remote sensing data classification (*Christovam et al., 2019).*

Random Forest Classifier (RF)

The RF algorithm is an ensemble of decision trees. It has been used in several remote sensing classification works because of its simplicity and good accuracy results.

Decision tree is a recursive split approach of the input data. The splits are performed starting from a root node (first level of tree) up to the leaf nodes, decreasing the entropy at each split. The leaves are the last level of the tree, and it is where the entropy is at its lowest possible. The intention is to have only samples from the same class in the leaves. There are several split nodes in the path that go from the root node to leaf node. These contain decision rules based on the available features and a threshold applied to the features chosen.

Despite the decision trees being extremely fast and simple, they are very sensible to noise and frequently overfit the training samples. Because of that, a decision tree can be classified as a weak learner. To overcome these drawbacks an ensemble of decision trees is combined in a RF (strong learner). In the RF algorithm the trees of the forest must be uncorrelated, each tree being unique, hence the random subspace (feature bagging) and bootstrap aggregating (bagging) techniques are applied.

Bootstrap aggregating consists of the random selection, with replacement, of a subset of samples from the training dataset. The random subspace consists of a randomly selected subset of features from all the input features at each node, and, from the new subset of features chosen, considering the one which splits the node that produces the smaller entropy at the next level.

As stated before, both techniques minimize the model variance without increasing the trend. So, while a single decision tree is sensitive to noise, the average forecast for an ensemble of trees is not sensitive, so long as the trees are uncorrelated. After growing the forest, each tree casts a vote for a class and the label is defined by the majority vote. The main advantage of the RF algorithm is in dealing well with noise, having fewer parameters to be tuned, and the low computational cost (Ho, 1995; *Leo Breiman, 2001; Hastie et al., 2001; Christovam et al., 2019*).

IV. EXPERIMENTAL SETUP

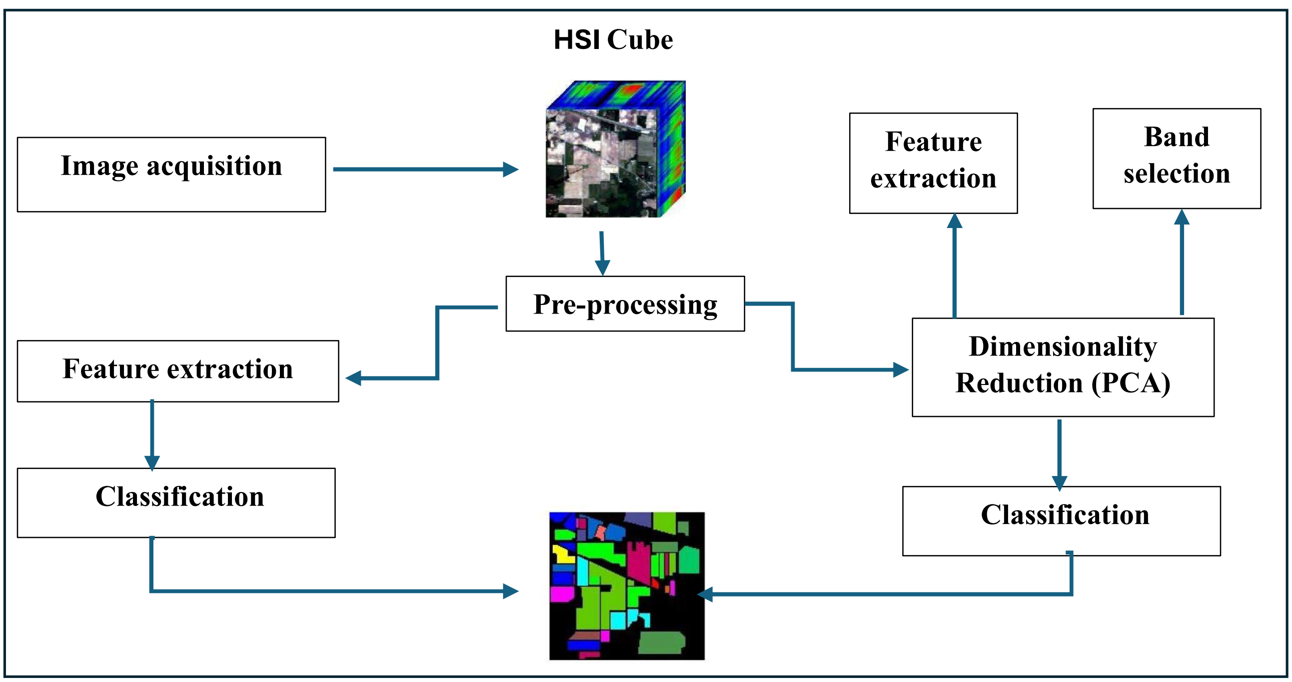

Fig. 2 Architecture

The architecture in fig. 2 describes the hyperspectral classification process with step-by-step representation.

Dataset

Indian Pines dataset used for the image classification. In Indian Pines dataset the data is taken by the AVIRIS spectrometer over the small area of North-Western Indiana, United States. The image is the subset of another image. The image consists of two-third agriculture land and one-third forest, houses, and roads. The dimensionality of the dataset is 145 X 145 pixels. It contains 145 rows, 145 columns, 16 classes, and 200 bands with the wavelength range from 400nm to 2500nm.

Pre-processing

The pre-processing is one of the major parts in the hyperspectral image classification because it shows significant impact based on the result classification. The main aim is to select those spectral bands having the maximum variance value since the maximum variance gives more proper information, So in this work PCA used to reduce the dimensionality of the dataset by eliminating the noise in the dataset.

Feature Selection

After completing the pre-processing phase of the dataset now it is required for the feature selection stage. In feature selection, work mainly focused on spectral signatures. In feature selection, it is found that the reflectance values by selecting the pixel intensity values. This can be done using a simple calculation i.e. by increasing the dimensions of the used dataset and it is remained unchanged the ground truth bands because they are all readily modified using PCA. The final spectral information will be available as result as 2-D or 1-D array.

V. RESULTS AND DISCUSSION

In this study the Indian Pines dataset classified using SVM and RF without PCA at first and evaluated the accuracy. Later SVM+PCA and RF+PCA were performed to classify the dataset and accuracy compared with initial classification methods.

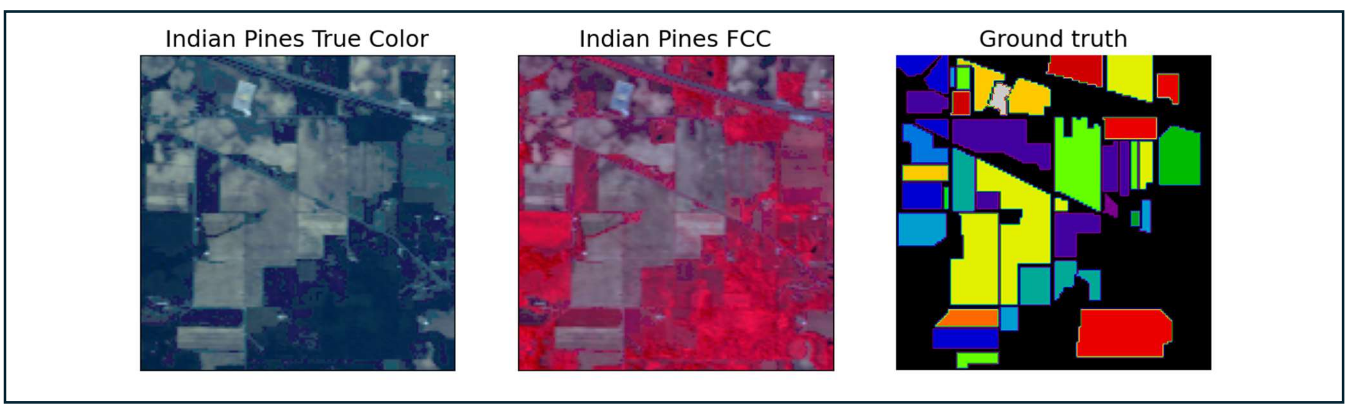

Fig. 3 Showing True color , False color composite (FCC) and Ground truth of Indian Pines dataset

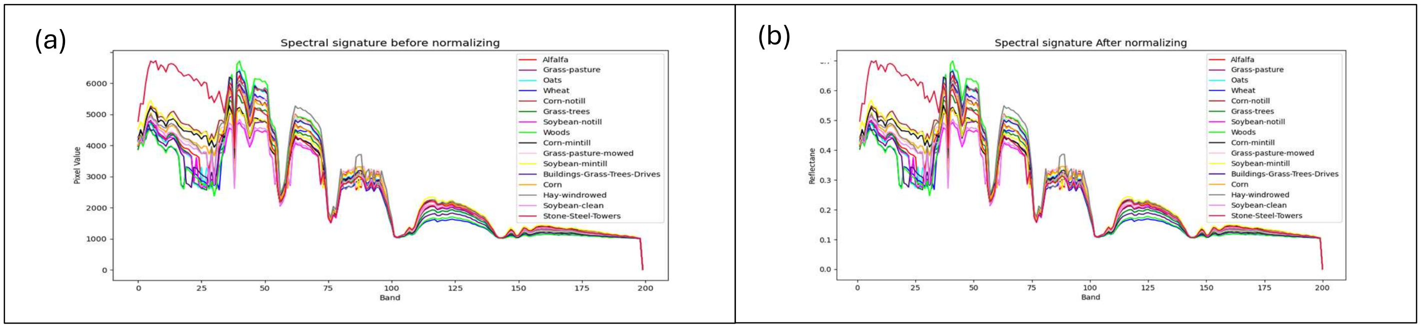

Fig. 4 Showing spectral signature of 16 classes before and after normalizing the dataset

The classified images were validated using ground truth on Indian Pines dataset. Fig.3 shows the ground truth of 'Indian Pines' image. The color portions represent different classes, whereas black portions for unclassified one. The black portions are not considered for training and testing.

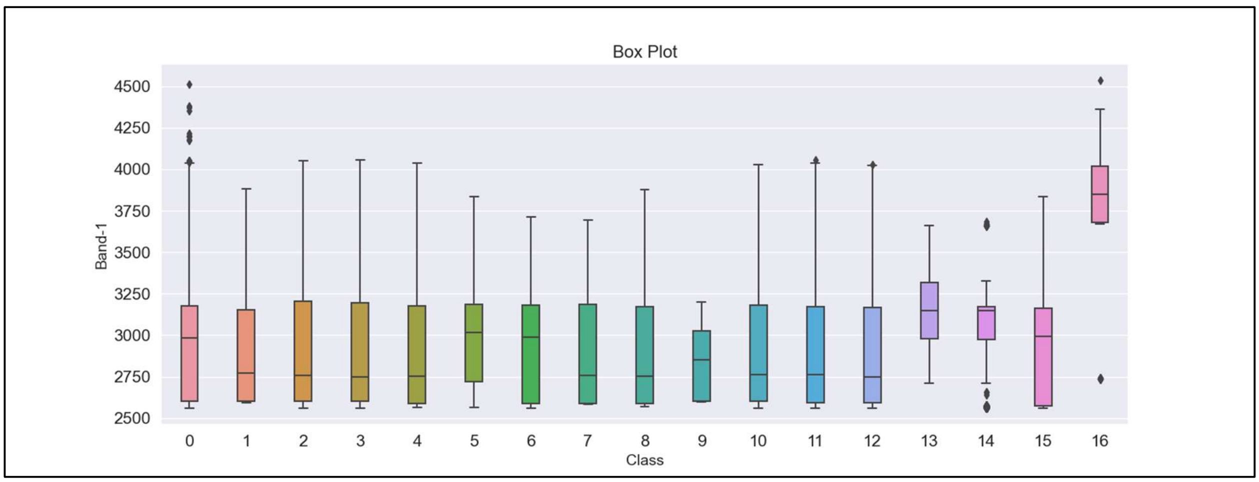

Fig. 5 Showing the boxplot of the 16 classes of the dataset

Fig. 4 (a) shows the spectral signature of 16 classes in the dataset with the pixel value in the y-axis and Fig. 4 (b) shows the reflectance value in y-axis after normalizing the dataset.

After implementing PCA using sklearn library in python, the relationship between signal and noise identified to understand the rate of information present in the first few PC bands. Cumulative explained variance plotted against the number of components to visualize the principal components in the dataset.

The PC bands are also plotted for visualization to understand which PC band gives more information of the dataset and identify at which PC band the noise in the data is prominent. From the results PC band 1 has more information (70%) compared to other PC bands. On PC band 4 and PC band 5 some noise has been introduced in the dataset. As PC band increases one can identify noisy dataset which can be neglected for feature extraction and band selection.

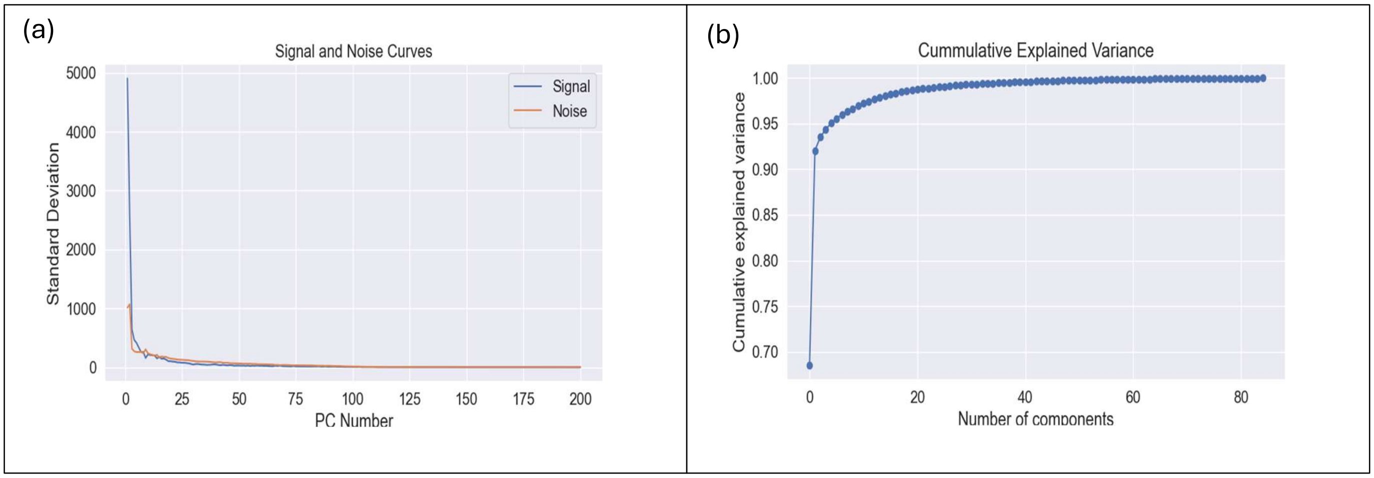

Fig. 6 (a) Showing Signal and noise relationship; (b) Showing the cumulative explained variance against number of components

SVM and RF Classification without PCA

Training and testing dataset taken from original dataset to classify the dataset using SVM and RF classifier. For this 70% taken as training and 30% taken for testing the dataset from original dataset. After running the classification algorithms Fig. 7 and Fig. 8 shows the classified image and confusion matrix of SVM and RF classification algorithms.

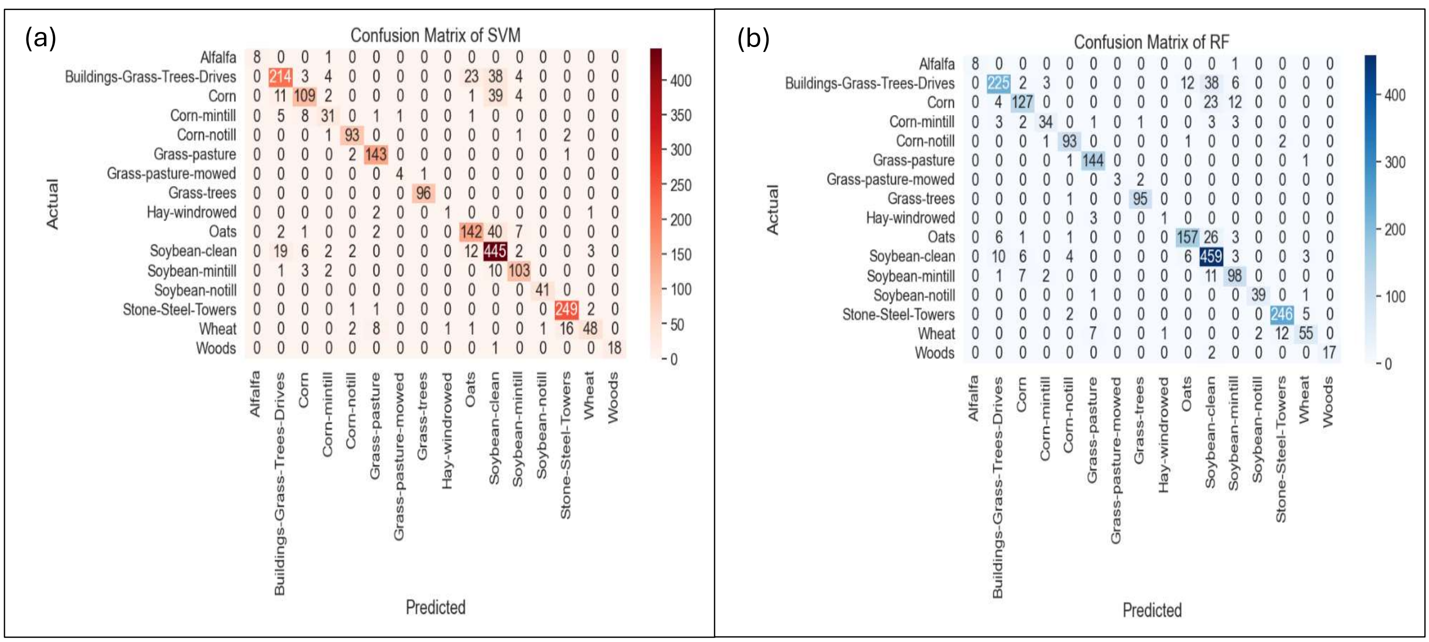

Fig. 7 (a) Showing Confusion matrix of SVM classifier (b) Showing Confusion matrix of RF Classifier

Confusion Matrix

The error matrix or confusion matrix compare the ground truth with the classification result. It checks for errors in comparison. The diagonal elements of the error matrix show the correctly classified pixels of each class out of total training samples. The off-diagonal elements value other than 0 show unclassified pixels. The rows in the error matrix show the actual samples of each class and columns show predicted pixels for each class.

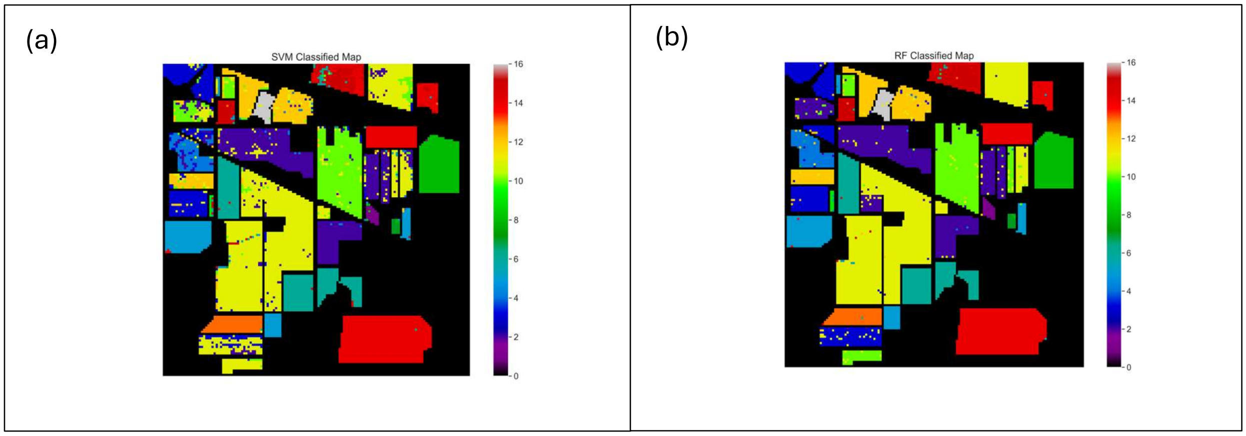

Fig. 8 (a) Showing SVM Classified Map; (b) Showing RF Classified Map

SVM and RF Classification with PCA

After calculating PCA, a total of 150 PC bands were selected from 200 bands to classify with SVM and RF. Of 150 PC bands 30 % data taken for testing and 70 % data for training the classifier.

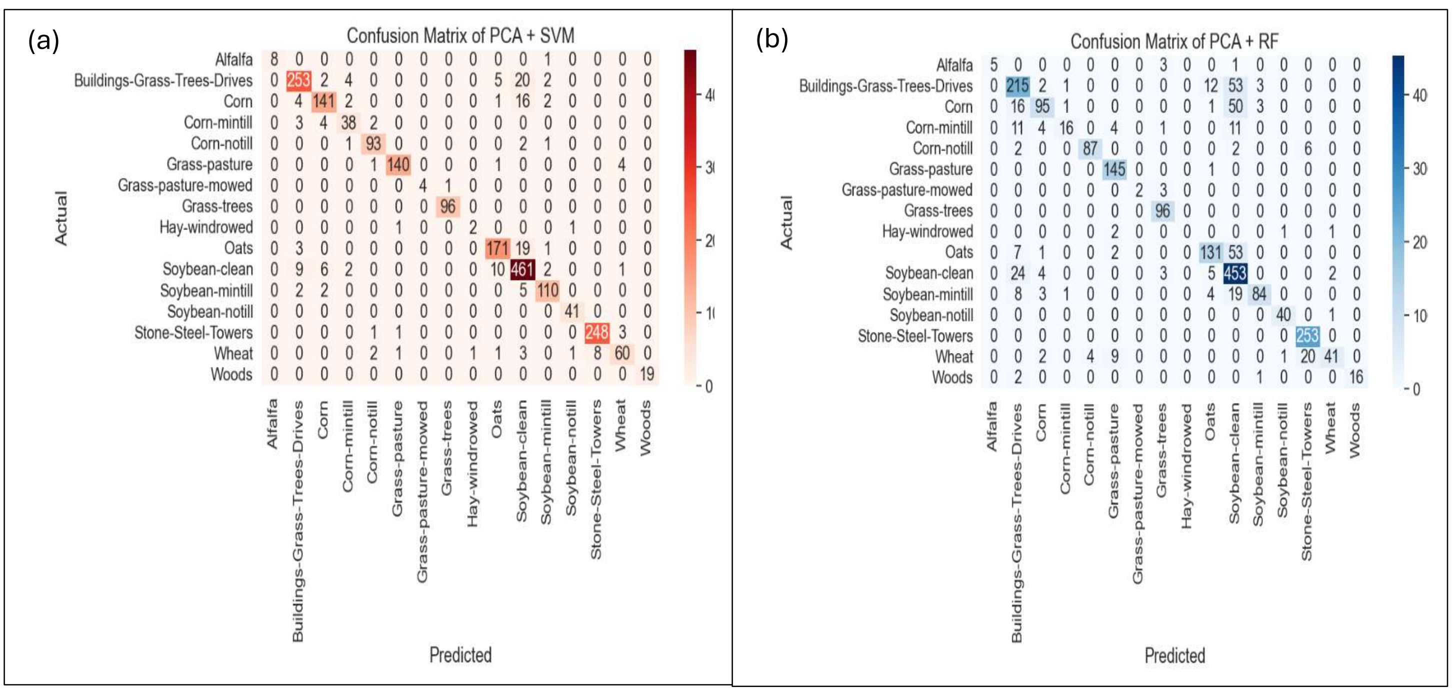

Fig. 9 (a) Showing confusion matrix of PCA + SVM; (b) Showing confusion matrix of PCA + RF

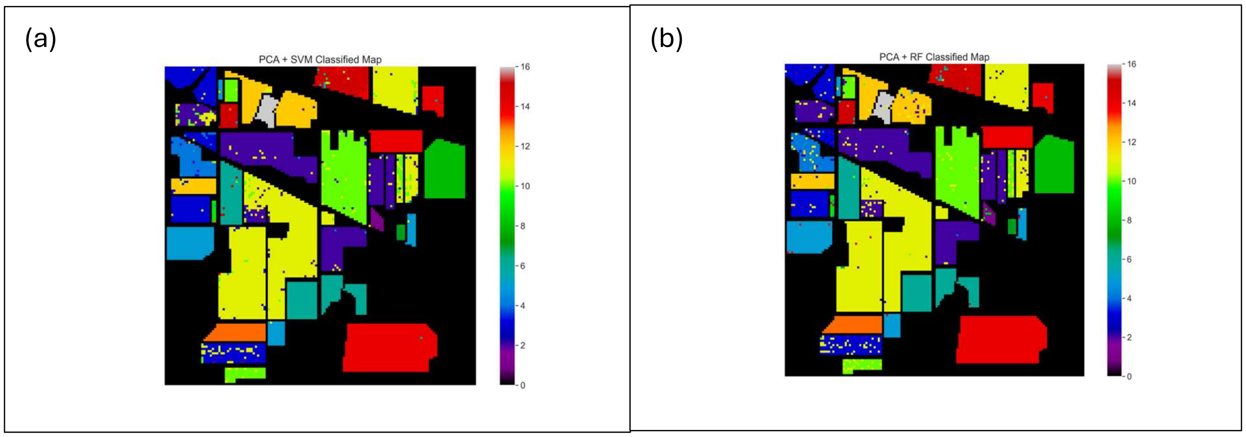

Fig. 10 (a) Showing PCA + SVM Classified Map; (b) Showing PCA + SVM Classified Map

Accuracy

The accuracy defines differences between the samples predicted values with the original set of given samples. The accuracy is shown in equation below

$$\{(T_P + T_N)\}/\{(T_P + T_N + F_p + F_N)\}$$

Table 1 Showing number of data points with respect to the image and comparison of classification accuracy of different methods for each class

Class	**Samples**	**Accuracy (f1-score)**			
		SVM	**RF**	**PCA + SVM**	**PCA + RF**
Alfalfa	46	0.94	0.94	0.94	0.50
Corn-notill	1428	0.80	0.84	0.90	0.77
Corn-mintill	830	0.74	0.83	0.88	0.66
Corn	237	0.69	0.78	0.81	0.44
Grass-pasture	483	0.94	0.93	0.95	0.92
Grass-tress	730	0.94	0.95	0.97	0.94
Grass-pasture-mowed	28	0.80	0.75	0.89	0.75
Hay-windrowed	478	0.99	0.98	0.99	0.96
Oats	20	0.33	0.33	0.57	0.00
Soyabean-notill	972	0.76	0.85	0.89	0.77
Soyabean-mintill	2455	0.84	0.87	0.91	0.79
Soyabean-clean	593	0.86	0.80	0.92	0.78
Wheat	205	0.99	0.95	0.98	0.99
Woods	1265	0.96	0.96	0.97	0.94
Buildings-Grass-trees-drivers	386	0.73	0.77	0.83	0.63
Stone-Steal-Towers	93	0.97	0.94	1.00	0.91

Table 2 Showing accuracy of different methods

Method	Overall classification accuracy
SVM	85.12 %
RF	87.85 %
PCA + SVM	**92.0 %**
PCA + RF	81.56 %

From Table 2 the overall classification accuracy of hyperspectral image classification for SVM and RF is less compared to PCA + SVM method. SVM method classified the image with 85.12 % accurate and RF with 87.85 % . In the case of classification without PCA, RF classified with greater accuracy compared to SVM but classification accuracy of PCA + RF is less compared to any other method. The PCA + SVM method gives 92.0 % overall accuracy for vegetation classification of hyperspectral image which means the PCA reduced the noise in the data and reduced the dimensionality which helped SVM to result in higher accuracy of classification.

Considering the computational time, the RF algorithm outperformed the SVM algorithm. However, taking into account just the time to fit the classification models, if the time to perform the PCA is considered, PCA + SVM and PCA + RF performed slower than SVM and RF.

VI. CONCLUSION

In this work, a brief description on image classification models using SVM, RF without and with PCA has been described. The study has been carried up on one common hyperspectral dataset i.e., Indian Pines comprise various landscape fields like dense vegetation, barren land, grasslands, etc., to reduce noisy band, PCA has been used. Also to achieve accuracy of classification using SVM some filter techniques (RBF) have been used. From the classification it shows that the overall classification accuracy of SVM improved from 85.12 % to 92.0 % by PCA + SVM method with RBF kernels with the same set of training data.

Although the results obtained for the machine learning algorithms, RF and SVM, are considered excellent, there is still room for improvement. One possibility is to perform the selection of bands that best contribute to the delimitation of the classes in the feature space so that the importance of each attribute generated by the RF algorithm can be used. Another possibility in the future work is to use deep NN in this dataset since it can improve the results because of its capacity for engineering new features. The results presented in this study can be used as a baseline for future research in hyperspectral image classification of Indian Pines dataset.

VII. REFERENCES

[1]. Ranjan, S., Nayak, D. R., Kumar, K. S., Dash, R., & Majhi, B. (2017, January). Hyperspectral image classification: A k-means clustering based approach. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1-7). IEEE.

[2]. Mounika, K., Aravind, K., Yamini, M., Navyasri, P., Dash, S., & Suryanarayana, V. (2021, October). Hyperspectral image classification using SVM with PCA. In *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 470-475). IEEE.

[3]. Christovam, L. E., Pessoa, G. G., Shimabukuro, M. H., & Galo, M. L. B. T. (2019). Land use and land cover classification using hyperspectral imagery: Evaluating the performance of spectral angle mapper, support vector machine and random forest. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 1841-1847.

[4]. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

[5]. Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., ... & Gong, P. (2014). Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, *35*(13), 4573-4588.

[6]. Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, *39*(9), 2784-2817.

[7]. Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote sensing of environment*, *177*, 89-100.

[8]. Dai, J., Vaughn, N. R., Seeley, M., Heckler, J., Thompson, D. R., & Asner, G. P. (2022). Spectral dimensionality of imaging spectroscopy data over diverse landscapes and spatial resolutions. *Journal of Applied Remote Sensing*, *16*(4), 044518-044518.

[9]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

[10]. Rodarmel, C., & Shan, J. (2002). Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, *62*(2), 115-122.

[11]. Arun, S. A., & Akila, A. S. (2023). Land-Cover Classification with Hyperspectral Remote Sensing Image Using CNN and Spectral Band Selection. *Remote Sens. Appl. Soc. Environ*, *31*, 100986.

[12]. Aburaed, N., Alkhatib, M. Q., Marshall, S., Zabalza, J., & Al Ahmad, H. (2023). A review of spatial enhancement of hyperspectral remote sensing imaging techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 2275-2300.

[13]. Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning. Springer series in statistics. *New York, NY, USA*.

[14]. Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

[15]. Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE transactions on geoscience and remote sensing*, *54*(10), 6232-6251.

[16]. Kang, X., Li, S., & Benediktsson, J. A. (2013). Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(6), 3742-3752.