# CNN with ICA-PCA-DCT Joint Preprocessing for Hyperspectral Image Classification

Aamir Naveed Abbasi* and Mingyi He†
International Research Center for Information Acquisition and Processing
School of Electronics and Information, Northwestern Polytechnical University, Xian, 710129, China
* amirnaveed@mail.nwpu.edu.cn † myhe@nwpu.edu.cn

***Abstract*—In this paper a simpler convolutional neural network with a joint pre-processing is proposed for hyperspectral image classification. Primarily the spectral dimension of raw hyperspectral data cube is reduced in a unique fashion by using PCA and DCT such that the data is reduced effectively but having much information intact for classification task. The raw data cube is divided into two small spectrally reduced cubes, the first cube (PCA cube) is a simple PCA based dimension reduction considering few top principal components and the second cube (PDCT cube) performing DCT as preliminary step which confined maximum energy into low frequencies and then subsequently applying PCA by selecting same number of principal components as in the first PCA cube. After that both PCA and PDCT cubes are fused together, furthermore ICA is carried out on fused data cube to make the output classes much independent for next steps. In the final pre-processing step, the ICA performed data cube is divided into small square patches having labeled center pixel and a fixed size neighboring pixels by considering that in hyperspectral image neighboring pixels are highly correlated and having same class label. These square patches are fed into our simpler convolutional neural network which effectively and automatically extract the suitable features for our classification prediction job. The results validated that our acclaimed model which mainly exploit a novel pre-processing tactic and simpler but effective CNN performs enormously well in comparison to the other compared models and can be used as an effectual classification model for hyperspectral images in particular.**

## I. Introduction

The vast spectral and spatial information embedded in hyperspectral images (HSI) make it more useful for various applications. The HSI data after slight revision, adaptation and processing can be useful for many fields. Few of them are agriculture [1], defense and security [2], food industry [3], medical [4], astronomy and environment monitoring. The HSI data cube contains spatial and spectral information and is quite different from typical RGB images in terms of information processing and labeled data availability. The classification is considered as the most widespread exploration area for HSI analysis, which ambitions at assigning a pre-defined class label to every pixel. Previously, many works have been suggested by employing spectral data alone [5], [6] but not gain much popularity and higher accuracy. Recently many works combined the spectral data and the spatial contextual information [7], [8] with better results and now it is implicit that both spectral and spatial data is imperative for better classification results.

Most recently, numerous deep learning (DL) architectures have emerged and have been successfully applied in the areas like audio recognition [9], natural language processing [10] and image classification [11], where they outperformed the traditional methods. Moreover for HSI classification in particular, DL based methods also reached up to the level where they used combined spectral-spatial methods for better classification. Few renowned DL methods for HSI classification are convolutional neural network (CNN) [12]–[14], stack autoencoders [15], deep belief network [16] and deep stacking network [17]. In particular, CNN has gained much fame due to their superior enactment in terms of automatic and appropriate feature extraction after exhaustive training. The CNN can acquire feature illustrations through several convolutional blocks and can hierarchically learn low level, mid-level and higher level features. The typical CNN architecture consists of several feature-extraction layer, one or more fully connected layer and a classifier. Where each feature extraction layer consists of a convolutional layer, a nonlinearity layer and optional pooling layer.

More recently, He [18] and Zhong [14] attained promising accuracy with 3D deep convolutional networks which are end-to-end approaches. However, the 3D CNN approach has higher complexity with deep 3D convolutional layers, excess of parameters, long training and testing time and most importantly the testing time is comparatively high which limits it for real time applications. In HSI many spectral bands are highly correlated and thus using dimension reduction approach is the most famed and direct solution to accommodate the Hughes phenomenon [19] using feature selection to find a suitable subset of the original spectral dimension of HSI. Few mathematical transformations are popular in HSI for feature selection or dimension reduction like principal component analysis (PCA) [20], independent component analysis (ICA) [21], wavelet transform (WT) and minimum noise fraction (MNF). Among those techniques, PCA is the most commonly used method which select the spectral bands after a transformation based on variance of data. The work [13] used randomized PCA to select principal components (PCs) containing 99% variance along spectral dimension and gain

[1]This work was supported in part by Natural Science Foundation of China (61420106007 and 61671387).

very promising accuracy with quite shallow CNN. On the other hand keeping the complexity issue aside, the classes are well separated by the ICA based feature selection and can contribute well for classification task. Recently, a discrete cosine transform (DCT) based pre-processing with ICA for HSI classification [22] was performed with better results. The DCT is widely used in image compression and video compression especially because of its ability to transform the data from a wide form to a compact form it is also suitable for hyperspectral data compression [23].

The end-to-end deep learning models are very famous for HSI classification because of the automatic feature engineering aspect. However these approaches employ big data for training and can cause accuracy declination otherwise. On the other side, apart from few advantages it is really challenging to find suitable and appropriate features using traditional methods (like PCA, ICA, etc.). In our work we put together traditional methods (ICA, PCA and DCT) and a shallow deep learning model in a way to exploit the advantages of both for the HSI classification task. The major contributions of our algorithm IPDCT-CNN are as follows:

1) The DCT compaction property is effectively used as a preliminary step before PCA, taking advantage of getting more variance of data in comparatively less number of spectral bands.
2) The novel approach of using all the three ICA, PCA and DCT in such a manner to get compact, distinctive and well separated features in pre-processing step. Which delivers best accuracy in comparison to use PCA or ICA alone as a feature selection transformation.
3) Our proposed simplified CNN which automatically and hierarchically extract spectral-spatial features for final classification using the IPDCT data.

## II. Mathematical Transformations

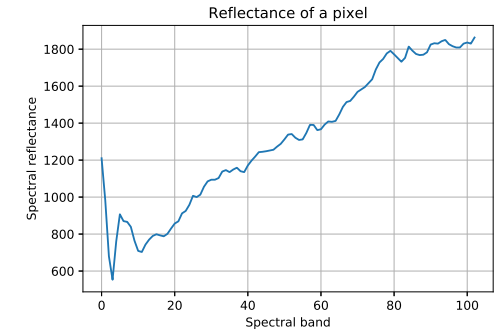
### A. DCT

The "energy compaction" eminence of DCT tends the transformed signal to be compacted in a few low-frequency components of the DCT. This compaction property can be suitable for HSI data especially along spectral dimension as suggested in the work [22]. Suppose there are $L$ spectral bands of HSI data, then for any pixel "$x$" we can say $x_1, x_2, ..., x_{L-1}$ are $L$ real values for reflectance and after the transformation the respective DCT coefficients $X_1, X_2, ..., X_{L-1}$ can be computed using the Eq. (1).

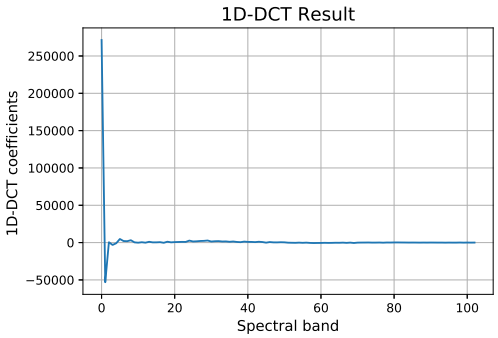
$$X_d = \sum_{n=0}^{L-1} x_n \cos\left(\frac{\pi}{L}\left(n+\frac{1}{2}\right)d\right) \tag{1}$$

where $d = 0, 1, ..., L-1$

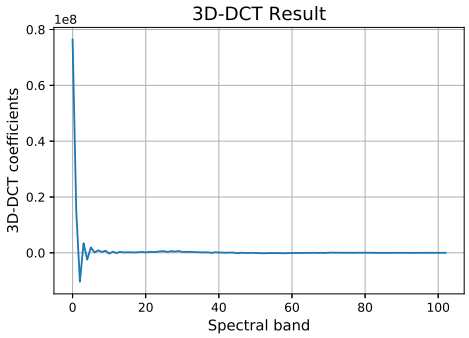
Moreover [23] performed 3D-DCT by repeating 1D-DCT three times and used it for HSI compression with discrete wavelet transform (DWT). Inspired with these two works we performed 3D-DCT along spectral dimension by repeating 1D-DCT three times using the Eq. (2).




(a) Reflectance of a pixel




(b) 1D-DCT of a HSI pixel




(c) 3D-DCT of a HSI pixel

Fig. 1: Pixel reflectance, 1D-DCT and 3D-DCT plots

$$X_{efg} = x_m x_n x_p \sum_{m=1}^{L-1}\sum_{n=1}^{L-1}\sum_{p=1}^{L-1} \cos\left(\frac{\pi}{L}(m+1/2)e\right) \cos\left(\frac{\pi}{L}(n+1/2)f\right)\cos\left(\frac{\pi}{L}(p+1/2)g\right) \tag{2}$$

The Figure 1 shows the HSI data pixel without DCT, with 1D-DCT and 3D-DCT. Although both 1D and 3D DCT looks quite similar however 3D-DCT provide further energy compaction.

### B. PCA

The transformation is performed by recognizing the principal directions called principal components (PCs) in which the

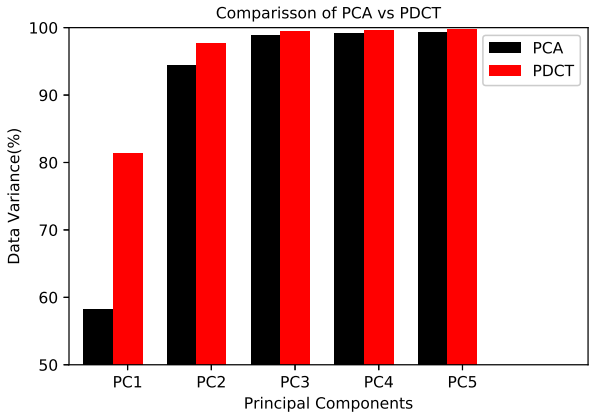



Fig. 2: Comparison of PCA vs PDCT

data varies the most and transforms original data dimension to a sub-space comparatively of low dimension. As the HSI spectral data is highly correlated so PCA is most suitable transformation for dimension reduction. Normally PCA is used alone for feature selection before a CNN based classifier like the work in [13], [24] by either selecting few top PCs or selecting PCs on the basis of certain percentage of data variance (99% or more) along spectral dimension. However we evaluated that if we perform suggested 3D-DCT prior to PCA (called it as PDCT) then it contains more variance of data in less no of PCs in comparison to the direct PCA based transformation. So this can further reduce the spectral dimension and select most pertinent features to tackle classification task. The Figure 2 shows the fair comparison between PCA and PDCT for first five PCs for university of Pavia dataset and it is clearly shown that PDCT contain more variance of data in comparison to PCA based method.

### C. ICA

The key approach in ICA is that the data is a linear mixture of separate and independent sources. The mixed data can be transformed into separate signal sources on the basis of their statistical distinctive properties. The ICA is widely used in HSI analysis in the area of spectral un-mixing in particular, for target detection [25] and as a pre-processing step [22] for HSI classification. The ICA transformation is established on a non-Gaussian postulation of independence between sources. Suppose we have an observation vector $y = [y_0, y_1, ..., y_{R-1}]$, which is linear mixture of $R$ independent elements of a random vector source $s = [s_0, s_1, ..., s_{R-1}]$. In matrix form the model will be:

$$Y = A.S \tag{3}$$

where $A$ indicates the mixing matrix.

So the ICA transformation estimates a matrix $W$ (i.e., the inverse of mixing matrix $A$) to calculate the best possible assumption of $S$.

$$Z = W.Y \approx S \tag{4}$$

TABLE I: CNN configuration

| Layer No | Input | Configuration | Filters/Units | Output |
|---|---|---|---|---|
| 1 | 7x7 | Conv(P)+BN+Relu | 5x5, 20 | 7x7x20 |
| 2 | 7x7x20 | Conv+BN+Relu | 5x5, 60 | 3x3x60 |
| 3 | | Dropout 50% | | |
| 4 | 3x3x60 | Conv+BN+Relu | 3x3, 100 | 1x1x100 |
| 5 | | Flatten | | |
| 6 | 1x1x100 | Fully Connected | 200 | 200 |
| 7 | | Dropout 50% | | |
| 8 | 200 | Softmax | N/A | 16/9 |

Normally PCA is a correlation based transformation however ICA not only de-correlates sources but also makes the signals independent which is pretty helpful for classification. It precisely identifies patterns and reduces noise from data effectively which makes it ideal for dimension reduction. Thus considering these properties we performed ICA on PDCT data to further process the reduced data in a way the classes are more separable retaining the rich patterns and features for further stages of classification. We referred this algorithm as IPDCT in the rest of the paper.

## III. Proposed Approach

### A. Dimension Reduction

The HSI data cube of dimension $L \times W \times B$ where "$L$" is height, "$W$"is width and "$B$" is number of spectral bands as shown in Figure 3. The raw data cube is divided into two sub cubes one is PCA cube and other is PDCT cube as explained earlier. Each cube having "$N$" number of PCs or spectral bands on the basis of better classification results. In the next step both cubes (each of dimension $L \times W \times N$) merged and ICA is performed so that IPDCT data cube of dimension $L \times W \times 2N$ contain reduced and more discriminate and separated data for classification. The spatial-spectral properties of adjacent pixels are quite similar and highly correlated to each other. So dividing the IPDCT cube (dimension $L \times W \times 2N$) into square patches of size $P \times P \times 2N$ to exploit the neighboring pixels correlation. In this way the whole IPDCT cube will be divided into $K$ patches each of dimension $P \times P \times 2N$. Suppose we have $G$ number of labeled pixels so considering only those patches centered with labeled pixels and discarding the rest of the patches. If the HSI raw cube shows the land cover map of "$F$" number of classes and $C$ is the matrix $C = C_1, C_2, \cdots, C_F$ of that class labels. Then the $G$ number of patches will be centered by any one of the class label of set $C$. For further process each patch will be taken such that the whole patch belong so same class label as center pixel of the patch.

### B. CNN Framework

The Figure 4 shows a very simplified and shallow IPDCT-CNN used for feature extraction and classification. The CNN mainly consists of three convolution layers (CL), one fully connected layer (FC) and a softmax classifier. We used the combined batch normalization (BN) and dropout as suggested in [14] as it gives much better performance even with higher

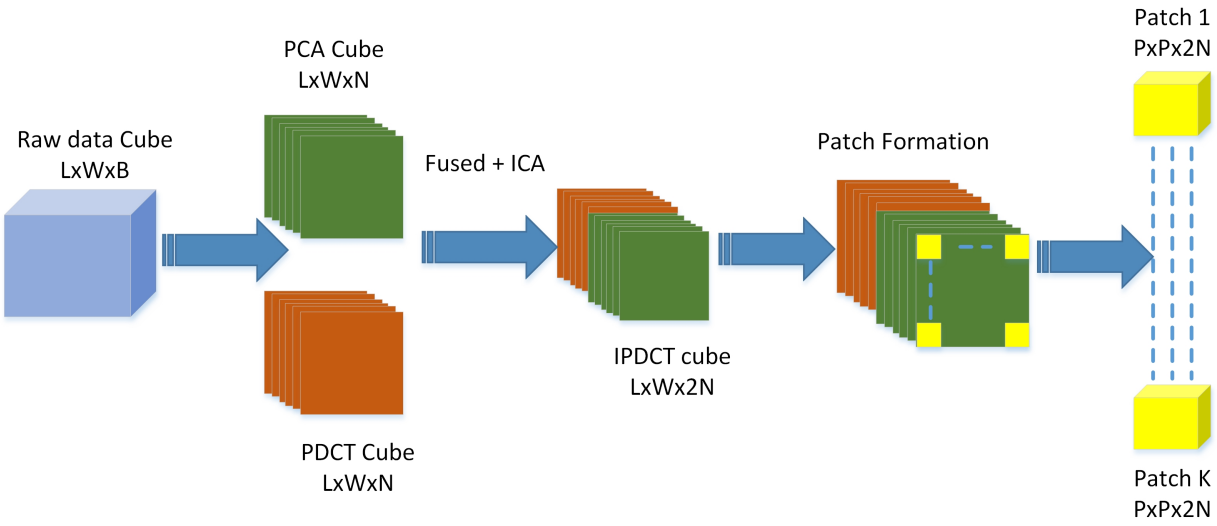



Fig. 3: IPDCT based feature selection/dimension reduction part

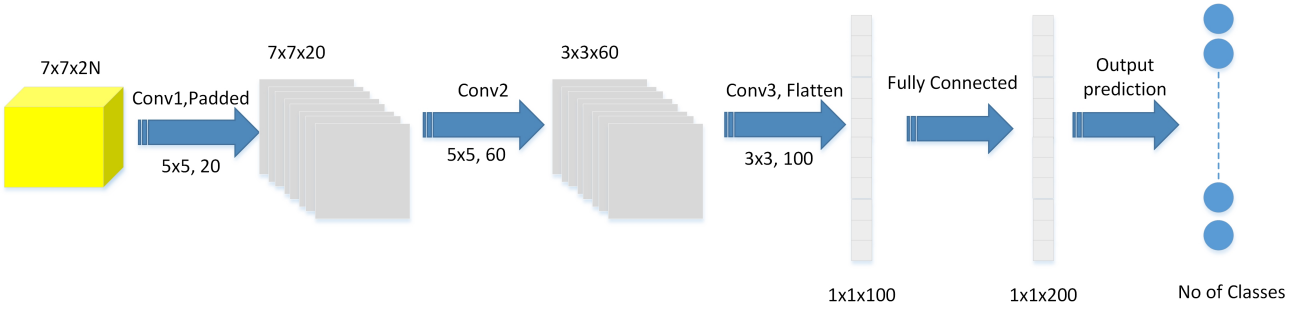



Fig. 4: Main configuration of CNN

learning rates. In our approach we have not used any pooling layer (as we ignore scale invariance and translation). We ignore the borders during convolution operation except the first convolutional layer where padding is applied to preserve the input dimension for further layers. The whole configuration of our CNN is explained in Table I. The selection of filter size and numbers are on the basis of performance and result will be shown in next section. Moreover, every patch of dimension $P \times P \times 2N$ is divided into $2N$ matrices of dimensions $P \times P$ which are given as input to our CNN which outs suitable high level features encoding the spectral-spatial individualities of pixels. Finally these feature maps help for classification after a training process.

TABLE II: Patch Size vs Accuracy

| Patch Size | IN | UP |
|---|---|---|
| 3x3 | 88.40 | 88.27 |
| 5x5 | 96.66 | 98.81 |
| 7x7 | 98.53 | 99.12 |
| 9x9 | 99.05 | 99.55 |
| 11x11 | 99.20 | 99.67 |

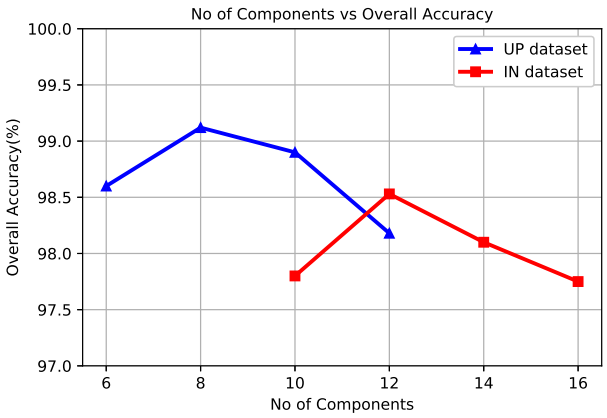



Fig. 5: Number of components vs overall accuracy

## IV. Experimental Settings

### A. Datasets and Parameters Setting

We used two benchmark datasets Indian Pines (IN) and University of Pavia (UP) to authenticate our approach. The UP dataset is collected by ROSIS sensor in northern part of

TABLE III: Effect of first CL filter and number of filters

| | **First CL filter size** | | | **Number of Filters(R)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **3x3** | **5x5** | **7x7** | **14** | **16** | **18** | **20** | **22** | **24** |
| **Indian Pines** | 98.118 | **98.53** | 97.67 | 97.35 | 98.06 | 97.72 | **98.53** | 98.22 | 98.24 |
| **University of Pavia** | 98.82 | **99.12** | 98.99 | 98.99 | 98.95 | 98.72 | **99.12** | 98.91 | 98.96 |

TABLE IV: Overall accuracy comparison with other methods

| | **CNN-PPF** | **DCPN** | **P-CNN** | **IDCT-CNN** | **IPDCT-CNN** |
|---|---|---|---|---|---|
| **Indian Pines** | 94.43 | 97.10 | 97.33 | 97.43 | **98.53** |
| **University of Pavia** | 96.48 | 98.51 | 98.77 | 98.87 | **99.12** |

Italy and it includes $610 \times 340$ pixels (after neglecting the no information part of image) having geometric resolution of 1.3 meters. They contain 103 spectral bands with the range of 430 to 860 nm and 9 classes. The IN dataset is achieved by AVIRIS sensor in north western part of Indiana and it comprises of $145 \times 145$ pixels with geometric resolution of 20 meters. The spectral data is divided into 200 bands (by discarding the bands with water absorption) within the range of 400 to 2500 nm. It has in total 16 vegetation land cover classes.

Training epochs are set as 500 and 1200 for IN and UP datasets respectively. Moreover 30% of labeled patches are randomly selected for training and 70% for testing for both the datasets. In order to further create some disparity in training samples, the data augmentation on 50% of training samples by using flip up, flip down and random rotation of patches. The back propagation algorithm is selected for training the network and the batch size is chosen as 32 for each training epoch. The network parameters are updated by minimizing the cross entropy loss function. The model with all the mentioned setting is trained for both the datasets for maximum number of epochs. Once the training is finished, the trained model is saved and test data is evaluated considering the already saved model and assessment is made on the basis of overall accuracy of testing samples. First of all following parameters of IPDCT-CNN are chosen on the basis of experiments.

*1) Patch size:* In order to evaluate the effect of the patch size $(P \times P)$, we verified the proposed models with different input patches of spatial sizes. Table II shows that the proposed IPDCT perform strongly for different spatial sizes and hierarchically conveys discriminative features. In both the datasets the accuracy increased with effective increase in the patch size keeping in mind that the more abstract features by including the neighboring pixels role in determining the features. However in order to select a suitable patch size we decided 7x7 patch size as it is most commonly used size for patch based schemes and also have effective classification results. Although larger patch size enhance accuracy a bit but create more complexity in terms of parameters and filter sizes.

*2) No of Components:* In the dimension reduction section the number of components (spectral bands) is very important such that it gives better dimension reduction with maximum information intact. In this regard experiments were conducted on IN and UP data set to choose the value of $N$. We select 6, 8, 10 and 12 for UP dataset and 10, 12, 14 and 16 for IN data set. On the basis of overall accuracy we select 8 and 12 components for UP and IN dataset respectively. The Figure 5 shows the comparison for number of components with the overall accuracy for both the selected datasets.

*3) Filters for CNN:* The selection for the size and number of filters for a CNN is very tricky as it is always a tradeoff between accuracy and complexity. So at first we evaluated the size of filter for only first CL among 3x3, 5x5 and 7x7 in terms of better accuracy and we found that for both the datasets 5x5 filter size gives best accuracy. Secondly, we conducted experiments for number of filters in our CNN, for simplicity we set number of filters in such a way that the first CL has "$R$" number of filters 2nd CL, 3rd CL and FC layer has $3 \times R$, $5 \times R$ and $10 \times R$ filters respectively. So by only changing the value of $R$ we can change the number of filters in all layers. We conducted experiments for different values of $R$ between 14 and 24 found that for both datasets the best results are achieved when $R$ is set as 20. The results are shown in Table III.

## V. Result Comparison

The usefulness of our method IPDCT-CNN is compared with other methods like CNN-PPF [26], DCPN [27]. Moreover we compare our result by performing only PCA before CNN (called P-CNN) and in other test performing DCT and ICA (IDCT-CNN) before CNN while retaining the rest of the parameters as in IPDCT-CNN. In all the comparisons the overall accuracy (OA) of classification for two datasets are matched.

The results are stated in the Table IV and it is quite clear that the best accuracy is achieved by IPDCT-CNN method. Thus we can conclude that the proposed method is quite considerate in selection of the most suitable, distinguished and well disjointed features prior to CNN due to the effective and unique combination of ICA, PCA and DCT as a pre-processing step. Furthermore the use of relatively simpler CNN helps in automatic and hierarchical extraction of the features for classification of the required classes.

## VI. Conclusion

In this paper a novel and quite effective joint ICA-PCA-DCT pre-processing approach for CNN based HSI classification is presented. The approach primarily based on successfully exploiting the advantages of DCT for compaction of spectral data of HSI which thus helpful for PCA based transformation in terms of better and slightly more dimension reduction. The further use of ICA not only make the feature more effective but also well separated in terms of class significance. Finally the CNN consists of 3 convolution layers and single fully connected layer effectively demonstrated the classification task on already pre-processed features. The results revealed that the advocated method of pre-processing along with CNN can bring a noticeable improvement in the classification as compared to the standard pre-processing transformations like PCA and ICA individually.

## References


[1] F. Martinelli, R. Scalenghe, S. Davino, S. Panno, G. Scuderi, P. Ruisi, P. Villa, D. Stroppiana, and M. Boschetti, "Advanced Methods of Plant Disease Detection. A review," *Agronomy for Sustainable Development*, vol. 35, no. 1, pp. 1–25, 2015.

[2] B. Uzkent, M. J. Hoffman, and A. Vodacek, "Real-time Vehicle Tracking in Aerial Video using Hyperspectral Features," 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 26-July 01, 2016.

[3] S. Suktanarak and S. Teerachaichayut, "Non-Destructive Quality Assessment of Hens' Eggs using Hyperspectral Images," *Journal of Food Engineering*, vol. 215, pp. 97–103, 2017.

[4] Z. Liu, H. Wang, and Q. Li, "Tongue Tumor Detection in Medical Hyperspectral Images," *Sensors*, vol. 12, no. 1, pp. 162–174, 2012.

[5] H. Yang, F. Meer, W. Bakker, and Z. Tan, "A Back-Propagation Neural Network for Mineralogical Mapping from AVIRIS Data," *International Journal of Remote Sensing*, vol. 20, no. 1, pp. 97–110, 1999.

[6] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[7] M. Fauvel, Y. Tarabalka, J. Benediktsson, J. Chanussot, and J. Tilton, "Advances in Spectral-Spatial Classification of Hyperspectral Images," *Proceedings of IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[8] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in Hyperspectral Image Classification," *IEEE Signal Processing Magazine*, vol. 31, pp. 45–54, Jan 2014.

[9] A. R. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, "Deep Belief Networks using Discriminative Features for Phone Recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague Congress Ctr, Prague, Czech Republic, May 22-27, 2011.

[10] R. Collobert and J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning," *Proceedings of the International Conference on Machine Learning, Helsinki, Finland*, 2008.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," International conference on Neural Information processing Systems (NIPS), Lake Tahoe, Nevada, United States, Dec 3-6, 2012.

[12] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep Convolutional Neural Networks for Hyperspectral Image Classification," *Journal of Sensors*, vol. 258619, 2015.

[13] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep Supervised Learning for Hyperspectral Data Classification Through Convolutional Neural Networks," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, July 26-31, 2015.

[14] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pp. 847–858, Feb 2018.

[15] Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-Spatial Classification of Hyperspectral Image Using Autoencoders," 9th International Conference on Information, Communications and Signal Processing (ICICS), Tainan, Taiwan, Nov 10-13, 2013.

[16] Y. Chen, X. Zhao, and X. Jia, "Spectral Spatial Classification of Hyperspectral Data Based on Deep Belief Network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.

[17] M. He, X. Li, Y. Zhang, J. Zhang, and W. Wang, "Hyperspectral Image Classification Based On Deep Stacking Network," 36th IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, July 10-15, 2016.

[18] M. He, B. Li, and H. Chen, "Multi-Scale 3D Deep Convolutional Neural Network For Hyperspectral Image Classification," 24th IEEE International Conference on Image Processing (ICIP), Beijing, China, Sep 17-20, 2017.

[19] D. L. Donoho, I. Johnstone, B. Stine, and G. Piatetsky-shapiro, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality," *Proceedings of AMS Math challenges*, pp. 1–33, 2000.

[20] V. Zubko, Y. J. Kaufman, R. I. Burg, and J. V. Martins, "Principal Component Analysis of Remote Sensing of Aerosols Over Oceans," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 730–745, 2007.

[21] A. Villa, J. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral Image Classification With Independent Component Discriminant Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4865–4876, 2011.

[22] K. Boukhechba, H. Wu, and R. Bazine, "DCT-Based Preprocessing Approach for ICA in Hyperspectral Data Analysis," *Sensors*, vol. 18, no. 4, 2018.

[23] S. Kala and D. Vasuki, "FPGA Based Hyperspectral Image Compression Using DWT and DCT," *Australian Journal of Basic and Applied Sciences*, vol. 8, no. 7, pp. 81–91, 2014.

[24] W. Zhao and S. Du, "Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

[25] E. Fakiris, G. Papatheodorou, M. Geraga, and G. Ferentinos, "An Automatic Target Detection Algorithm for Swath Sonar Backscatter An Automatic Target Detection Algorithm for Swath Sonar Backscatter Imagery , Using Image Texture and Independent Component Analysis," *Remote Sensing*, vol. 8, no. 5, 2016.

[26] W. Li, G. Wu, and Q. Zhang, F.and Du, "Hyperspectral Image Classification using Deep Pixel-Pair Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 844–853, 2017.

[27] L. Shu, K. McIsaac, and G. Osinski, "Hyperspectral Image Classification with Stacking Spectral Patches and Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5975–5984, 2018.