Hyperspectral Image Classification using Hybrid Deep Convolutional Neural Network

Hyperspectral Image Classification using Hybrid Deep Convolutional Neural Network

Omprakash Nayak[1], Hrishikesh Khandare[2], Nikhil Kumar Parida[3], Ramnivas Giri[4], Rekh Ram Janghel[5] and Himanshu Govil[6]

National Institute of Technology, Raipur, Chhattisgarh, India

[1]omnayak27199@gmail.com ,[2]hrishikesh0408@gmail.com[3] , nkparida97@gmail.com, [4]ramnivasgiri7@gmail.com , [5]rrjanghel.it@nitrr.ac.in, [6]hgovil.geo@nitrr.ac.in

Abstract. The Hyperspectral Images (HSI) are now being widely popular due to the evolution of satellite imagery and camera technology. Remote sensing has also gained popularity and it is also closely related to HSI. HSI possesses a wide variety of spatial and spectral features. However, HSI also has a consider-able amount of useless or redundant data. This redundant data causes a lot of trouble during classifications as it possesses a huge range in contrast to RGB. Traditional classification techniques do not apply efficiently to HSI. Even if somehow the traditional techniques are applied to it, the results produced are inefficient and undesirable. The Convolutional Neural Network (CNN), which are widely famous for the classification of images, have their fair share of trouble when dealing with HSI. 2D CNNs is not very efficient and 3D CNNs increases the computational complexity. To overcome these issues a new hybrid CNN approach is used which uses sigmoid activation function at the output layer, using a 2D CNN with 3D CNN to generate the desired output. Here, we are using HSI classification using hybrid CNN i.e., 2D and 3D. The dataset used is the Indian pines dataset sigmoid classifier for classification. And we gain the Overall accuracy 99.34 %, average accuracy 99.27%, kappa 99.25%.

1. Introduction

The practical benefits of Hyperspectral images (HSI) analysis in real life make it very important topic of study [1]. HSI are made up of number of image bands. Which makes its analysis tricky because large amount of data being available. Since the late 1980s, spectral images with reflective wavelengths ranging from 400 to 2400 nm has been accessible, with information and data. HSIs can be used distance measuring in various research domains such as agriculture, monitoring of environment, forestry, mapping of land surface. HSIs can be utilized to their full potential in future research [1] by employing the classification method. Hyperspectral data is rich both in respect of spatial as well as spectral features. Simultaneously, correlation between the spectral and special information makes major part of this information redundant or useless. Techniques like band extraction and selection [2], multimodal learning model [3], and dimensionality reduction [4] can be used to create efficient representations. Existing traditional classification approaches are mainly based upon spectral resolution in the beginning stages, which usually consists of two primary components: feature engineering technique and classifier (SoftMax, Sigmoid) [5]. Feature engineering is used to create discriminant information or bands and decrease the HSIs Dimensions. Feature engineering consists of two common methods: first is Feature selection and the second one is feature extraction.[6]. Its primary goal is to convert the High -Dimension image to a low-Dimension image so that the categories may well be easily distinguished[7]. There are various methods of feature extraction such as so that the categories may well be easily distinguished[8-

10]. There are many feature extraction methods eg. independent component analysis [11] ,, Minimum noise fraction [12], linear discriminant analysis (LDA) [13], principal component analysis (PCA) [14] All are used for categorizing the features. The second technique of feature engineering is feature selection. Which is used to keep the spectral features of the most significant bands from the input HSIs while discarding the bands which has less contribution in HSI classification. Some general approaches for feature selection are spectral angle mapper [16], Jeffries–Matusita distance [15], and other feature selection approaches are common. The classifier uses features developed using feature engineering as input. Representative classifiers comprise different methods like support vector machine (SVM) [19], k-nearest neighbour [17], random forest [18] etc. The whole spatial information of HSIs signatures is not used in traditional classification algorithms based on spectral features, therefore They're all part of the same land cover types. Based on this hypothesis, approaches for extracting spectral-spatial features, like Gabor filtering [20] and wavelet modification [21], have been developed to improve class classification. Deep learning has recently emerged as a cutting-edge ML technology with immense potential for HSI classification [22]. It can learn the hierarchical feature from the given raw input image without any manual assistance. In any machine vision algorithms, the learned features have achieved the great success. The Existing classification approaches provide a wide range of classification solution which is based on the CNN-HIS classification framework [18]. the convolution operation is divided into three categories: 1D, 2D, and 3D CNN.
The 1-D CNN network architecture use the pixel vector with the spectral information of image as a input sample for training in order to select and extract the deep feature. It’s known as spectral-based classification technique. The usage of multilayer CNN for classification of HSI straight into the spectral domain was demonstrated for the very first time in [23]. To efficiently analyze the hyperspectral pixel as sequential data in order to extract the essential feature, a recurrent neural network was presented which employed a novel classifier to train the network model eliminating the risk of deviation [24].

The 2D CNN more commonly known as the spatial based classification attempts on learning features with the help of approaches similar to the traditional approaches used which were with the colorful patterns of RGB, which discovered the setback caused due to lack of attention towards the spectral-spatial attributes of spectral hypercube.

HSI originally being 3D hypercube which possess spatial and spectral continuity, the methods which made use of both spectral and spatial attribute have become popular. HSI CNN classifications can be dealt with the help of 3D convolutions easily, also known as spectral-spatial classification method. 3D areas with coupled spectral-spatial data can be processed at the same time using this approach. CNNs have an added benefit in comparison to other models. It is an popular model that used in the HSI classification due to its ability to its high ability to pick up spectral and spatial features. However, it still has flaws. For illustration, during the process of gradient descent, making the results converge to a local minimum is simple, However, the pooling layer ends up sacrificing ample amount of information which could have proven to be useful. We know that in the accuracy of classification of model the preprocessing stage plays a important role. PCA is a feature extraction technique among HSI models which removes non-linear features from an image. The task of picking up differentiating feature maps from spectral dimensions cannot be done by 2D CNN alone. On the other hand, a deep 3D CNN on its own is bad at predicting classes that have similarities in textures over may spectral bands apart from having computational complexity. However, 2D and 3D CNN clubbed together have a capability to achieve high accuracy.

In the next section 2 literature survey we will discuss about the previous methods and technique proposed by different authors in their papers. Furthermore, in section 3 proposed method for hybrid spectral CNN (HybridSN) using Indian pines Dataset is described. in section 4 we will analyze the result of our method and this paper will end in the section of conclusion.

2. Literature Survey

In [24] P.Ghamisi et al, suggested a Recurrent Neural Net- work was presented for efficient analysis of hyperspectral pixels as sequential data and subsequently identify feature categories with the help of network reasoning. In this paper, anew classifier PRetanh (parametric rectified tanh) was used for analysis of hyperspectral sequential data. The overall accuracy achieved by this approach for Indian Pines data set is 88.63%.

In [25] Hao et al, developed a model for classification of HIS images called SRCL model, in which a spatially enhanced image can be prepared by using a super-resolution-aided approach. This approach perform better when the input training data sample is small. The overall accuracy achieved in this research is 97.62%

In [26] F. Zhou et al, by using the posterior class distributions, the author proposes a CNN-Markov Random Field model to combine spectral and spatial resolution in a unified Bayesian framework. Overall accuracy of the proposed methodology is 99.32%

In [27] Chen et al.The author has proposed a deep feature extraction and classification techniques for HSI with the help of Convolutional neural network. Regularization and dropout techniques are also used for prevention of overfitting. The overall accuracy for the proposed method is 98.53%

In [28] Feng et al. The authors presented a multiclass spatial-spectral Generative adversarial networks (GAN) ap- proach that uses generators to create samples and a discrim- inator to extract joint spatial-spectral features. The overall accuracy of this approach is 98.2% with 9% percent ratio of testing samples.

In [29] Pauletti et al, the authors present a new deep CNN framework for classification of HSIs. in this approach a shortcut layer connection are introduced, were the information gathered from the lower layer is act as a input of current layer, which then provides output to the remaining upper layers. This research shows an overall accuracy of 97.57%.

in [30] Hong et al, suggested a semi-supervised approach for exploiting multimodal data in order to achieve better inference is presented in this research. Label information can flow from labelled nodes to unlabeled nodes using Graph Convolutional Networks (GCN). MiniGCNs are also proposed in this re-search. MiniGCNs are a new type of GCN that uses regular patches of the original HSI to train the GCN model, resulting in decreased computing costs. For spatial-spectral classifica- tion, a number of popular deep learning-based approaches have recently been used. The proposed MiniGCN shows and overall accuracy of 75.11% with the Indian Pines dataset.

in [31] Roy et al, In this paper, the author has used a HyrbridSN for HSI classification and compared it with some existing techniques. The author has shown overall accuracy of 98.39 for Indian Pines data set and also compared the proposed approach with some of the state of art technologies.

In [32] A. Krizhevsky,Using two RNN layers, the cascaded RNN model was created to investigate the redundant and complimentary information of HSIs. In this research an overall accuracy 91.79% can be observed for Indian Pines data set.

In [33] C. Shi, The authors of this research describe an unique hyperspectral image (HSI) classification frame- work that uses hierarchical recurrent neural networks to explore multi-scale spectral-spatial characteristics. The suggested method takes into account not only the HSI's local spectral- spatial properties, but also the spatial dependency of non-adjacent picture patches at various scales.

3. Proposed methodology

*3.1 Dataset Description:*In this paper we are using only one classical hyperspectral dataset for our propose technique namely "INDIAN PINES".

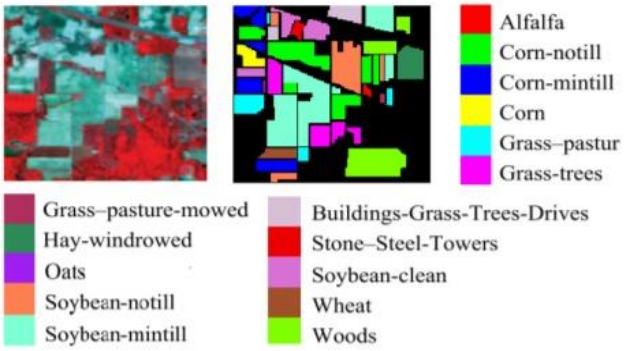

Figure 1. Indian Pines Dataset along with false color image, ground truth and classes

The "Indian Pines" dataset is captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). It comprises of images have 145 X 145 pixels resolution with a spatial resolution of 20m per pixel and 200 spectral bands whose wavelength range is between 0.4–2.5 micrometers after discard 20 bands which were affected by atmosphere absorption. The available ground truth has 16 classes of vegetation as shown in Fig.1.

3.2 Proposed method

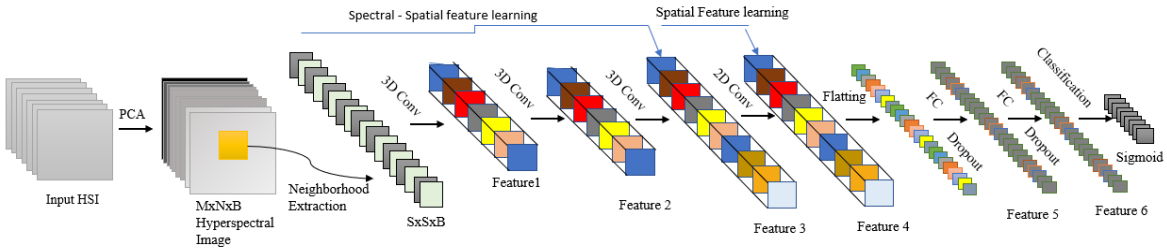

Figure 2. Proposed Hybrid CNN flow Diagram

Let us denote the hypercube by $\mathbf{I} \in \mathcal{R}^{M\times N\times D}$ Where, the actual image input is denoted by I, width is represented by M, height is indicated by N, and number of spectral bands is represented by D. Each HSI pixel in image input I contains D number of spectral measures which forms one-hot label vector Where, the actual image input is denoted by I, width is represented by M, height is indicated by N, and number of spectral bands is represented by D. Each HSI pixel in image input I contains D number of spectral measures which forms one-hot label vector $Y = (y_1, y_2, \dots, y_C) \in \mathcal{R}^{1\times 1\times C}$, Where, C stands for the land-cover classes. But the hyperspectral pixels show a variety of land-cover classifications, resulting in considerable variability in the same class and similarity between multiple classes in I. Taking on this problem is a huge challenge for any model. To eliminate spectral redundancy, PCA is first performed to the genuine HSI data (I) over spectral bands. While keeping the spatial dimensions intact, the number of spectral bands are reduced from D to B using PCA.The width M and height N still remain the same since we reduce only the spectral bands keeping spatial information as it is which will further aid in recognizing the object. Hence, now the data cube after reduction by PCA can be represented as $\mathbf{X} \in$

$\mathcal{R}^{M\times N\times B}$ Where, the updated input after applying PCA is **X**, width is denoted by M, N stands for the height and the B represent the number of spectral bands.

Table 1. Model Summary (Window Size 25 x 25)

Type of Layer	Shape of output	Parameters
input_1 (Input Layer)	(25,25,30,1)	0
conv3d_1 (Conv3D)	(23,23,24,8)	512
conv3d_2 (Conv3D)	(21,21,20,16)	5776
conv3d_3 (Conv3D)	(19,19,18,32)	13856
reshape_1 (Reshape)	(19,19,576)	0
conv2d_1 (Conv2D)	(17,17,64)	331840
flatten_1 (Flatten)	(18496)	0
dense_1 (Dense)	(256)	4735232
dropout_1 (Dropout)	(256)	0
dense_2 (Dense)	(128)	32896
dropout_2 (Dropout)	(128)	0
dense_3 (Dense)	(16)	2064

To use image classification, The hypercube is divided into 3-D-patches which overlap each other and are small and their truth labels are decided on the label of the pixel at the center. Thus, now we have created neighboring 3D patches $P \in \mathcal{R}^{S\times S\times B}$ from **X**, Whose center is located at the spatial location (α, β). It spans over the $S \times S$ window or spatial extent and all spectral bands i.e., B. (n) is the total number of 3D patches generated from X is given by$(M - S + 1) \times (N - S + 1)$. Thus, the location of 3 -D-patch (α, β) is represented by $P_{\alpha,\beta}$, spreads over the width from $\alpha - (S-1)/2$ to $\alpha + (S-1)/2$ and the height from $\beta - (S-1)/2$ to $\beta + (S-1)/2$, and B spectral bands of X i.e., which is the PCA reduced data cube. The raw data is convolved with 2D kernels in the 2D CNN. Convolution is done performing addition of dot product between input data and kernel. To cover the entire spatial dimension of the input data, the kernel is slided across the same. In order to introduce non-linearity in the model, The activation function receives the convolved features. In 2-D convolution, at spatial co-ordinate (x, y) the value of activation in the j^{th} feature map of the i^{th} layer, is denoted by $v_{i,j}^{x,y}$, which is produced by the equation 1:

$$v_{i,j}^{x,y} = \phi\left(b_{i,j} + \sum_{\tau=1}^{d_{l-1}} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j,\tau}^{\sigma,\rho} \times v_{i-1,\tau}^{x+\sigma,y+\rho}\right) \text{---------(1)}$$

Where: ϕ is Activation function,$b_{i,j}$ is Bias parameter for i$^{\text{th}}$ layers of j$^{\text{th}}$ feature map , d_{l-1} is total feature map in $(l-1)^{\text{th}}$ layer and kernel depth $w_{i,j}$ for i$^{\text{th}}$ layer's j$^{\text{th}}$ feature map. $2\gamma + 1$ is Kernel width, $2\delta + 1$ is Kernel height, $w_{i,j}$ is weight parameter value of j$^{\text{th}}$ feature map of the i$^{\text{th}}$ layer.

3D convolution can be done by convolving a 3D kernel over 3-dimensional data [32]. The feature maps of the convolution layer are obtained after convolving the 3-D kernel over many continuous bands in the input layer in the proposed model for HSI data which facilitates extraction of the spectral information. The activation value at spatial dimension (x,y,z) in the i th layer's j th feature map in 3-D convolution, represeted as $v_{i,j}^{x,y,z}$, is generated as equation 2:

$$v_{i,j}^{x,y,z} = \phi\left(b_{i,j} + \sum_{\tau=1}^{d_{l-1}} \sum_{\lambda=-\eta\rho=-\gamma}^{\eta} \sum_{\sigma=-\delta}^{\gamma} w_{i,j,\tau}^{\sigma,\rho,\lambda} \times v_{i-1,\tau}^{x+\sigma,y+\rho,z+\lambda}\right) \text{----------(2)}$$

Depth of kernel along spectral dimension is represented by $2\eta + 1$ and remaining parameters are the identical to those used in equation for 2D CNN.

For training the parameters of CNN like kernel weight *w* and the bias b, generally supervised techniques [12] aided by gradient descent optimization method are used. Traditionally, convolutions are applied only on spatial dimensions that cover all of the feature maps from previous layer in 2D CNNs. Whereas, in view of the HSI classification problem, the requirement is to be able to pick up the spectral information along with the spatial information. Handling the spectral information cannot be done by 2D CNN. Whereas, the 3D kernel is capable of obtaining both spectral and spatial feature in parallel but at the same time spiking up the complexity of computation. To take advantage of both 2-D and 3-D CNNs' automated feature learning capabilities, a hybrid feature learning framework called HybridSN for HSI classification is hereby proposed. However, the classification used at last layer is done by the sigmoid activation function

Hybrid SN network consist three -3-D convolutions, one -2-D convolution , and three fully connected layers which is shown in figure 2.In HybridSN model, the 3-D convolution kernels dimension are represented like $8 \times 3 \times 3 \times 7 \times 1$ (i.e., $K_1^1 = 3, K_2^1 = 3$, and $K_3^1 = 7$ in fig 2),$8 \times 3 \times 3 \times 7 \times 1$ (i.e., $K_1^1 = 3, K_2^1 = 3$, and $K_3^1 = 7$ in Fig. 2),$16 \times 3 \times 3 \times 5 \times 8$ (i.e., $K_1^2 = 3$, $K_2^2 = 3$, and $K_3^2 = 5$ in Fig. 2)$32 \times 3 \times 3 \times 3 \times 16$ (i.e., $K_1^3 = 3, K_2^3 = 3$, and $K_3^3 = 3$ in Fig. 2).In the next first, second, and third convolution layers, respectively, where $16 \times 3 \times 3 \times 5 \times 8$ stands for 16 number of 3-dimensional kernels each of $3 \times 3 \times 5$ dimensions (i.e., two for spatial and one for spectral dimension) for all eight 3-D input feature maps. The dimension of 2-D convolution kernel is $64 \times 3 \times 3 \times 576$ (i.e., $K_1^4 = 3$ and $K_2^4 = 3$ in Fig. 1), here 64 represents number of 2-D-kernels and 3×3 stands for the spatial dimensions of 2-D-kernel, and the number of 2-D input feature maps is represented by 576 for simultaneously

4. Experimental and Discussion:

All experiments are conducted on Google Colab server GPU and Dell system of 16 GB RAM. The optimal learning rate of 0.001 is selected with reference to the outcomes of classification.

4.1Performance measurement criteria

To judge the performance of HSI classification we use different parameters namely:

- Overall Accuracy (OA)=Area correctly labelled/Total number of labels
- Average Accuracy (AA) = refers to the mean value of overall accuracy counter over every class
- Overall Accuracy (OA)=Overall accuracy of each class/Total number of classes
- Kappa Coefficient(kappa) = stands for the interrater agreement between ground truth map and clas sification map.

$$K = Po\text{-}Pe/1\text{-}Pe$$

4.2 Activation function

At the last layer of convolutional neural network, activation function works as the classifier for prediction. In this paper we have compared results from SoftMax activation function with sigmoid activation function.

- SoftMax: SoftMax activation function can be used for predict multinomial probability distribution at the last layer of the neural network i.e. output layer.
- *Sigmoid:* Sigmoid function is nonlinear and special type of activation function. It is specifically used for the model where probability prediction as an output is required. Since probability varies in the range between o and 1. Sigmoid function is appropriate for the purpose of classification.

Table 2. Classification Performance of Hybrid CNN model with SoftMax function on IP

Class name	Precision	Recall	f1 score	Support	Classwise accuracy
Alfalfa	1	1	1	7	100
Corn-notill	1	0.98	0.99	214	98.13084
Corn-mintill	0.98	1	0.99	125	100
Corn	1	1	1	36	100
Grass-pasture	0.99	1	0.99	72	100
Grass-trees	0.99	1	1	109	100
Grass-pasture-mowed	1	1	1	4	100
Hay-windrowed	1	1	1	72	100
Oats	1	1	1	3	100
Soybean-notill	0.98	1	0.99	146	100
Soybean-mintill	0.99	1	1	368	100
Soybean-clean	1	0.99	0.99	89	98.8764
Wheat	1	0.94	0.97	31	93.54839
Woods	1	1	1	190	100
Buildings-Grass-Trees-Drives	1	0.95	0.97	58	94.82759
Stone-Steel-Towers	1	1	1	14	100

Table 3. Classification Performance of Hybrid CNN model with sigmoid function on IP

Class name	Precision	Recall	f1 score	Support	Classwise accuracy
Alfalfa	1	1	1	7	100
Corn-notill	1	0.98	0.99	214	97.6635514
Corn-mintill	0.97	0.99	0.98	125	99.2
Corn	1	1	1	36	100
Grass-pasture	0.95	1	0.97	72	100
Grass-trees	1	1	1	109	100
Grass-pasture-mowed	1	1	1	4	100
Hay-windrowed	1	1	1	72	100
Oats	1	1	1	3	100
Soybean-notill	1	1	1	146	100
Soybean-mintill	1	1	1	368	100
Soybean-clean	1	1	1	89	100
Wheat	1	0.97	0.98	31	96.77419355
Woods	1	1	1	190	100
Buildings-Grass-Trees-Drives	0.98	0.95	0.96	58	94.82758621
Stone-Steel-Towers	1	1	1	14	100

4.3 Classification Results

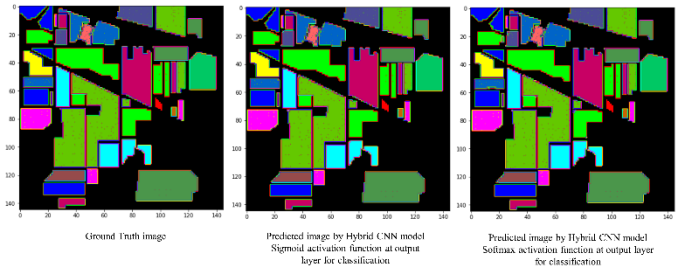

Figure 3. Comparison between ground truth and classified images of both models

In this study we have used 15 % of the sample for training of the model and 85% samples for testing purpose. We have compared the accuracies of results by using softmax at the output layer with the results obtained by sigmoid function at output layer for classification. Figure 3 shows the comparison between ground truth of the Indian Pines dataset with the predicted image generated by hybrid CNN with softmax as classifier with predicted image of hybrid CNN with sigmoid as classifier.

Table 4. comparison between performance of our model with state- of- art methods with testing ratio15%

Parameters	WaveletCNN [28]	HybridSN [29]	**Our approch**
OA	97.34	99.3498	**99.34980**
AA	83.056	99.0864	**99.27908**
kappa	96.97	99.25861	**99.258905**

Table 4 shows comparison between the overall accuracy, average accuracy and kappa accuracy for and WavletCNN model having softmax activation function[28]and Hybrid CNN with softmax as an activation function using by author in [29] with our Hybrid CNN model having sigmoid as an activation for classification. And the performance of HbridSN with sigmoid is better than other model proposed by authors in their paper.

5. Conclusion

In this paper, we have use Indian pines dataset for image samples and applied a 3D and 2D convolutions over HSI data also known as HybridSN. The HybridSN basically clubs information obtained from spatial-spectral and spectral information in the form of 3-D and 2-D convolutions, respectively. We have evaluated existing HybridSN model which used SoftMax activation function for predicting the class of output image labels with the proposed model of this paper which uses sigmoid activation function at the output layer. The proposed model is computationally efficient and the accuracy is up to the mark with the Hybrid CNN model with SoftMax for classification i.e. OA for the model stands at par with the existing model whereas, there is a slight improvement in kappa coefficient (99.258905) as compared to the existing model. Both the models work well for small size of data as we have used only 15% of Indian Pines dataset for the model.

References

1. Lv, W., & Wang, X. (2020). Overview of hyperspectral image classification. *Journal of Sensors*, *2020*.

2. Cai, Y., Liu, X., & Cai, Z. (2019). BS-Nets: An end-to-end framework for band selection of hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(3), 1969-1984.
3. Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., & Zhang, B. (2020). More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(5), 4340-4354
4. K. G. Willson, L. E. Cox, J. L. Hart, and D. C. Dey, "Three-dimensional light structure of an upland quercus stand post-tornado disturbance," J.Forestry Res., vol. 31, no. 1, pp. 141–153, 2020.
5. Y. Xu, B. Du, and L. Zhang, "Beyond the patchwise classification:Spectral-spatial fully convolutional networks for hyperspectral image classification," IEEE Trans. Big Data, vol. 6, no. 3, pp. 492–506,Sep. 2020
6. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis,"IEEE Trans. Geosci. Remote Sens., vol. 44, no. 6, pp. 1586–1600.
7. A. Sharifi, "Using sentinel-2 data to predict nitrogen uptake in maize crop," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13,pp. 2656–2662, 2020,
8. A. Sharifi, "Using sentinel-2 data to predict nitrogen uptake in maize crop," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13,pp. 2656–2662, 2020.
9. A. Sharifi and M. Hosseingholizadeh, "Application of sentinel-1 data to height and biomass of rice crop in Astaneh-ye Ashrafiyeh, Iran,"J. Indian Soc. Remote Sens., vol. 48, pp. 11–19, Oct. 2019.
10. A. Sharifi, "Yield predictionwith machine learning algorithms and satellite images," J. Sci. Food Agriculture, vol. 101, pp. 891–896, Feb. 2021. B. Heinrichs and S. B. Eickhoff, "Your evidence? Machine learning
11. algorithms for medical diagnosis and prediction," Hum. Brain Mapping, vol. 41, no. 6, pp. 1435–1444, 2020J.
12. T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 3, pp. 862–873, Mar. 2009.
13. A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," IEEE Trans. Image Process., vol. 20, no. 3,pp. 612–624, Mar. 2011.
14. A. Kosari, A. Sharifi, A. Ahmadi, and M. Khoshsima, "Remote sensing satellite's attitude control system: Rapid performance sizing for passive scan imaging mode," Aircr. Eng. Aerosp. Technol., vol. 92, no. 7,pp. 1073–1083, Jun. 2020.
15. L. Bruzzone, F. Roli, and S. B. Serpico, "An extension of the jeffreysmatusita distance to multiclass cases for feature selection," IEEE Trans. Geosci. Remote Sens., vol. 33, no. 6, pp. 1318–1321, Nov. 1995.
16. N.Keshava, "Distance metrics and band selection in hyperspectral processingwith applications to material identification and spectral libraries," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 7, pp. 1552–1565, Jul. 2004.
17. E. Blanzieri and F. Melgani, "Nearest neighbor classification of remotesensing images with the maximal margin principle," IEEE Trans. Geosci.Remote Sens., vol. 46, no. 6, pp. 1804–1811, Jun. 2008.
18. J. Xia, L. Bombrun, Y. Berthoumieu, C. Germain, and P. Du, "Spectralspatial rotation forest for hyperspectral image classification," in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2016, pp. 5126–5129.
19. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," IEEE Trans. Geosci. RemoteSens., vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
20. S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," IEEE Trans. Cybern., vol. 48, no. 4, pp. 1176–1188, Apr. 2018.

21. D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu,"Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens.,vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
22. L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data:A technical tutorial on the state of the art," IEEE Geosci. Remote Sens.Mag., vol. 4, no. 2, pp. 22–40, Jun. 2016.
23. W. Hu, Y. Huang, L.Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," J. Sensors, vol. 2015, vol. 2015, Art. no. 258619.
24. G. Cheng, Z.Li,J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 11, pp. 6712-6722, Nov. 2018.
25. Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," IEEE Trans. Geosci. Remote Sens., vol. 56, no, 2, pp. 847-858, Feb. 2018.
26. Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6232-6251, Oct. 2016.
27. L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral-spatial feature learning via deep residual Conv-Deconv network for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 1, pp, 391 – 406, Jan. 2018.
28. Chakraborty, T., & Trehan, U. (2021). SpectralNET: Exploring Spatial-Spectral WaveletCNN for Hyperspectral Image Classification. *arXiv preprint arXiv:2104.00341*.
29. Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2019). HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), 277-281.