

# Hybrid 3DNet: Hyperspectral Image Classification with Spectral-spatial Dimension Reduction using 3D CNN

Zareen Binta Zakaria

Department of Computer Science and Engineering  
Hajee Mohammad Danesh Science and Technology  
University, Bangladesh

Md. Rashedul Islam

Department of Computer Science and Engineering  
Hajee Mohammad Danesh Science and Technology  
University, Bangladesh

## ABSTRACT

Hyperspectral image classification (HSI) is a fantastic approach for assessing diverse land cover utilizing remotely sensed hyperspectral images and has been an established research topic. The term classification is used in remote sensing to refer to the process of assigning individual pixels to a group of classes. The utilization of CNN for HSI classification is likewise noticeable in ongoing works. These approaches are generally founded on 2-D CNN. For practical purposes, a 2D Convolutional Neural Network (CNN) is a viable option; however, these models do not provide high-quality feature maps because a 3D data cube, a Hyperspectral image, contains both two-dimensional spatial information (image feature) and one-dimensional spectral information (spectral-bands). Therefore, 3D CNN can be another option, yet it has high computational complexity because of the volume and spectral dimensions. This paper proposed a 3D CNN model that achieves excellent results by combining spatial and spectral feature maps. The performance of our proposed method is approved using three standard HSI datasets (Pavia University, Indian Pines, and Salinas), and the outcomes are further compared with several state-of-the-art methods.

## Keywords

3D Convolutional Neural Network (CNN); Dimension Reduction; Hyperspectral Images (HSI); HSI Classification;

## 1. INTRODUCTION

Hyperspectral images are made up of hundreds of spectral bands, each of which holds comprehensive spectral information. Hyperspectral sensors and imaging spectrometers have made hyperspectral imaging systems one of the most important remote sensing technologies, and they are now widely utilized in urban mapping, military operations, agriculture, geology, mineral detection, and many other fields [1]. A hyperspectral image (HSI) is multidimensional and has a huge set of information that can be used to identify special features in a scene. As a result, hyperspectral sensors are being put on a growing number of satellites, allowing them to produce imagery with extensive spectral information.

There are two ways of classifying hyperspectral image (HSI): One is with a handcrafted feature extraction method and another with a learning-based feature extraction method. In the past time, several HSI classification techniques have been developed. Such as, Qian and Yang developed Hyperspectral Image Classification by Multiscale Joint Collaborative Representation and Locally Adaptive Dictionary [2]. That locally adaptive dictionary is used to reduce the influence of irrelevant pixels on representation, and The MLJCR method is used for classifying a hyperspectral image. Fang [3] has proposed a new Spatial-Spectral feature

extraction Method and has utilized the local covariance matrix to develop the relationship between different spectral bands. They classified using a support vector machine and used those covariance matrices to train HSI. For combining spatial and spectral information, a composite kernel is used for HSI classification [4]. Jun Li and Xin Huang [5] have developed a model for multiple feature learning for the classification of hyperspectral images. Some other approaches of handcrafted are a joint sparsity model by Qishuo Gao [6], 3D discrete cosine transform by Shrutika & Prabukumar [7], entropy-based classification for hyperspectral band selection, and Spatial residual clustering by P. Gao, J. Wang [8], multiple scales and superpixels with guided filter [9], etc. The convolutional neural network (CNN) has gained popularity as a state-learning-based feature extraction method due to its significant improvement over hand-designed techniques.

Convolution neural networks (CNNs) are widely used in image identification and segmentation that considers pixels spatial correlation and have a uniquely built deep learning architecture. Successful examples of CNN's include AlexNet: ImageNet Classification with Deep Convolutional Neural Networks [10], VGG [11], InceptionNet/GoogleNet [12], ResNet: Deep Residual Learning for Image Recognition [13] and DenseNet: Densely Connected Convolutional Networks [14]. Existing CNNs, on the other hand, are used for conventional image classification tasks instead of hyperspectral image classification tasks, which require efficient subjugation of both spatial and spectral correspondences. Very auspicious performance has been shown where visual information processing is required by the CNN in many applications, such as image classification [15], [16], semantic segmentation [17], object recognition [18], face anti-spoofing [19], colon cancer cataloging [20], depth guesstimate [21] and so on. This study demonstrates the huge improvements being made in deep learning for HSI analysis.

Therefore, this paper proposes a Hybrid3DNet classification method for Hyperspectral Image (HSI). This procedure first divides the HSI data cube into small 3D patches that overlap. These patches are used to construct 3D feature maps over various adjacent bands using the 3D kernel function, preserving joint spatial and spectral information for the feature learning process and exploiting important discrimination information for HSI classification. Factor Analysis (FA) is used in the preprocessing step to eliminate redundant bands and extract the few relevant bands from the entire HSI data cube. Later the model is trained by the 3D CNN classifier in an end-to-end fashion, which contains a smaller amount of parameters than other 2D or 3D CNN models. Finally, the experimental results of this comparative study demonstrated that the proposed method outperforms existing state-of-the-art 2D or 3D CNN-based HSI classification methods presented in the literature.

The other sections of the paper are divided as follows: the proposed methodology is presented in Section 2. The experimental datasets, results, and discussion are depicted in Section 3. Finally, Section 4 concludes the work by outlining possible future research directions.

## 2. PROPOSED METHOD

### 2.1 Factor analysis (FA)

Factor analysis (FA) consists of procedures describing each variable's contribution (band) from the exploration of covariance. The underlying structure of large variables (bands) is examined by Factor analysis. Therefore, it is also a statistical technique that determines variability in terms of fewer unobserved variables (band) among observed variables (band) and interprets using factor scores or components. This approach aims to redirect the variables (band) so that many core variables (band) can be comparatively compressed with a few factors that capture the most potential data variation from the original dataset. So, Factor analysis calculates the amount of variability in the data due to common factors [22].

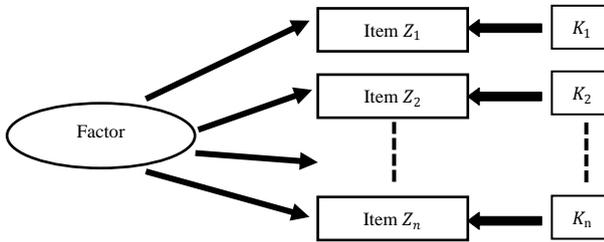


Fig. 1. Latent factor drive from observed variables (band).

Figure 1 shows  $Z_n = [Z_1, Z_2, Z_3, \dots, Z_n]$  is a set of observable random variables with the mean vector  $\mu = \mu_1, \mu_1, \dots, \mu_n$ . The factor analysis equation assumes as follows:

$$Z = \mu + \lambda K + e$$

Where  $\lambda$  is denoted as the matrix of factor loadings with a vector of latent factor scores  $K_n = [K_1, K_2, K_3, \dots, K_n]$  and  $e_n = [e_1, e_2, e_3, \dots, e_n]$  denotes the vector of latent error terms. Under the factor analysis process, the covariance matrix of the observable random variable  $Z$  is estimated as follows:

$$Cov(Z) = \lambda \lambda' + \varphi$$

Here,  $\varphi$  is a diagonal matrix. The  $k^{\text{th}}$  diagonal element of  $\lambda \lambda'$ , the addition of the values of the squared loading, is called the  $k^{\text{th}}$  commonality, which shows the percent of variability explained by the common factors. The  $k^{\text{th}}$  diagonal element of  $\varphi$  is called the  $k^{\text{th}}$  specific variance.

### 2.2 Hybrid 3DNet

Let's consider a hyperspectral data cube can be signified by  $I \in R^{L \times S \times B}$ , where  $I$  is the original input,  $L$  is the Height,  $S$  is the

Width and  $B$  are the numbers of spectral bands/distance. For all the pixels of the input image, data cube  $I$ , has spectral measures denoted by  $B$  that form a one-hot label vector  $Y = (Y_1, Y_2, Y_C) \in R^{1 \times 1 \times C}$ , where  $C$  represents the land-cover categories. However, the pixels of the hyperspectral image

exhibit the mixed land-cover classes and introducing the high interclass variability and interclass similarity into  $I$ . Any classification model faces the challenge to get rid of this problem. To overcome the aforementioned issue, we remove the spectral redundancy first. For this, Factor analysis (FA), which was previously mentioned, the original HSI data ( $I$ ) and spectral bands had gone through this dimension reduction procedure. For maintaining the same spatial dimensions (i.e., width  $L$  and height  $S$ ), the number of spectral bands is reduced from  $B$  to  $M$ . This procedure has reduced only spectral bands and preserves spatial information which is very important for recognizing an object. The FA reduced data cube is denoted by  $X \in R^{L \times S \times M}$ , where  $X$  is the modified input after FA,  $L$  is the width,  $S$  is the height, and  $M$  is the number of spectral bands after FA.

Table 1. Layer-wise summary of the proposed Hybrid3Dnet architecture. With window size  $25 \times 25$ . The last layer is based on the PU data set

Layer (type)	Output Shape	No of Param #
input_1 (Input Layer)	(None, 25, 25, 7, 1)	0
conv3d (Conv3D)	(None, 23, 23, 3, 8)	368
conv3d_1 (Conv3D)	(None, 21, 21, 1, 16)	3472
conv3d_2 (Conv3D)	(None, 19, 19, 1, 32)	4640
conv3d_3 (Conv3D)	(None, 17, 17, 1, 64)	18496
flatten (Flatten)	(None, 184960)	0
dense (Dense)	(None, 256)	4735232
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (dense)	(None, 16)	2064
In total, <b>9,528,937</b> trainable parameters are required		

For the application of image classification techniques, the HSI data cube was partitioned into small corresponding 3-D-patches whose truth positions are specified by the position of the central pixel. Here are the three-dimensional neighboring patches  $P \in R^{L \times S \times B}$  from  $X$ , with their centers located at the spatial coordinates  $(a, b)$  is covering the  $W \times W$  window or spatial region and all  $M$  spectral bands. The total number of three-dimensional patches ( $n$ ) is denoted by the expression  $(L - W + 1) \times (S - W + 1)$ . As a result, the 3-D-patch at  $(a, b)$ , indicated by  $P_{a,b}$  covers the width from  $a - (W - 1)/2$  to  $a + (W - 1)/2$ , height from  $b - (W - 1)/2$  to  $b + (W - 1)/2$ , and all  $M$  spectral bands of FA reduced data cube  $X$ .

The 3-D convolution method was carried out using a 3-D kernel and 3-D data [23]. Using the 3-D kernel throughout many neighboring bands in the input layer to build the feature maps of the convolution layer is our proposed model for HSI data. The spectral information and the activation values at spatial position  $(x, y, \text{ and } z)$  in the  $j^{\text{th}}$  feature map of  $i^{\text{th}}$  layer are captured and denoted as  $v_{ij}^{x,y,z}$ , is generated as follows:

$$v_{ij}^{x,y,z} = \mathcal{F} \left( \sum_{\tau=1}^{d_{i-1}} \sum_{\lambda=-\nu}^{\nu} \sum_{\rho=-\gamma}^{\gamma} \sum_{\phi=-\delta}^{\delta} \omega_{ij,\tau}^{v,\rho,\lambda} \times v_{(i-1),\tau}^{(x+\nu),(y+\rho),(z+\lambda)} + b_{ij} \right)$$

Here, the activation function is  $\mathcal{F}$ , the bias parameter for the  $j^{\text{th}}$  feature map of  $i^{\text{th}}$  layer is  $b_{ij}$ , the number of feature maps in

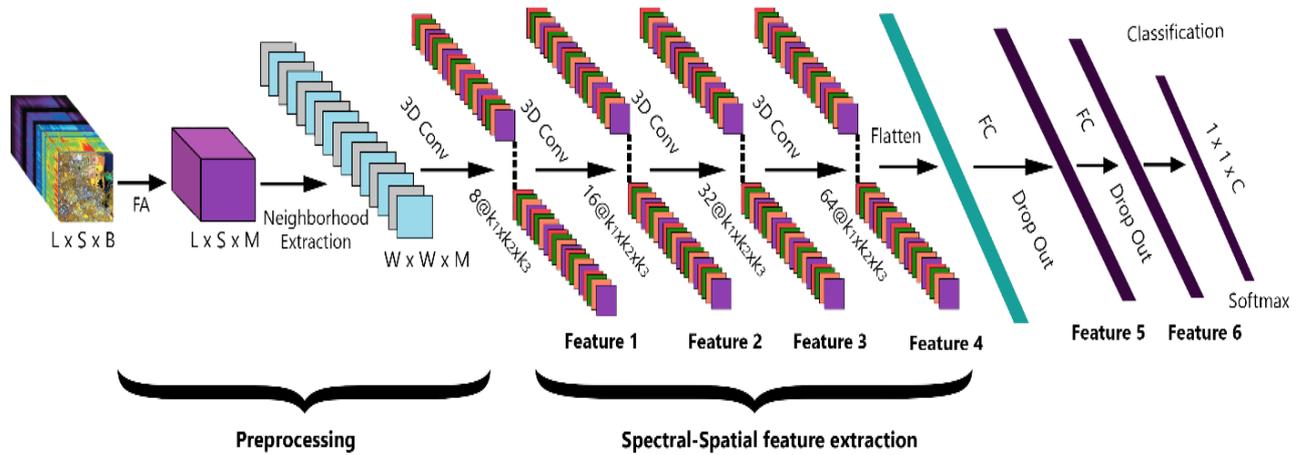


Fig. 2. Proposed Hybrid3DNet model for HSI classification

$(t-1)^{th}$  layer  $d_{t-1}$ , the depth of kernel for the  $j^{th}$  feature map of  $i^{th}$  layer is  $w_{i,j}$ , and width, height and depth of the kernel is  $2\gamma+1$ ,  $2\delta+1$ , and  $2v+1$  respectively.

In the Hybrid3DNet framework, the proposed 3D CNN convolutional [22] kernels are as follows: 3D\_conv\_layer1 =  $8 \times 3 \times 3 \times 7 \times 1$ , where  $\frac{1}{1}K=3, \frac{1}{2}K=3, \frac{1}{3}K=7$ . 3D\_conv\_layer2 =  $16 \times 3 \times 3 \times 5 \times 8$ , where  $\frac{1}{1}K=3, \frac{1}{2}K=3, \frac{1}{3}K=5$ . 3D\_conv\_layer3 =  $32 \times 3 \times 3 \times 3 \times 16$ , where  $\frac{1}{1}K=3, \frac{1}{2}K=3, \frac{1}{3}K=3$ . In order to ensure that the model is able to classify spatial information throughout different spectral bands without any loss, the model operate four 3D convolutional layers before the flatten layer to enhance the amount of spatial-spectral feature maps. Furthermore, detailed information about the proposed model is provided in Table 1 and the block diagram is illustrated in figure 2. The Hybrid3DNet model has 9,528,937 trainable parameters based on the last layer of the PU dataset. A softmax loss function combined with Adam analyzer back-propagation randomly distributes the initial weights. By utilizing a mini-batch size of 256, the weights are updated, and the network is trained for 200 epochs without any batch normalization or data augmentation.

### 3. EXPERIMENTS AND DISCUSSION

#### 3.1 Data set Description

Pavia University Dataset (PU) was collected using a Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor over Pavia, Northern Italy. With a spatial dimension of 1:3 meters, the PU dataset is related to  $610 \times 610$  spatial and 103 spectral bands. The PU dataset has a total of 9 ground truth classes.

An Infrared Imaging Spectrometer (AVIRIS) sensor was used to collect the Indian Pines Dataset (IP) over the test site in northwest Indiana. The IP data set consists of  $145 \times 145$  spatial dimensions and 224 spectral bands in the wavelength range of 400 to 2500 nm, 24 of which are excluded because they cover the water immersion region. The available data on the ground shows 16 different types of vegetation. More than a quarter of the entire coverage is focused on some crops that are still in

the early phases of growth. IP also owned low-density homes, buildings, tiny roads, a two-lane highway, and a railroad.

The Salinas dataset (SA) is obtained over Salinas Valley, California using the AVIRIS sensor. Vineyard fields, vegetables, and barren soils are among the photos in the SA data collection, which has  $512 \times 217$  spatial dimensions and 224 spectral bands in the wavelength range of 360–2500 nm SD. The 20 spectral bands that absorb water have been discarded. There are 16 classes in total in SA. Further details about the experimental datasets can be found in [24]. The ground images of all the experimental datasets are shown in Figure 3.

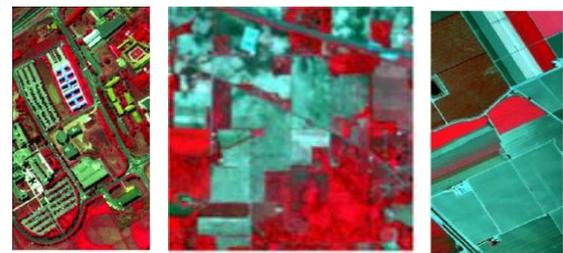


Fig3: Ground images of experimental datasets PU, IP and S Arespectively used in this work

#### 3.2 Experimental Configuration

All the experiments were conducted on Google Colab [25], an online platform. This online Platform requires a fast internet connection to run in any setting. The codes can be executed on a Python 3 notebook with a Graphical Processing Unit (GPU), 25 GB of RAM, and 358:27 GB of cold storage for data computation, according to Google Colab. The initial Test/Train set is divided into a 20-80% ratio in our experimental work.

#### 3.3 Classification Results

For evaluation purposes, the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) are used, which are computed from the confusion matrices to determine the

nondiscriminatory comparisons of the performance of the HSI classification. Here, OA is represented by the number of correctly classified samples out of the total test samples; the average of class wise classification accuracies represents AA, and Kappa delivers shared information regarding a solid agreement among the ground truth map and classification map, which is a metric of statistical dimension. The results of the proposed Hybrid3DNet model are compared with the most widely used supervised methods, such as SVM [26], 2D-CNN [27], and 3D-CNN [28]. The data are randomly divided into training and testing groups of 20% and 80% respectively. The publicly available code is used to perform the comparisons among the classification methods.

**Table 2. Classification accuracies (in percentages) using the proposed and State-of-the-Art methods**

Methods	University of Pavia Dataset			Indian Pines Dataset			Salinas Scene Dataset		
	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa	AA
SVM	82.77	74.21	75.45	78.58	68.99	72.98	90.75	87.23	90.69
2D-CNN	99.82	99.76	99.67	94.63	93.86	87.50	99.96	99.95	99.97
3D-CNN	99.93	99.91	99.91	98.93	98.79	98.82	99.97	99.97	99.99
Hybrid 3DNet	99.96	99.95	99.93	99.24	99.14	98.16	99.99	99.99	99.98

**Table 3. Training time in minutes (m) and test time in seconds (s) over IP, UP, and SA data sets using 2D-CNN, 3D CNN, and Hybrid3Dnet architectures**

Data	2DCNN		3DCNN		HYBRID3DNet	
	Train(m)	Test(s)	Train(m)	Test(s)	Train(m)	Test(s)
PU	1.98m	10.64s	11.00m	7.914s	4.37m	3.362S
IP	2.39m	5.818s	10.40m	6.32s	4.37m	2.404s
SA	1.00m	6.344s	10.38m	11.54s	4.37m	6.052s

**Table 4. Impact of the Spatial Window Size on the Performance of HYBRID3DNet**

Window	PU(%)	IP(%)	SA(%)	Window	PU(%)	IP(%)	SA(%)
19×19	99.94	99.09	99.74	23×23	99.95	99.23	99.99
21×21	99.95	99.04	99.92	25×25	99.84	98.16	99.98

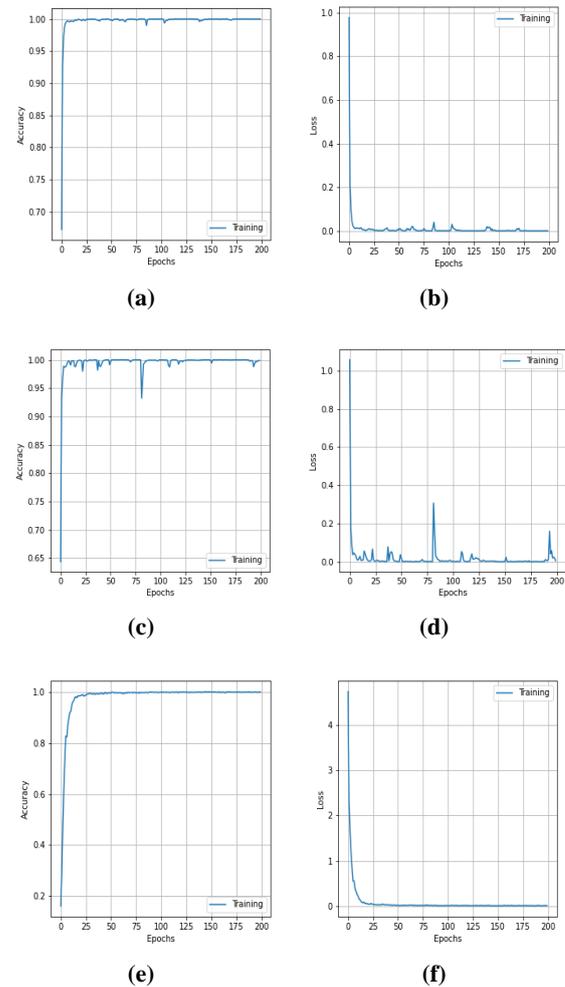
**Table 5. Classification accuracies (in percentages) using the proposed and State-of-the-Art methods on less amount of training data, i.e., 10% only**

Methods	University of Pavia Dataset			Indian Pines Dataset			Salinas Scene Dataset		
	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa	AA
SVM	82.39	73.48	74.16	75.74	64.88	70.74	90.24	86.56	88.54
2D-CNN	99.61	99.48	99.20	92.48	91.68	86.13	99.84	99.82	99.90
3D-CNN	99.74	99.65	99.43	96.96	96.53	96.29	99.96	99.95	99.97
Hybrid 3DNet	99.99	99.93	99.94	97.95	97.66	96.52	99.99	99.99	99.97

Table 2 illustrates the findings for different methods in terms of OA, AA, and Kappa and shows that the Hybrid3DNet outperforms all other approaches over each data set while keeping the standard deviation to a minimal. Based on the spectral-spatial 3D-CNN hierarchical representation and dimensionality reduction method FA, the Hybrid3DNet is proposed. These results also show that our proposed model outperforms 3D-CNN and 2D-CNN more handsomely when it comes to analyzing IP data. According to the prediction, this is due to specific classes in the Indian Pines data set that have highly comparable textures throughout most spectral bands

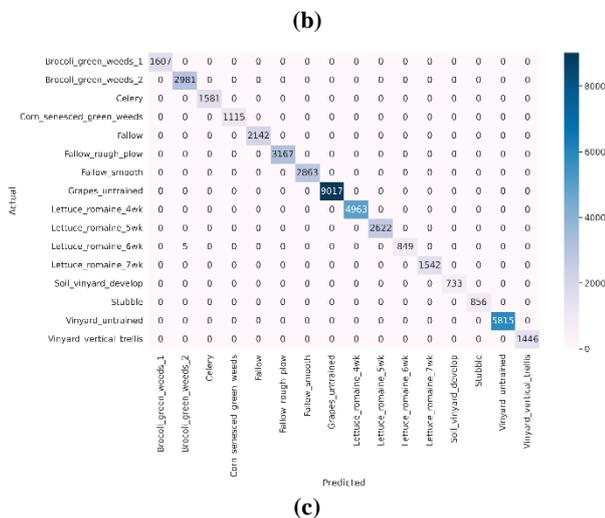
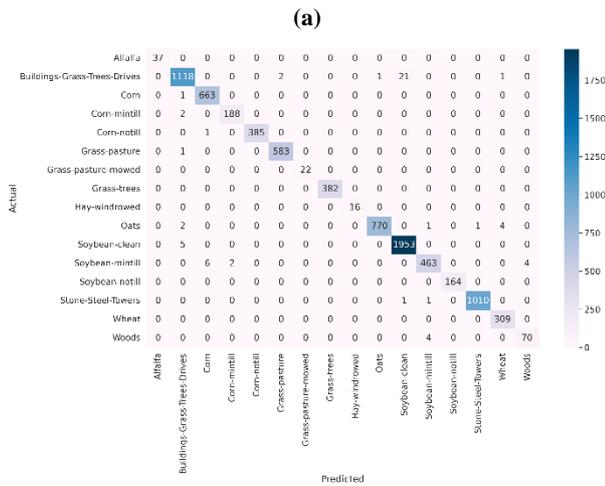
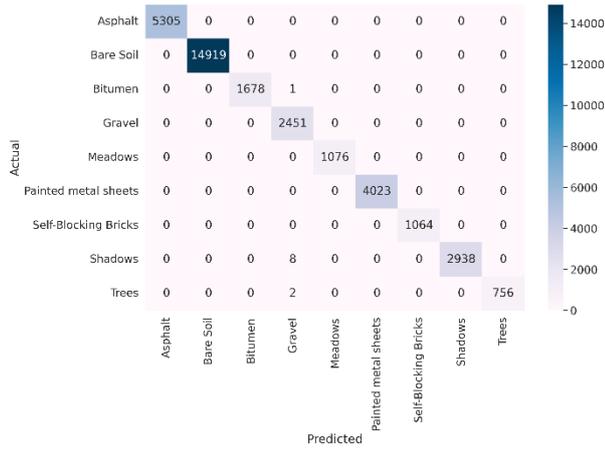
which is increased redundancy among the spectral bands. Hybrid3DNet, on the other hand, consistently outperforms SVM, 2D-CNN, and 3D-CNN in terms of performance. Compared to the proposed methodology, the highly discriminative feature cannot be represented in SVM, 3-D or 2-D convolution.

Table 3 shows up the computational effectiveness of the Hybrid3DNet model in terms of training and testing periods, and the proposed model is more efficient than 3-D-CNN and 2D-CNN. Furthermore, table 4 is an analysis of how the operation of the Hybrid3DNet model is affected by a variety of different spatial dimensions. It has been observed that, among the several spatial dimensions, 25×25 is the one that works best with the proposed method. The outcomes were obtained by using only 10% of the total training data that we have assembled here are presented in Table 5.



**Fig. 4. Accuracy and Loss convergence epochs over the PU dataset (a) and (b), SA dataset (c) and (d), IP dataset (e) and (f)**

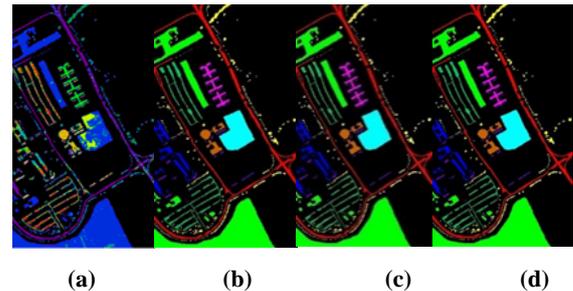
The proposed method's accuracy and loss convergence over 200 epochs of training and validation sets are represented in Figure 4. The confusion matrices of PU, IP, and SA datasets for the proposed methodology is shown in Figure 5. A confusion matrix is a table that is used to define the performance of a classification algorithm that visualizes and summarizes the performance of the classification algorithm also. The classification map for an example of HSI classification is presented in Figures 6, 7, and 8 and was generated with the



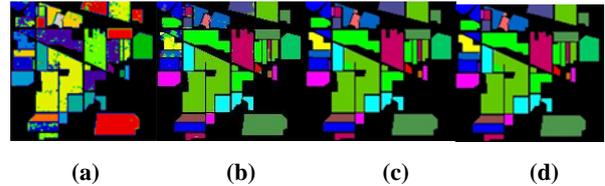
**Fig. 5.** Confusion matrix using proposed Hybrid3DNet method with actual and predicted classes over PU (a), IP (b), and SA (c) datasets in the first, second, and third matrices, respectively

supportvector machine (SVM), as well as the 2-D-CNN, 3-D-CNN, and Hybrid3DNet techniques for each of the dataset's PU, IP, and SA, in that order. In comparison to these previous methods, the quality of the classification map produced by Hybrid3DNet is significantly higher. As a result of this examination, it is clear that the performance level of each model declines slightly, but the Hybrid3DNet that was proposed is still able to outperform the others in almost all

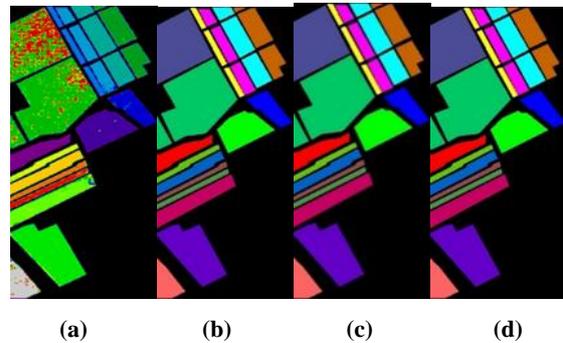
cases.



**Fig. 6.** Classification Map PU Dataset using (a) SVM, (b) 2DCNN, (c) 3DCNN, (d) Hybrid3DNet (FA+3DCNN).



**Fig. 7.** Classification Map IP Dataset using (a) SVM, (b) 2DCNN, (c) 3DCNN, (d) Hybrid3DNet (FA+3DCNN).



**Fig. 8.** Classification Map SA Dataset using (a) SVM, (b) 2DCNN, (c) 3DCNN, (d) Hybrid3DNet (FA+3DCNN).

#### 4. CONCLUSION

The proposed Hybrid3DNet, together with a dimension reduction method for HSI classification, has been presented in this paper. The complementary spatial and spectral information is merged in the form of 3D convolutions in the model, which is then followed by the dimension reduction method factor analysis (FA) and term as Hybrid3DNet. Experiments were run on three separate benchmark data sets, and the results were compared to the most recent state-of-the-art methodologies to validate that the proposed method is superior to those other options. In contrast to the 3D Convolutional Neural Network Model, the model that was proposed exhibited a computational efficiency that was noticeably superior. In addition to that, it exhibits outstanding performance despite having a limited amount of training data.

#### 5. ACKNOWLEDGMENTS

This project of this work is funded by IRT, HSTU, Dinajpur, Bangladesh.

#### 6. REFERENCES

- [1] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, vol. 1. Springer Science & Business Media, 2003.
- [2] J. Yang and J. Qian, "Hyperspectral image classification via multiscale joint collaborative representation with locally adaptive dictionary," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 112–116, Jan. 2018.

- [3] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
- [4] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [5] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.
- [6] Qishuo Gao & Samsung Lim (2019) A probabilistic fusion of a support vector machine and a joint sparsity model for hyperspectral imagery classification, *GIScience& Remote Sensing*, 56:8, 1129-1147, DOI: 10.1080/15481603.2019.1623003
- [7] Shrutika S. Sawant &Prabukumar Manoharan (2020) Unsupervised band selection based on weighted information entropy and 3D discrete cosine transform for hyperspectral image classification, *International Journal of Remote Sensing*, 41:10, 3948-3969, DOI: 10.1080/01431161.2019.1711242
- [8] P. Gao, J. Wang, H. Zhang, and Z. Li, "Boltzmann entropy-based unsupervised band selection for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 462–466, Mar. 2019.
- [9] Liu, Sicong& Hu, Qing & Tong, Xiaohua& Xia, Junshi& Du, Qian &Samat, Alim & Ma, Xiaolong. (2020). A Multiscale Superpixel-Guided Filter Feature Extraction and Selection Approach for Classification of Very-High-Resolution Remotely Sensed Imagery. *Remote Sensing*. 12. 862. 10.3390/rs12050862.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, arXiv: 1409.1556. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions," 2014, arXiv:1409.4842. [Online]. Available: <https://arxiv.org/abs/1409.4842>.
- [13] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," 2015, arXiv:1512.03385. [Online]. Available: <https://arxiv.org/abs/1512.03385>.
- [14] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks," 2016, arXiv:1608.06993. [Online]. Available: <https://arxiv.org/abs/1608.06993>.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [16] S. K. Roy, S. Manna, S. R. Dubey, and B. B. Chaudhuri, "LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," 2019, arXiv:1901.05894. [Online]. Available: <https://arxiv.org/abs/1901.05894>
- [17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [19] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Mar. 2019, pp. 1–8.
- [20] S. H. S. Basha, S. Ghosh, K. K. Babu, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "RCCNet: An efficient convolutional neural network for histological routine colon cancer nuclei classification," in *Proc. 15th Int. Conf. Control, Autom., Robot., Vis. (ICARCV)*, Nov. 2018, pp. 1222–1227.
- [21] V. K. Repala and S. R. Dubey, "Dual CNN models for unsupervised monocular depth estimation," 2018, arXiv: 1804.06324. [Online]. Available: <https://arxiv.org/abs/1804.06324>
- [22] Chakraborty, Tanmay, and Utkarsh Trehan. "Spectralnet: Exploring spatial-spectral waveletcnn for hyperspectral image classification." arXiv preprint arXiv:2104.00341 (2021).
- [23] Y. Li, H. Zhang, and Q. Shen, "Spectral spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, p. 67, 01 2017.
- [24] Hyperspectral Datasets Description, 2020 (accessed 2020-01-2), [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).
- [25] T. Carneiro, R. V. M. Da N'obrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. Reboucas Filho, "Performance analysis of google colabouratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61 677–61 685, 2018.
- [26] 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, Aug. 2004.
- [27] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.
- [28] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4420–4434, Aug. 2018.