A Deep Spectral–Spatial Residual Attention Network for Hyperspectral Image Classification

Koushikey Chhapariya, *Graduate Student Member, IEEE*, Krishna Mohan Buddhiraju, *Member, IEEE*, and Anil Kumar

***Abstract*—In recent years, deep learning algorithms, particularly convolutional neural networks, have significantly improved the performance of the hyperspectral image (HSI) classification. However, due to the high dimensionality of HSI and limited training samples, the deep neural network causes model overfitting. In addition, considering all the bands of HSI datasets equally for feature learning and being unable to distinguish between the edge and the center pixels of a neighborhood reduces classification accuracy. Thus, in this article, we propose an end-to-end deep spectral–spatial residual attention network (DSSpRAN) motivated by the attention mechanism of the human visual system for HSI classification. The DSSpRAN considers input HSI data as a 3-D cube instead of using dimensionality reduction methods. The proposed model simultaneously incorporates spectral and spatial features by considering a spectral residual attention network (SRAN) and a spatial residual attention network (SpRAN). In SRAN, the weights are assigned and learned adaptively to select essential features from each band. The SpRAN enhances the importance of classifying each nearby pixel to the center pixel. It assigns the same label as that of the center pixel to the surrounding pixels, thus limiting pixels with different labels. The proposed method has been evaluated on five different datasets to prove the state of the art for various land use land cover scenarios. A comprehensive qualitative and quantitative analysis of the results shows that the proposed method significantly outperforms other state-of-the-art methods.**

***Index Terms*—Attention network, convolutional neural network (CNN), deep learning framework, hyperspectral image (HSI) classification, remote sensing, residual network, spectral–spatial classification.**

I. Introduction

HYPERSPECTRAL images (HSI) consist of hundreds of contiguous and spectrally narrow wavelength bands. Each pixel in the HSI consists of a different spectrum in terms of wavelength, which relates to the material composition of the features [1]. HSI, with its high spectral resolution, has been widely used in many applications, such as crop monitoring [2], [3], urban development [4], [5], the mining industry [6], [7], environmental and natural resources management [8], [9], surveillance [10], [11], etc. This article mainly focuses on HSI classification, which intends to assign every pixel in the HSI to a specific thematic class. In HSI classification, having abundant spectral information and ensuring a precise pixelwise classification are important characteristics. However, HSI with limited labeled training samples, high dimensionality, and redundancy in adjacent spectral bands increases the challenges of high-accuracy image classification. The Hughes phenomenon (curse of dimensionality) also leads to overfitting on the training data, reducing the generalization ability of the classifier [12]. Thus, many feature extraction and dimensionality-reduction methods, such as principal component analysis [13], linear discriminant analysis [14], local linear embedding [15], sparse representation [16], manifold learning [17], [18], etc., have been proposed to reduce the noise and redundancy in the data. However, these methods require experts to manually tune parameters and select spectral bands with maximum information to avoid significant information loss [19].

These traditional methods can extract only low-level handcrafted features that fail to handle unknown/unseen and complex scenes [20]. Moreover, factors, such as sensor configuration, atmospheric hindrance, and spectral variability, influence HSI data to deal with high intraclass variability and interclass similarity [21], [22]. This increases the challenges for HSI classification, thus demanding more developed techniques. Learning-based algorithms, such as deep learning, models have been widely considered recently due to their adaptive characteristics and enhanced classification results. Deep learning models can extract high-level abstract features using end-to-end hierarchical frameworks. With flexible architectural characteristics, deep neural networks incorporate spectral and spatial information from the acquired data [20], [23], [24], [25].

In recent years, many HSI classification frameworks have incorporated spatial information to maximize the usage of hyperspectral data and retrieve better classification performance. Extended morphological profiles (EMPs) [26], [27], stacked autoencoder (SAE) [28], [29], deep belief networks (DBNs) [30], [31], recurrent neural networks (RNNs) [32], [33], convolutional neural networks (CNNs) [34], [35], [36], generative adversarial networks [37], [38], etc., are such examples of a few deep learning methods. SAE and DBN are unsupervised methods where Chen et al. [39] proposed to use a multilayer architecture with SAE to classify HSI. Li et al. [40] also proposed a single restricted Boltzmann machine for classifying HSI using multilayer DBN. However, SAE and DBN methods focus only

Received 30 May 2023; revised 15 September 2023 and 8 December 2023; accepted 6 January 2024. Date of publication 17 January 2024; date of current version 6 September 2024. The work of Koushikey Chhapariya was supported by the Ministry of Education, Government of India, through Prime Minister Research Fellowship Funding. *(Corresponding author: Koushikey Chhapariya.)*

Koushikey Chhapariya and Krishna Mohan Buddhiraju are with the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: chhkoushikey@gmail.com).

Anil Kumar is with the Indian Institute of Remote Sensing, ISRO, Dehradun 248001, India.

Digital Object Identifier 10.1109/JSTARS.2024.3355071

on spectral features by ignoring nearby pixel information, thus reducing the overall accuracy (OA) in terms of spatial features. Hang et al. [33] proposed a cascaded recurrent neural network to generate discriminative spectral features while keeping the same patch size as the input, leading to the loss of spatial information. Liu et al. [41] introduced a hybrid approach that combines CNN and graph convolutional networks for hyperspectral image classification. This article includes a pixel and superpixel-level feature fusion by leveraging both CNN and GNN mechanisms, showcasing a synergistic approach for improved classification performance in hyperspectral image analysis. However, the feature fusion in their method is relatively simple, thus limiting the model's ability to capture complex spatial-spectral features. EMP with 3-D CNN proposed by Chhapariya et al. [42] shows an effective way to incorporate spectral and spatial features simultaneously but is limited in extracting salient features from an image.

Among all the different methods, the CNN has shown efficacy in achieving satisfactory performance for HSI classification. With its weight-sharing capability, the CNN reduces the required parameters for image classification [43]. Hu et al. [44] introduced 1-D CNN (1-D CNN) to extract spectral information of each pixel. However, the number of training pixels was limited, thus restricting the extraction of complex features. Many researchers combined 1-D CNN with other efficient networks, such as Mou et al. [32] proposed RNN with 1-D CNN. However, this method lacks in extracting spatial information required for land use land cover classification. Furthermore, joint spectral–spatial feature extraction by combining 1-D CNN and 2-D CNN was introduced [45]. A 3-D CNN has also been used in recent works for HSI classification considering spectral and spatial features [21], [46], [47], [48]. Spectral–spatial residual network (SSRN) [49] is an example that learns deep discriminative features using spectral features and spatial information from HSI. Similarly, Tu et al. [50] proposed a fusion of spectral–spatial features using a global–local hierarchical weighted fusion end-to-end classification architecture. Recently, Roy et al. [51] proposed a self-attention mechanism with a vision transformer to introduce a morphological transformer and implement the learnable spectra-spatial network. The attention network considering spectral and spatial information is also being used to solve many hyperspectral unmixing problem statements. Spatial attention-weighted unmixing network [52] and spatial–spectral attention bilateral network for hyperspectral unmixing [53] are examples of recent work in this domain. Similarly, Yu et al. [54] proposed a spatial–spectral dense CNN framework with a feedback attention mechanism for HSI classification. Adding to this, Shu et al. [55] introduced a spectral–spatial split attention network for local features with long-range dependencies. However, many parameters in dense CNN to integrate local features increase the memory requirement and, thus, the memory cost.

All the abovementioned CNN-based HSI classification methods have shown fulfilling results in one way or another, but some questions still need definite answers.

1) Is there any method to optimize the parameters for CNN-based HSI classification?
2) How much spectral information does each band in HSI contribute? Is the contribution equally distributed?
3) How does the classification accuracy change considering nearby pixel information with the center pixel?

This article proposes an end-to-end deep spectral–spatial residual attention network (DSSpRAN) for HSI classification to answer these questions. The main focus of our work is to propose a network that can enhance the spectral and spatial features of the HSI. For extracting discriminative spectral features, we design a spectral residual attention network (SRAN) using 1-D convolutional layers. In SRAN, the weights are assigned and learned adaptively for selecting features from each band. The spatial residual attention network (SpRAN) is designed for extracting spatial features using 2-D convolutional layers. The SpRAN enhances the importance of classifying each nearby pixel to the center pixel. It assigns the same label as the center pixel to the surrounding pixels, thus limiting pixels with different labels. The SRAN and SpRAN models are embedded parallelly into a CNN block to simultaneously emphasize spectral and spatial features. The residual connection addresses the problem of vanishing and exploding gradients in the deep neural network. The batch normalization (BN) and full preactivation rectified linear unit (ReLU) have been used along with the CNN block to avoid model overfitting and increase the training speed of the proposed model. To validate the model's accuracy and generalization ability, we tested our model on five different datasets. Experimental studies show that the proposed model can achieve excellent classification results on all five datasets.

The main contribution of this article is summarized as follows.

1) A two-branch spectral-SpRAN is proposed for HSI classification. The first branch is an SRAN proposed to enhance the importance of individual spectral bands by generating a spectral weight vector. The second branch is a SpRAN, which extracts the spatial features by emphasizing the importance of nearby pixels with the center pixel for image classification.
2) An end-to-end DSSpRAN is proposed, parallelly stacking the spectral and spatial features from SRAN and SpRAN. This channel escalates the classification accuracy by combining spectral and spatial information simultaneously to extract more discriminative features.
3) The effectiveness and generalization ability of the proposed model are demonstrated experimentally on five different datasets that outperform eight compared methods.

The rest of this article is organized as follows. Section II explains the proposed model of this research work. Section III presents the quantitative and qualitative results with the proposed model. Sections IV shows the discussions. Finally, Section V concludes this article with future directions.

II. Methodology

In this section, we first give an overview of the proposed architecture. Further, we explain residual attention networks and illustrate SRAN and SpRAN along with the main framework, i.e., DSSpRAN. We further explain the loss function and optimization method for the proposed work.

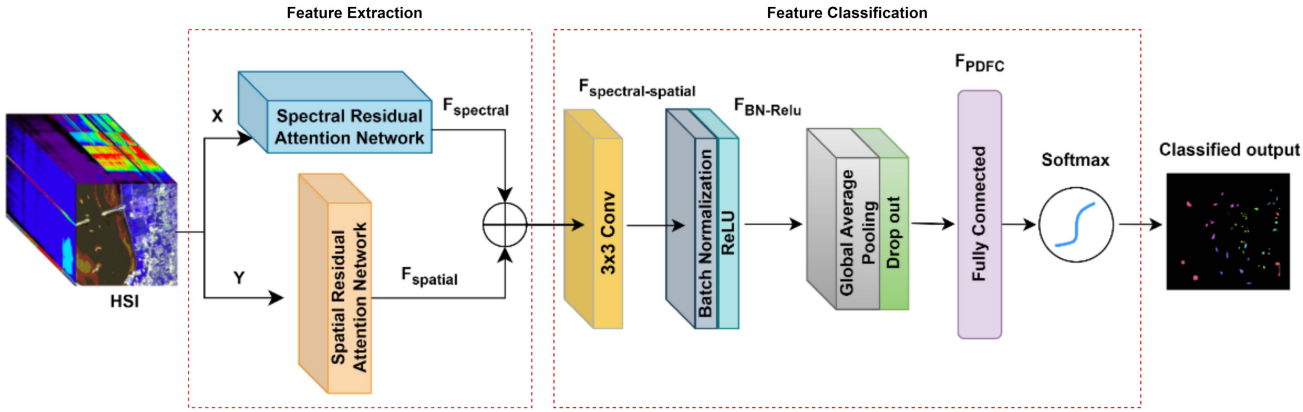

Fig. 1. Proposed DSSpRAN for HSI classification.

A. Framework of the Proposed Architecture

The overall architecture of the proposed method is divided into two sections, feature extraction and feature classification, as shown in Fig. 1. Further, the feature extraction section consists of two core modules, SRAN and SpRAN, which extract spectral and spatial features, respectively. The features extracted from spectral and spatial attention networks are stacked simultaneously in the feature classification module. Residual learning maintains the original spectral and spatial information intact without increasing the feature operations. The feature classification block consists of a convolutional layer, BN, ReLU, and dropout layer to learn low-level features. Further, to learn high-level joint spectral–spatial features, a fully connected layer is used with a softmax function for prediction and classification.

Let HSI data be represented as $I \in \mathbb{R}^{H\times W\times D}$, where I represents input HSI data, and H, W, and D represent height, width (spatial dimension), and the number of spectral bands in HSI, respectively. Suppose there are n labeled pixels with $P = (p_1, p_2, \ldots\ldots, p_n) \in \mathbb{R}^{1\times 1\times D}$ and their corresponding one-hot encoding vectors as $V = (v_1, v_2, \ldots\ldots, v_n) \in \mathbb{R}^{1\times 1\times C}$, where C represents number of classes. The labeled datasets are randomly divided into training, validation, and test sets. The trained models are validated, and the best trained models are used for the performance evaluation on the test dataset. We have not used any dimensionality reduction method to avoid information loss. The spectral vector represented as P_i is considered as input for the SRAN. A 1-D CNN is used to extract the spectral features, which can be formulated as follows:

$$X = \text{Cov}_{1-\text{D}}(P_i) \tag{1}$$

$$F_{\text{spectral}} = P_i \bigoplus \text{SRAN}(X) \tag{2}$$

where F_{spectral} represents extracted spectral features from SRAN. $\text{Cov}_{1-\text{D}}$ is the 1-D convolutional block, $\bigoplus$ is the elementwise addition, and the SRAN(.) denotes the extracted spectral features from the spectral residual attention network. A spatial patch centering the pixel P_i represented as S_{p_i} is considered as an input for SpRAN. A 2-D CNN is used to extract spatial features, which can be formulated as follows:

$$Y = \text{Cov}_{2-\text{D}}(S_{p_i}) \tag{3}$$

$$F_{\text{spatial}} = S_{p_i} \bigoplus \text{SpRAN}(Y) \tag{4}$$

where F_{spatial} represents extracted spatial features form SpRAN. $\text{Cov}_{2-\text{D}}$ is the 2-D convolutional block, $\bigoplus$ is the elementwise addition, and the SpRAN(.) denotes the extracted spatial features from the SpRAN.

B. Residual Attention Network

In HSI, with a limited number of labeled training samples, there is a decrease in classification accuracy with a large number of convolutional layers because of its representation capacity and increased training complexity [56]. However, the decreasing accuracy with increasing convolutional layers can be resolved using a skip connection after every CNN block. Using elementwise addition, these shortcut connections between two CNN blocks build a residual network. The residual network has also solved the vanishing gradient problem. A typical residual connection is represented as

$$R_l = R_{l-1} + F(R_{l-1}) \tag{5}$$

where R_l and R_{l-1} represent the input and output of the residual unit of lth layer. The function $F(.)$ represents nonlinear transformation operations, such as pooling, ReLU, BN, etc. In the proposed architecture, skip connections are used to deepen the network where errors of each layer can be transferred to the previous layer. The skip connections through its shortcut path use a backpropagation stage for faster training as they do not have any additional parameters, thus resolving the vanishing gradient problem.

The attention mechanism has played an essential role in human visual perception. As the name implies, the attention mechanism was introduced into deep learning models to redirect the models to focus more on important features and discard unnecessary features. The attention mechanism is commonly used in computer vision and natural language processing applications. In a typical attention network, a weight matrix is generated from

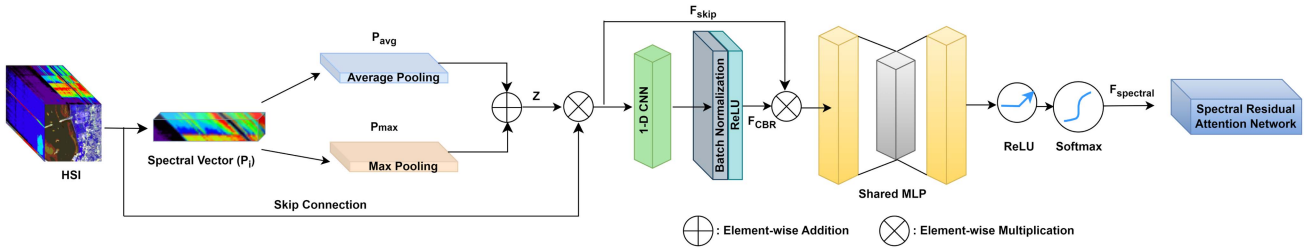

Fig. 2. Proposed SRAN using 1-D convolutional layers.

the input and is applied back to itself. Woo et al. [57] introduced convolutional block attention module (CBAM), a simple yet effective attention network for CNNs. An attention network can be formulated as follows:

$$R_l = R_{l-1}.A(R_{l-1}) \tag{6}$$

where the function $A(.)$ represents operations, such as pooling, fully connected, ReLU, and sigmoid function. In this work, we have proposed the combination of attention network and residual network called residual attention network, which can be represented as

$$R_l = R_{l-1} + F(R_{l-1}).A(F(R_{l-1})). \tag{7}$$

Thus, in the proposed work, we have integrated residual and attention networks to design SRAN and SpRAN. In this attention model, the spectral features are recalibrated by utilizing 1-D CNN with the pooling layer and sigmoid function, thus improving the classification performance. A 2-D convolutional layer, pooling layer, and sigmoid function generate a spatial weight matrix, thus improving the classification performance. It consists of BN and ReLU activation functions and a dropout layer to avoid model overfitting.

C. Spectral Residual Attention Network

The SRAN focuses on increasing the weight of the spectral vector, which is useful for feature representation and classification for the given HSI datasets. Given this, we have developed a mapping function for input to the spectral vectors that can be used to highlight the importance of each spectral band. Consider a feature map $F \epsilon \mathbb{R}^{h\times w\times d}$, where $h \times w$ denotes spatial size and d denotes the number of spectral bands. Different spectral band information from the feature map is combined using two pooling operations illustrated as follows:

$$P_{\text{avg}} = \text{Avg}(F) = \frac{1}{h \times w}\sum_{i=1}^{h}\sum_{j=1}^{w} F_{i,j} \tag{8}$$

$$P_{\max} = \max(F) \tag{9}$$

where $F_{i,j}$ denotes the value at position (i, j) of the input F. The SRAN is shown in Fig. 2. The pooling feature maps are combined and fed as an input to the attention network. These are calculated as follows:

$$Z = P_{\text{avg}} + P_{\max} \tag{10}$$

$$F_{\text{skip}} = X \otimes Z \tag{11}$$

where F_{skip} denotes the skip connection and $\otimes$ denotes the elementwise multiplication. A 1-D CNN is used for mapping spectral vectors and extracting spectral features, followed by the BN layer and ReLU. BN normalizes the layer input of each training cycle and overcomes the covariant shift phenomena, while ReLU learns the nonlinear representation of the extracted feature map. This is given as

$$F_{\text{CBR}} = \text{CBR}(F_{\text{skip}}) \tag{12}$$

where CBR$(.)$ represents features extracted using convolution, BN, and ReLU, respectively. Two bottleneck fully connected layers are considered to enhance the extracted spectral features. This layer tends to reduce model complexity and support model generalization ability. The first layer, represented as W_1, is a dimensionality reduction layer, and the second layer, represented as W_2, is a dimensionality increment layer. The spectral features are calculated and can be formulated as follows:

$$F_{\text{spectral}} = \sigma(W_2(\delta(W_1 F_{\text{CBR}}))) \tag{13}$$

where σ and δ represent ReLU and sigmoid function, respectively.

D. Spatial Residual Attention Network

The SpRAN focuses on increasing spatial information for the nearby pixels with the same class label as the center pixel and restricting pixels with different class labels. Thus, the ideal output of SpRAN will be a matrix of the same height and width as input feature F', where the value of a pixel in this location with the same label as the center is equal to 1 or else 0. The SpRAN architecture can be shown in Fig. 3. Considering CBAM [57] as a reference, two pooling operations, namely, average pooling and max pooling, represented as $P_{\text{avg(Sp)}}$ and $P_{\max(\text{Sp})}$ are performed, and fed as input to the attention network as follows:

$$P_{\text{avg(Sp)}} = \text{Avg}(F') = \frac{1}{d}\sum_{i=1}^{d} F'_{i,j} \tag{14}$$

$$P_{\max(\text{Sp})} = \max(F') \tag{15}$$

$$Z' = P_{\text{avg}} + P_{\max} \tag{16}$$

$$F'_{\text{skip}} = Y \otimes Z' \tag{17}$$

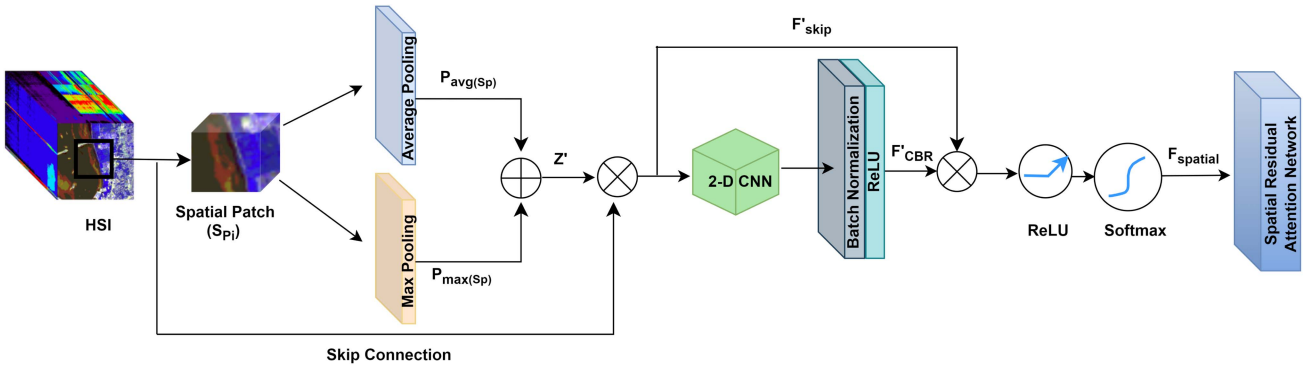

Fig. 3. Proposed SpRAN using 2-D convolutional layers.

where d denotes the number of channels. A 2-D CNN is used for mapping and extracting spatial features, followed by the BN layer and ReLU. The processed spatial features followed by a sigmoid function can be represented as

$$F'_{C_{2-\mathrm{D}}BR} = C_{2-\mathrm{D}}BR(F'_{\mathrm{skip}}) \tag{18}$$

$$F_{\mathrm{spatial}} = \sigma([F'_{C_{2-\mathrm{D}}BR}]) \tag{19}$$

where $F'_{C_{2-\mathrm{D}}BR}$ represents features extracted using 2-D CNN, BN, and ReLU, respectively.

E. Deep Spectral-SpRAN

The extracted spectral–spatial features are fed to a convolution block having a kernel size of 3×3 with stride 1. This is followed by BN and ReLU activation functions to handle nonlinearity in the model. We have also used a global average pooling layer and a dropout layer of 0.4. Since there is no parameter to optimize in the global average pooling, overfitting is avoided at this layer. The output is flattened and input to a fully connected layer along with a softmax function for probability distribution and image classification. The DSSpRAN architecture can be seen in Fig. 1 and can be formulated as follows:

$$F_{\mathrm{BN}} = \mathrm{BN}(F_{\mathrm{spectral-spatial}} * (W + b)) \tag{20}$$

$$F_{\mathrm{BN-ReLU}} = \delta(F_{\mathrm{BN}}) \tag{21}$$

$$F_{\mathrm{PDFC}} = \mathrm{PDFC}(F_{\mathrm{BN-ReLU}}) \tag{22}$$

$$F_{\mathrm{output}} = \mathrm{Sof}(F_{\mathrm{PDFC}}) \tag{23}$$

where $*$ represents the convolution operation, and W and b represent the weight and biases, respectively. δ represents the ReLU function, and PDFC$(.)$ represents the pooling, dropout, and fully connected operation, respectively. Sof$(.)$ denotes the softmax function.

F. Loss Function and Optimization

A loss function is needed to optimize the model for classification problems. In this article, the cross-entropy loss function has been used. The loss function for the cross-entropy is given as follows:

$$\mathrm{Loss}_{\mathrm{CE}} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} y_c^n \log(\hat{y}_c^n) \tag{24}$$

where y and $\hat{y}$ denotes the actual and predicted labels, respectively. C is the number of classes, and N is the number of samples. The model parameters are updated using backpropagation and stochastic gradient descent.

III. Experiments and Results

The following section describes five different datasets considered for the experiment. The factors influencing the model performance, different parameters, and model configurations were discussed. Further, the proposed method is compared with different state-of-the-art deep learning models by calculating standard evaluation metrics.

A. Dataset Details

For this research work, we have considered five standard HSI datasets, namely, Indian Pines (IP), Botswana (BW), Pavia University (PU), Pavia Centre (PC), and Kennedy Space Center (KSC). All five datasets considered have different sizes of available labeled samples and classes. This is to analyze the model classification performance with different sizes of datasets and classes. A 10%, 10%, and 80% of the labeled data were considered randomly for training, validation, and testing sets, respectively.

1) Indian Pines (IP): The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor acquired the IP dataset over the IP test site in North-Western Indiana. It consists of 224 spectral bands, among which 24 bands covering the water absorption region were removed. It has 145×145 pixels, a spectral resolution of 10 nm ranging from 0.4 to 2.5 μm, and a spatial resolution of 20 m/pixel. The ground truth contains 16 vegetation classes with 10 249 labeled pixels. Table I lists the numbers of training, validation, and test samples for each class.

2) Botswana (BW): BW dataset was acquired by a Hyperion sensor over the Okavango Delta, BW, in 2001–2004. It consists of 242 spectral bands, among which 97 uncalibrated and noisy bands that cover water absorption features were removed, and

TABLE I
TRAINING, VALIDATION, AND TEST SAMPLES FOR DIFFERENT CLASSES OF IP DATASETS

No.	Class Name	Train	Val	Test
1	Alfalfa	5	5	37
2	Corn-notill	143	143	1142
3	Corn-mintill	83	83	664
4	Corn	24	24	190
5	Grass-pasture	49	49	386
6	Grass-trees	73	73	584
7	Grass-pasture-mowed	3	3	22
8	Hay-windrowed	97	97	778
9	Oats	2	2	15
10	Soybean-notill	97	97	778
11	Soybean-mintill	246	246	1964
12	Soybean-clean	60	60	474
13	Wheat	21	21	164
14	Woods	127	127	1012
15	Building-Grass-Trees-Drives	39	39	309
16	Stone-Steel-Towers	9	9	74
	Total	979	979	8593

TABLE II
TRAINING, VALIDATION, AND TEST SAMPLES FOR DIFFERENT CLASSES OF BW DATASETS

No.	Class Name	Train	Val	Test
1	Water	76	76	609
2	Hippo grass	24	24	193
3	Floodplain grasses1	26	26	205
4	Floodplain grasses2	26	26	201
5	Reeds	17	17	129
6	Riparian	23	23	183
7	Firescar	11	11	84
8	Island interior	43	43	345
9	Acacia woodland	52	52	416
10	Acacia shrublands	38	38	302
11	Acacia grasslands	42	42	335
12	Short mopane	47	47	370
13	Mixed mopane	91	91	726
14	Exposed soils	10	10	76
	Total	526	526	4174

the remaining 145 bands are considered. It has a 30 m pixel resolution over a 7.7 km strip covering the 400–2500 nm portion of the spectrum in 10 nm windows. The ground truth contains 14 vegetation classes with 3232 labeled pixels. Table II lists the numbers of training, validation, and test samples for each class.

3) PU and PC: These are two scenes acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) over the urban area of the PU and the center city of Pavia, Northern Italy, on 8 July 2002. The number of spectral bands is 102 with 1096×1096 pixels for PC and 103 spectral bands with 610×340 pixels for PU. It has a spatial resolution of 1.3 m with ground truth containing ten vegetation classes for both datasets. Tables III and IV list the numbers of training, validation, and test samples for each class of PU and PC, respectively.

4) Kennedy Space Center (KSC): The KSC dataset was acquired by the National Aeronautics and Space Administration

TABLE III
TRAINING, VALIDATION, AND TEST SAMPLES FOR DIFFERENT CLASSES OF PU DATASETS

No.	Class Name	Train	Val	Test
1	Asphalt	663	663	5305
2	Meadows	1865	1865	14 919
3	Gravel	210	210	1679
4	Trees	306	306	2451
5	Painted metal sheets	135	135	1076
6	Bare Soil	503	503	4023
7	Bitumen	133	133	1064
8	Self-Blocking Bricks	369	369	2946
9	Shadows	95	95	758
	Total	4279	4279	34 221

TABLE IV
TRAINING, VALIDATION, AND TEST SAMPLES FOR DIFFERENT CLASSES OF PC DATASETS

No.	Class Name	Train	Val	Test
1	Water	83	83	659
2	Trees	82	82	656
3	Asphalt	82	82	653
4	Self-Blocking Bricks	81	81	646
5	Bitumen	81	81	646
6	Tiles	126	126	1008
7	Shadows	48	48	381
8	Meadows	82	82	659
9	Bare Soil	82	82	656
	Total	747	747	5964

(NASA) AVIRIS instrument over the KSC, FL, USA, on 23 March 1996. It consists of 224 spectral bands, among which 48 bands containing water absorption and low SNR were removed and the remaining 176 bands are used for analysis. It has a spatial resolution of 18 m and spectral bands of 10 nm width with center wavelengths from 400–2500 nm. The ground truth consists of 13 classes representing the various land cover type with 5211 labeled pixels. Table V lists the numbers of training, validation, and test samples for each class.

B. Experimental Configuration and Parameter Settings

In this article, the experiment has been conducted on a computer with Intel Xeon Silver 4214R CPU at 2.40 GHz with 64 GB RAM and an NVIDIA GeForce RTX A6000 graphical processing unit with 51 GB RAM. The software environment is the Ubuntu 14.04 Ultimate 64-bit system with a deep learning framework of PyTorch. The different hyperparameters used for this research include a learning rate of 0.001 set for stochastic gradient descent with a momentum of 0.7 and weight decay of 0.0005. Full-resolution images have been used to train the network for 100 epochs with a minibatch of size 128.

Root Mean Squared Propagation, or RMSProp, has been used for optimizing the parameters and random initialization to initialize the weights and biases for all CNNs. The performance of the proposed method is evaluated using class-specific accuracy, OA, average accuracy (AA), and kappa coefficient (κ).

TABLE V
TRAINING, VALIDATION, AND TEST SAMPLES FOR DIFFERENT CLASSES OF KSC DATASETS

No.	Class Name	Train	Val	Test
1	Scrub	76	76	609
2	Willow swamp	24	24	193
3	CP hammock	26	26	205
4	Slash pine	26	26	201
5	Oak/Broadleaf	17	17	129
6	Hardwood	23	23	183
7	Swamp	11	11	84
8	Graminoid marsh	43	43	345
9	Spartina marsh	52	52	416
10	Cattail marsh	38	38	302
11	Salt marsh	42	42	335
12	Mud flats	47	47	370
13	Water	91	91	726
	Total	516	516	4098

TABLE VI
CLASSIFICATION RESULTS OF DIFFERENT STATE-OF-THE-ART METHODS FOR LABELED PIXELS OF THE IP DATASET

Class	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
1	91.27	91.83	82.12	99.81	97.98	98.99	99.87	**100**
2	83.45	82.45	81.97	98.23	95.91	94.99	98.79	99.87
3	81.83	94.93	76.94	**100**	93.01	93.19	99.02	99.45
4	95.81	95.95	67.11	99.54	91.02	97.92	98.99	98.4
5	71.23	98.04	93.91	91.02	99.87	100	92.95	99.78
6	79.01	96.93	92.78	97.93	94.94	99.81	**100**	99.56
7	61.93	87.67	93.47	92.01	95.91	92.94	97.65	99.56
8	89.12	90.01	81.78	90.01	**100**	94.87	98.79	98.87
9	99.01	**100**	84.07	95.91	94.91	99.56	**100**	95.78
10	39.01	94.16	82.33	90.04	**100**	99.08	99.87	96.89
11	64.32	100	94.78	93.55	99.82	**100**	99.87	98.82
12	85.92	86.93	96.23	95.91	92.93	97.93	91.02	**100**
13	71.12	98.25	90.56	90.99	90.04	96.92	93.67	99.19
14	49.02	79.23	84.98	91.93	98.06	94.03	93.77	**100**
15	78.06	91.08	89.01	92.98	95.95	97.55	**100**	94.99
16	87.98	81.37	91.98	**100**	99.84	99.76	95.97	**100**
OA (%)	79.17	84.79	90.08	94.76	97.45	96.45	97.34	**99.03**
AA (%)	78.54	86.43	92.54	92.89	96.45	95.01	97.98	**98.08**
Kappa × 100	79.01	84.12	91.58	92.23	96.94	95.97	98.90	**98.98**

The best results are highlighted in bold.

C. Comparison With State-of-the-Art Methods

In this research work, we have compared our proposed model DSSpRAN with different state-of-the-art methods. We have considered two traditional models: support vector machine (SVM) [58] and 3-D CNN [59] and five deep learning-based models specific to residual and attention networks for HSI classification. These models are SSRN [49], residual spectral–spatial attention network RSSAN [60], adaptive spectral–spatial multiscale network (ASSMN) [38], cross-attention spectral–spatial network (CASSN) [61], and spatial proximity feature selection with residual spatial–spectral attention network (SPFS-RSSAN) [62]. SSRN is one of the earliest works to use residual networks for the spectral–spatial classification of HSI. The latter four networks are based on residual and attention networks using different convolution layers with PyTorch or Keras deep learning framework.

1) Quantitative Evaluation: The classwise accuracy and quantitative metric comparison with different state-of-the-art methods are shown in Tables VI–X. Compared with the traditional SVM method, the CNN-based deep learning models perform better with high accuracy. This could be because of the hierarchical feature learning capacity of deep learning methods with the classifiers. Thus, the deep learning methods extract high-level features by learning discriminative features required for HSI classification. It can be observed from the table that the accuracy of 3-D CNN is lower than other deep learning methods since it does not use a residual network. It can be observed that the proposed model performs better for all five datasets than

TABLE VII
CLASSIFICATION RESULTS OF DIFFERENT STATE-OF-THE-ART METHODS FOR LABELED PIXELS OF THE BW DATASET

Class	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
1	90.45	96.23	91.76	96.45	97.93	96.43	97.45	**98.95**
2	93.94	92.96	89.07	91.23	93.45	92.89	94.67	97.89
3	86.98	**100**	94.23	90.43	92.89	94.78	98.23	**100**
4	77.08	84.87	90.45	95.98	91.2	91.27	91.29	98.94
5	**100**	99.05	**100**	93.76	**100**	**100**	99.45	**100**
6	85.45	92.76	93.45	92.19	94.89	95.89	94.78	96.98
7	98.08	86.92	92.9	90.87	98.91	97.98	94.99	**100**
8	76.45	**100**	95.78	95.93	95.98	**100**	97.89	**100**
9	98.42	92.56	97.29	91.29	96.91	91.94	93.78	95.89
10	59.53	94.23	98.89	98.72	96.92	93.28	97.59	**100**
11	97.35	91.45	93.47	92.2	99.91	94.78	92.94	99.76
12	92.38	95.4	91.23	92.99	91.99	93.09	96.95	99.89
13	**100**	97.92	99.03	**100**	99.79	**100**	99.56	**100**
14	93.96	89.03	93.85	97.56	96.32	96.97	95.99	99.79
OA (%)	89.04	92.89	94.34	95.34	95.98	97.89	97.94	**98.92**
AA (%)	88.78	90.28	92.39	93.43	94.34	97.6	96.56	**98.89**
Kappa × 100	89.01	91.98	92.3	93.98	94.56	96.21	97.12	**98.05**

The best results are highlighted in bold.

TABLE VIII
CLASSIFICATION RESULTS OF DIFFERENT STATE-OF-THE-ART METHODS FOR LABELED PIXELS OF THE PU DATASET

Class	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
1	90.12	91.23	92.34	94.62	92.78	95.76	93.45	97.65
2	87.45	90.21	91.03	97.83	96.12	92.98	99.87	**100**
3	78.92	87.98	93.23	96.99	91.18	94.67	99.76	98.99
4	91.23	93.32	91.23	91.78	96.89	92.89	97.89	97.89
5	85.45	97.97	90.02	**100**	**100**	96.97	94.56	**100**
6	90.09	91.23	91.78	98.99	99.34	**100**	90.98	98.23
7	67.98	89.57	94.34	93.56	98.76	**100**	**100**	**100**
8	70.34	96.89	93.98	99.64	**100**	91.26	94.67	99.87
9	89.03	91.29	92.95	98.76	93.45	99.88	91.23	97.93
OA (%)	81.23	89.98	93.89	96.23	96.78	96.99	97.34	**98.76**
AA (%)	80.43	91.23	92.45	94.56	95.98	96.01	95.67	**98.12**
Kappa × 100	79.48	89.94	90.09	92.45	94.89	95.82	97.12	**97.95**

The best results are highlighted in bold.

TABLE IX
CLASSIFICATION RESULTS OF DIFFERENT STATE-OF-THE-ART METHODS FOR LABELED PIXELS OF THE PC DATASET

Class	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
1	83.22	89.45	92.98	94.56	93.89	96.87	95.34	98.76
2	91.2	94.65	90.87	91.27	94.99	92.26	97.89	97.99
3	78.93	91.23	89.67	93.73	97.89	96.09	94.69	99.89
4	65.78	89.76	89.9	93.26	91.22	93.89	99.02	97.45
5	80.93	97.34	93.92	90.65	96.34	91.27	95.9	99.08
6	93.67	97.87	91.29	97.13	94.27	90.87	95.98	96.21
7	95.89	**100**	99.87	94.2	93.67	**100**	99.89	**100**
8	87.92	98.34	**100**	98.23	**100**	94.67	**100**	99.85
9	93.67	86.78	95.98	92.93	91.05	98.79	93.8	**100**
OA (%)	86.59	91.2	92.81	94.34	94.94	96.12	97.08	**98.9**
AA (%)	88.9	90.09	91.87	91.23	93.89	94.34	96.98	**97.89**
Kappa × 100	86.21	90.98	92.01	92.91	92.67	94.98	96.59	**97.63**

The best results are highlighted in bold.

TABLE X
CLASSIFICATION RESULTS OF DIFFERENT STATE-OF-THE-ART METHODS FOR LABELED PIXELS OF THE KSC DATASET

Class	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
1	86.12	90.23	92.1	91.98	93.89	91.95	94.87	97.45
2	90.26	87.69	91.29	90.93	91.2	90.98	93.58	98.82
3	89.23	92.34	93.67	96.89	98.76	97.87	**100**	**100**
4	79.12	91.21	89.89	**100**	92.79	93.56	97.56	98.99
5	82.2	95.34	95.9	94.78	89.07	97.92	92.45	94.67
6	74.51	90.01	91.26	98.72	96.81	**100**	98.87	98.12
7	85.23	93.23	97.76	99.87	93.91	97.67	91.24	94.01
8	87.34	96.51	98.92	91.23	91.98	92.39	92.46	97.89
9	91.23	91.27	98.01	90.01	91.12	90.19	94.81	98.7
10	85.86	93.12	92.07	93.06	97.29	95.98	94.71	99.81
11	93.5	94.59	93.7	92.93	92.32	97.92	96.48	**100**
12	96.12	99.87	**100**	**100**	99.87	96.78	95.68	**100**
13	81.29	92.34	97.89	98.71	91.99	98.78	94.79	98.7
OA (%)	81.27	91.98	92.39	94.21	93.98	96.43	96.79	**97.98**
AA (%)	83.24	90.74	91.19	92.97	94.83	94.67	96.02	**98.23**
Kappa × 100	81.98	91.81	91.39	93.29	94.21	94.92	95.2	**96.98**

The best results are highlighted in bold.

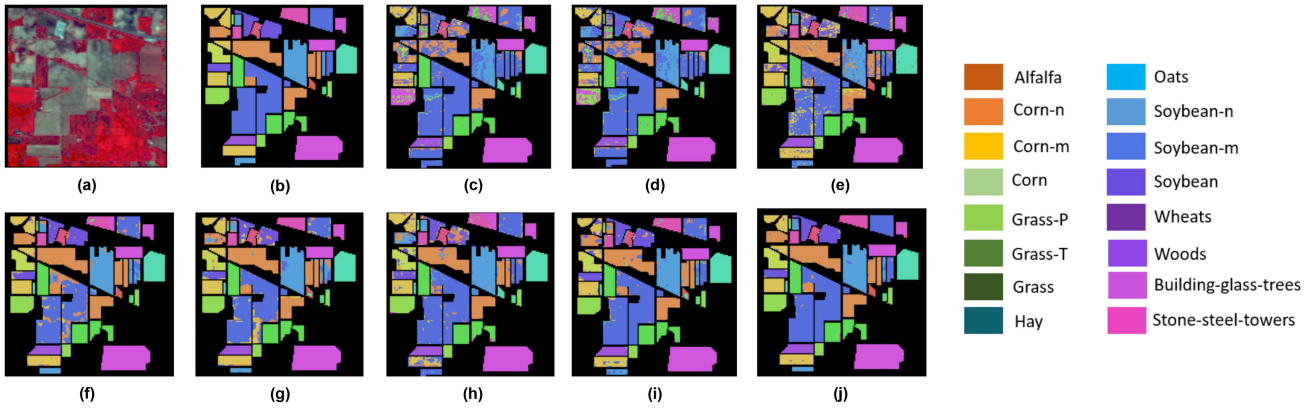

Fig. 4. Classification maps for the IN dataset. (a) False-color image. (b) Ground-truth map. (c)–(j) Classification maps of SVM, 3-D CNN, SSRN, RSSAN, ASSMN, CASSN, SPFS-RSSAN, and Proposed DSSpRAN.

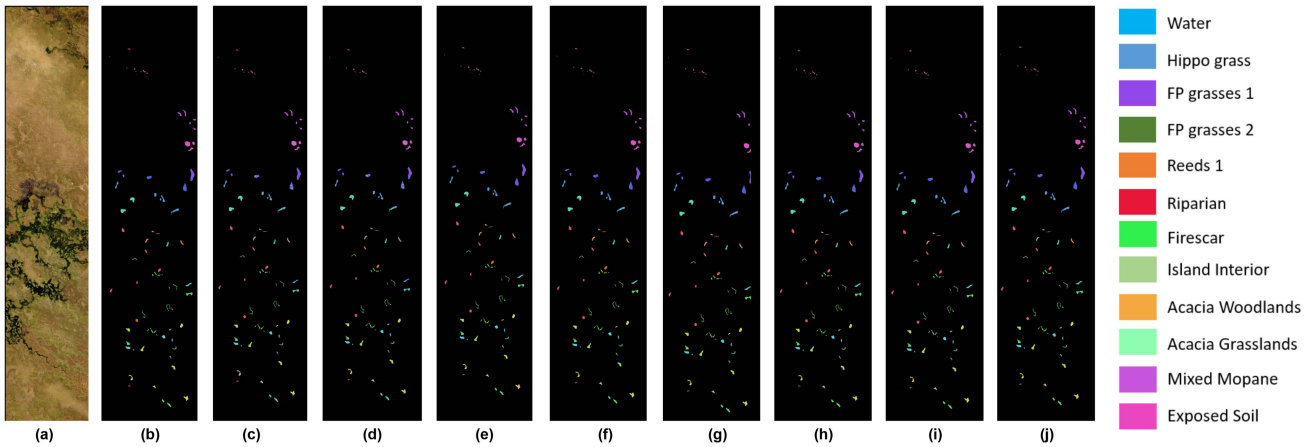

Fig. 5. Classification maps for the BW dataset. (a) False-color image. (b) Ground-truth map. (c)–(j) Classification maps of SVM, 3-D CNN, SSRN, RSSAN, ASSMN, CASSN, SPFS-RSSAN, and Proposed DSSpRAN.

other residual and attention networks in terms of OA, AA, and kappa coefficient. The SSRN excludes attention networks, while the RSSAN uses the dimensionality reduction method, thus reducing spectral information. ASSMN uses a long short-term memory network for HSI classification, which is not well-suited for classification tasks where input data are not a sequence. At the same time, CASSN and SPFS-RSSAN have a large number of computational parameters, thus requiring more training data to learn effectively. On the other hand, the proposed method uses SRAN and SpRAN to enhance spectral bands having the most important features and weaken the remaining insignificant bands, thus achieving higher accuracy with less computational parameters.

2) Qualitative Evaluation: The visual comparison of the proposed model with different state-of-the-art methods is shown in Fig. 4–8. The false color composition and their respective ground truth for five different datasets are considered for qualitative evaluation. The SVM, 3-D CNN, and SSRN show pixel-level misclassification for some classes, visible in Fig. 4–8 for different datasets. In comparison, the remaining methods generate smoother classification maps, specifically at the boundary of two different classes. The proposed model produces a better classification map with distinct boundaries between two different classes as it uses the SpRAN layer to learn the correlation between the center pixel and surrounding neighboring pixels.

D. Ablation Study

In this section, the proposed algorithm's performance has been analyzed by considering different factors. These factors are the effect of different proportions of training samples, spatial input patch size, and learning rate. We have analyzed the classification performance of PU and PC as they share similar pixel labels considering training on PU and testing it with PC and vice versa. We also observed the effect of the attention network on the classification accuracy of the proposed network.

1) Effect of Training Sample: We have compared the effect of different proportions of training samples on OA. We have randomly considered 1%, 3%, 5%, 10%, 15%, 20%, 25%, and 30% of labeled training samples as the training set to train DSSpRAN. The OA, as shown in Fig. 9, increases with the increase in training samples. It can be observed that at 10% of the training sample, OA is the highest for all the datasets.

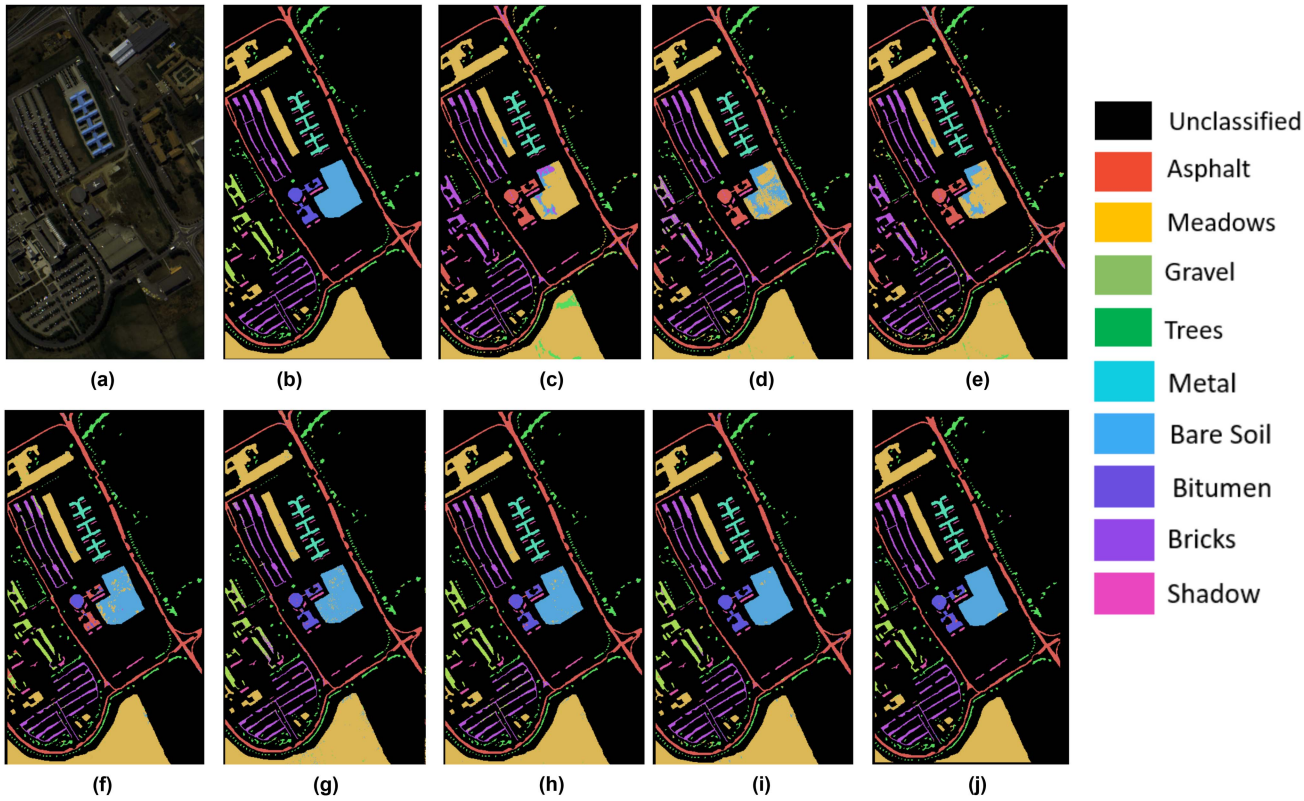

Fig. 6. Classification maps for the PU dataset. (a) False-color image. (b) Ground-truth map. (c)–(j) Classification maps of SVM, 3-D CNN, SSRN, RSSAN, ASSMN, CASSN, SPFS-RSSAN, and Proposed DSSpRAN.

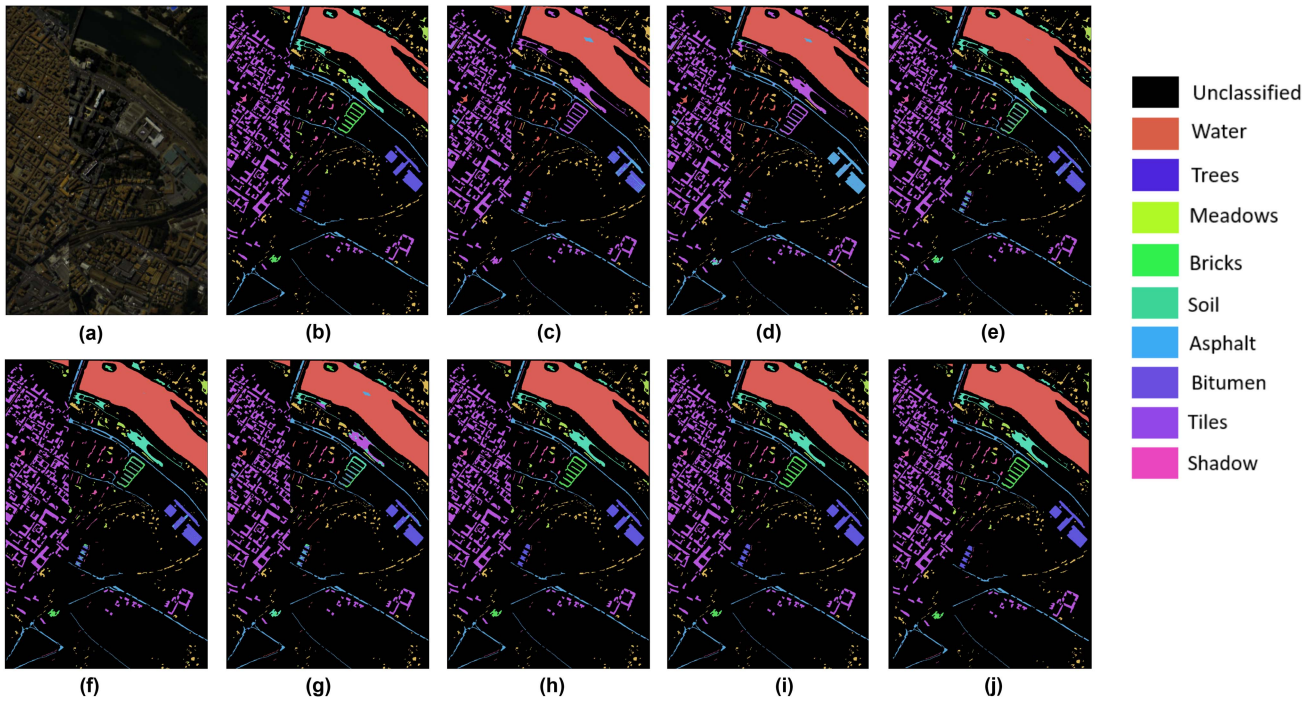

Fig. 7. Classification maps for the PC dataset. (a) False-color image. (b) Ground-truth map. (c)–(j) Classification maps of SVM, 3-D CNN, SSRN, RSSAN, ASSMN, CASSN, SPFS-RSSAN, and Proposed DSSpRAN.

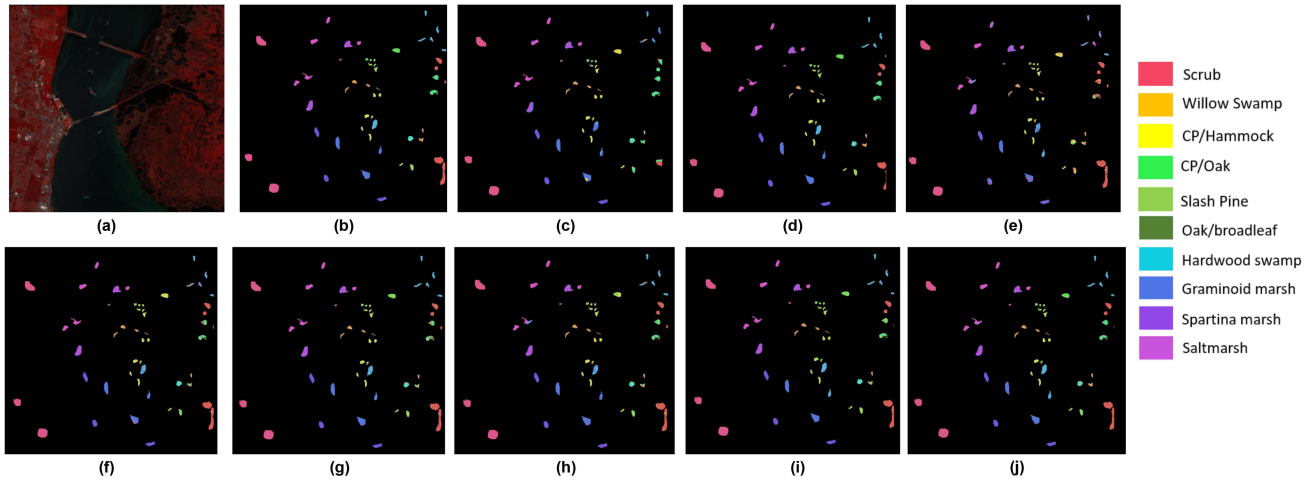

Fig. 8. Classification maps for the KSC dataset. (a) False-color image. (b) Ground-truth map. (c)–(j) Classification maps of SVM, 3-D CNN, SSRN, RSSAN, ASSMN, CASSN, SPFS-RSSAN, and Proposed DSSpRAN.

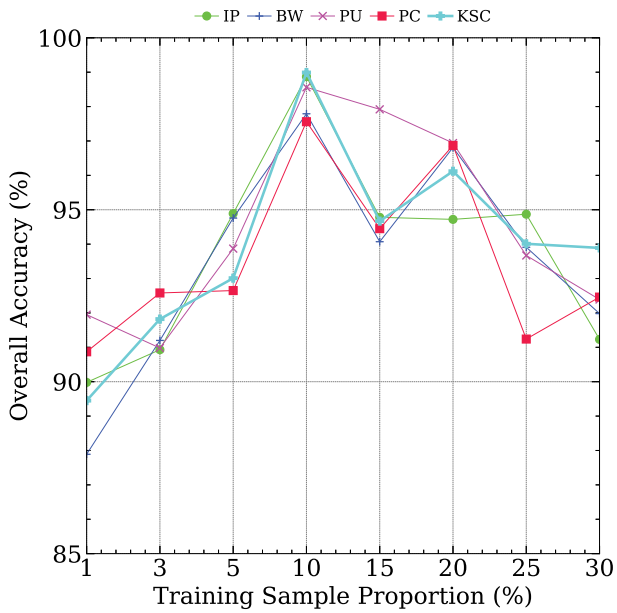

Fig. 9. OA (%) of DSSpRAN with different training sample proportions in the IN, BW, PU, PC, and KSC datasets.

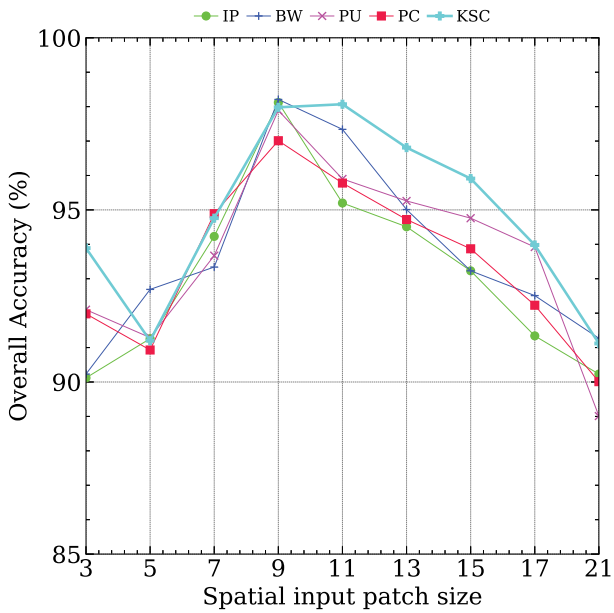

Fig. 10. OA (%) of DSSpRAN with different input patch sizes for the IN, BW, PU, PC, and KSC datasets.

2) Effect of Spatial Input Patch Size: Spatial input patch size tells about the amount of spatial information required by nearby pixels to classify one center pixel. We have compared the effect of different spatial input patch sizes with OA. We considered patch size as small as 3 and increased it to 5, 7, 9, 11, 13, 15, 17, and 21. From Fig. 10, it can be observed that for the patch size 9, the OA for all the datasets was high compared to others. Thus, a patch size 9 has been considered for this research work.

3) Effect of Learning Rate: The learning rate has a significant role in the convergence of training with minimum loss. It controls the steps of gradient descent and, thus, the training speed. We have considered different learning rates to find their effect on OA. We considered a grid search of 0.00003, 0.0001, 0.0003, 0.001, 0.003, and 0.01. From Fig. 11, it can be observed that for the learning rate of 0.001, the OA for all the datasets was higher than others. Thus, a learning rate of 0.001 has been considered for this research work.

4) PU for Training and PC for Testing and Vice Versa: The PU and PC datasets have been acquired by the same sensor (ROSIS) over an overlapping area (Northern Italy). Thus, we have considered these two datasets for training at one and testing at another to analyze the performance accuracy. From Table XI, it can be observed that training the model on PC and testing it on PU give better OA, AA, and kappa accuracy. The PC is a

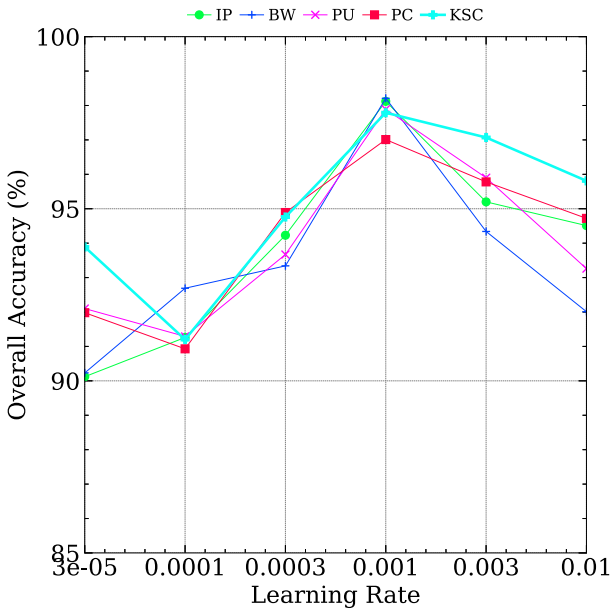

Fig. 11. OA (%) of DSSpRAN with different learning rates for the IN, BW, PU, PC, and KSC datasets.

TABLE XI
COMPARISON RESULT BETWEEN PU AND PC CONSIDERING PU FOR TRAINING AND PC FOR TESTING AND VICE VERSA

Training	Test	OA (%)	AA (%)	Kappa x 100
PC	PU	97.48	95.9	95.89
PU	PC	94.92	94.18	93.74

TABLE XII
EFFECT OF ATTENTION NETWORK

Attention Module	IP	BW	PU	PC	KSC
None	91.24	90.29	90.87	87.12	91.39
SRAN	89.02	89.27	87.92	84.91	90.82
SpRAN	92.93	92.49	93.13	93.59	90.92
SRAN + SpRAN	96.82	94.98	94.91	93.69	97.96

1096×1096 pixels image while PU is 610×340 pixels where some features of PU are a subset of PC that explains the increase in OA for PC as training and PU as a testing site.

5) Effect of Attention Network: We have compared the effect of the attention module on OA for different datasets. We considered the classification accuracy by removing the attention module, i.e., just the convolution block, similar to HSI classification with CNN. Further, we consider only the spectral attention module, the spatial attention module, and the spectral+spatial attention module for all datasets. From Table XII, it can be observed that the OA is higher when we use both spectral and spatial attention modules simultaneously. We also observed that OA for PU and PC is better with the spatial attention module than the spectral attention module. This can be justified as these datasets have a high spatial resolution of 1.3 m/pixel, thus explaining the effect of neighboring pixels on the classification of the center pixel.

TABLE XIII
COMPARISON OF TRAINING AND TESTING TIME OF DIFFERENT METHODS ON EACH DATASET

Datasets	Time (sec)	SVM	3-D CNN	SSRN	RSSAN	ASSMN	CASSN	SPFS-RSSAN	Proposed
IP	Train	5.92	298.17	731.64	1278.41	502.86	318.76	413.78	216.34
	Test	3.16	28.64	18.92	62.83	13.63	11.31	5.51	3.16
BW	Train	4.21	218.63	431.60	1310.09	468.76	301.86	269.11	118.79
	Test	2.16	19.88	16.28	48.17	8.31	10.97	3.86	2.18
PU	Train	7.52	343.87	1028.69	2783.91	894.79	517.81	501.87	389.27
	Test	3.16	45.73	72.84	103.24	19.16	14.63	7.63	5.91
PC	Train	6.93	312.64	942.16	1912.86	684.63	412.31	319.17	284.63
	Test	4.16	29.70	64.36	89.73	16.01	12.07	4.91	3.92
KSC	Train	2.92	114.29	401.11	1189.3	343.61	181.68	264.82	186.87
	Test	0.89	12.62	14.23	42.87	11.03	7.69	4.12	2.08

6) Analysis for Computation of Time: The efficiency of the proposed method DSSpRAN is compared with other methods for each dataset by computing training and testing time, as shown in Table XIII. The training time reflects the complexity of the model while the testing time defines the model's efficiency in practical application. Among all compared methods, the conventional method SVM has the least cost computation with the simplest architecture. The 3-D CNN network having higher parameters takes much more time to train compared to SVM. Attention networks, such as RSSAN and SSRN, with their complex architecture have lengthy computation time. The other three methods, i.e., ASSMN, CASSN, and SPFS-RSSAN, consider all the bands of HSI and thus take longer training time with all the convolutional layers and the hundreds of thousands of self-attention parameters. There is always a tradeoff between accuracy and the computation cost in accordance with the model's architecture. Nevertheless, in our proposed model, we have focused on specific bands with nearby pixel information having less number of convolutional layers to extract the spectral–spatial features. Thus, DSSpRAN results in a relatively fast and efficient performance on all five datasets with high accuracy.

IV. DISCUSSION

In this study, the effectiveness of the proposed method DSSpRAN is validated through experimental results. It is observed that the performance accuracy increases by focusing on nearby pixel information to enhance spatial resolution and by considering the most effective spectral bands. Moreover, a residual network adding skip layers and feedback connection solves the vanishing gradient problem and the attention network redirects the models to focus more on important features and discard unnecessary features. The proposed DSSpRAn considers spectral and spatial features simultaneously in two different blocks (SRAN and SpRAN), thus extracting more discriminative features. In addition, using BN at each convolutional layer reduces the need for thousands of iterations for training in our model compared to [49] and [63].

Traditional methods, such as SVM, have limitations, as they need manual feature engineering and face challenges with the complex spectral-spatial patterns in hyperspectral data. In contrast, CNNs perform well by independently learning distinctive features from raw data, demonstrating their ability to extract

high-level features. Thus, the proposed DSSpRAN extends the capabilities of contemporary deep learning methods by incorporating attention mechanisms, enabling the model to focus on salient spectral and spatial features. This enhances the network's adaptability by surpassing the limitations of traditional methods. The comparison highlights the significant impact of contemporary deep learning methodologies, thus demonstrating our proposed model as an advanced and effective solution for hyperspectral image classification and analysis.

However, the method is limited to smaller datasets with limited spatial resolution. It will be interesting to observe the efficiency of the model on a larger dataset with a higher spatial resolution and complex features. In addition, in most cases, a higher training percentage increases the model's accuracy. However, in our case, a larger patch size leads to a broader receptive field and thus is more prone to learn noise, outliers, and irrelevant information as the training data percentage increases. Thus, with extensive computation and analysis, we have considered a training size of 10% and a patch of 9 for optimum and best accuracy. Nevertheless, the proposed work demonstrates that even with a limited number of training samples available, DSSpRAN avoids overfitting and achieves state-of-the-art classification accuracy.

V. Conclusion

In this research, we have proposed an end-to-end CNN-based architecture for HSI classification using spectral and spatial information. The DSSpRAN deep learning framework consists of an SRAN and a SpRAN module that help in selecting the most effective spectral bands and focus on increasing the spatial information for the nearby pixels, respectively. The SRAN considers spectral features across spatial dimensions to distribute the weight of each spectral band and select only the useful ones. The SpRAN considers a spatial patch as input to enhance surrounding pixels with the same class labels as the center pixel while constraining pixels with different class labels for HSI classification. The training of the proposed model was accelerated by concatenating SRAN and SpRAN to a CNN block, thus reducing the possibility of model overfitting. Compared with the other residual attention network model, the proposed DSSpRAN model achieves better classification accuracy even with the limited number of training samples. The experimental result on five different datasets proves the state of the art for various land use land cover scenarios and, thus, can be generalized to other remote sensing datasets. Further work is being taken up where the model can be trained to learn more discriminative spectral–spatial features for HSI classification and reduce the computation time by an iterative process, thus increasing algorithm efficiency.

References

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] S. Alahmari et al., "Hybrid multi-strategy aquila optimization with deep learning driven crop type classification on hyperspectral images," *Comput. Syst. Sci. Eng.*, vol. 47, no. 1, pp. 375–391, 2023.

[3] J. G. A. Barbedo, "A review on the combination of deep learning techniques with proximal hyperspectral images in agriculture," *Comput. Electron. Agriculture*, vol. 210, 2023, Art. no. 107920.

[4] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[5] S. Pattem and S. Thatavarti, "Hyperspectral image classification using machine learning techniques-a survey," in *Proc. IEEE Int. Students' Conf. Elect. Electron. Comput. Sci.*, 2023, pp. 1–14.

[6] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.

[7] Y. Zhang et al., "Mapping soil available copper content in the mine tailings pond with combined simulated annealing deep neural network and UAV hyperspectral images," *Environ. Pollut.*, vol. 320, 2023, Art. no. 120962.

[8] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[9] A. Alboody et al., "A new remote hyperspectral imaging system embedded on an unmanned aquatic drone for the detection and identification of floating plastic litter using machine learning," *Remote Sens.*, vol. 15, no. 14, 2023, Art. no. 3455.

[10] B. Uzkent, A. Rangnekar, and M. Hoffman, "Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 39–48.

[11] X. Xu, B. Ban, H. R. Howard, S. Chen, and G. Wang, "Mapping and dynamic monitoring of military training-induced vegetation cover loss using Sentinel-2 images and method comparison," *Environ. Monit. Assessment*, vol. 195, no. 2, 2023, Art. no. 320.

[12] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[13] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying Land Inf. Sci.*, vol. 62, no. 2, pp. 115–122, 2002.

[14] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[16] X. Zhang, Q. Song, Z. Gao, Y. Zheng, P. Weng, and L. Jiao, "Spectral–spatial feature learning using cluster-based group sparse coding for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4142–4159, Sep. 2016.

[17] D. L. Donoho et al., "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*,2000. [Online]. Available: https://dl.icdst.org/pdfs/files/236e636d7629c1a53e6ed4cce1019b6e.pdf

[18] Y. Duan, H. Huang, and Y. Tang, "Local constraint-based sparse manifold hypergraph learning for dimensionality reduction of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 613–628, Jan. 2021.

[19] L. Lin, C. Chen, J. Yang, and S. Zhang, "Deep transfer HSI classification method based on information measure and optimal neighborhood noise reduction," *Electronics*, vol. 8, no. 10, 2019, Art. no. 1112.

[20] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[21] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Visual attention-driven hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8065–8080, Oct. 2019.

[22] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.

[23] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *J. Imag.*, vol. 5, no. 5, 2019, Art. no. 52.

[24] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[25] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.
[26] A. Plaza, P. Martínez, J. Plaza, and R. Pérez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.
[27] K. Chhapariya, K. M. Buddhiraju, and A. Kumar, "CNN-based salient object detection on hperspectral images using extended morphology," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6015705.
[28] J. Zabalza et al., "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, 2016.
[29] Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-spatial classification of hyperspectral image using autoencoders," in *Proc. 9th Int. Conf. Inf. Commun. Signal Process.*, 2013, pp. 1–5.
[30] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.
[31] C. Chen, Y. Ma, and G. Ren, "Hyperspectral classification using deep belief networks based on conjugate gradient update and pixel-centric spectral block features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4060–4069, Jul. 2020.
[32] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
[33] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
[34] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 120–147, 2018.
[35] S. K. Roy, R. Mondal, M. E. Paoletti, J. M. Haut, and A. Plaza, "Morphological convolutional neural networks for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8689–8702, 2021.
[36] K. Chhapariya, K. M. Buddhiraju, and A. Kumar, "Spectral-spatial classification of hyperspectral images with multi-level CNN," in *Proc. 12th Workshop Hyperspectral Imag. Signal Processing: Evol. Remote Sens.*, 2022, pp. 1–5.
[37] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.
[38] J. Wang, F. Gao, J. Dong, and Q. Du, "Adaptive dropblock-enhanced generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5040–5053, Jun. 2021.
[39] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[40] T. Li, J. Zhang, and Y. Zhang, "Classification of hyperspectral image based on deep belief networks," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 5132–5136.
[41] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657–8671, Oct. 2021.
[42] K. Chhapariya, K. M. Buddhiraju, and A. Kumar, "Hyperspectral salient object detection using extended morphology with CNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 429–432.
[43] W. N. Khotimah, M. Bennamoun, F. Boussaid, F. Sohel, and D. Edwards, "A high-performance spectral-spatial residual network for hyperspectral image classification with small training data," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3137.
[44] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, 2015, Art. no. 258619.
[45] L. Mou and X. X. Zhu, "Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 110–122, Jan. 2020.
[46] 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.
[47] B. Liu, X. Yu, P. Zhang, X. Tan, R. Wang, and L. Zhi, "Spectral–spatial classification of hyperspectral image using three-dimensional convolution network," *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art. no. 016005.
[48] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3904–3908.
[49] 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, Oct. 2017.
[50] B. Tu, W. He, W. He, X. Ou, and A. Plaza, "Hyperspectral classification via global-local hierarchical weighting fusion network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 184–200, Dec. 2021.
[51] S. K. Roy, A. Deria, C. Shah, J. M. Haut, Q. Du, and A. Plaza, "Spectral–spatial morphological attention transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5503615.
[52] L. Qi, X. Qin, F. Gao, J. Dong, and X. Gao, "SAWU-Net: Spatial attention weighted unmixing network for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, May 2023, Art. no. 5505205.
[53] Z. Yang, M. Xu, S. Liu, H. Sheng, and H. Zheng, "Spatial-spectral attention bilateral network for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Jul. 2023, Art. no. 5507505.
[54] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2021, Art. no. 5501916.
[55] Z. Shu, Z. Liu, J. Zhou, S. Tang, Z. Yu, and X.-J. Wu, "Spatial–spectral split attention residual network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 419–430, 2023.
[56] X. Mei et al., "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 963.
[57] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[58] B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, Jul. 2010.
[59] 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.
[60] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, May 2021.
[61] K. Yang, H. Sun, C. Zou, and X. Lu, "Cross-attention spectral–spatial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5518714.
[62] X. Zhang and Z. Wang, "Spatial proximity feature selection with residual spatial–spectral attention network for hyperspectral image classification," *IEEE Access*, vol. 11, pp. 23268–23281, 2023.
[63] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 67.

Koushikey Chhapariya (Graduate Student Member, IEEE) is currently working toward the Ph.D. degree in the satellite image processing lab with the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India.

She is currently building a robust deep-learning model for object detection on hyperspectral images, considering the spatio-spectral aspects. Her research focuses on the processing and analysis of satellite images, such as hyperspectral, multispectral, and thermal data.

Dr. Chhapariya was the recipient of the Raman-Charpak Fellowship 2022, the Fulbright Nehru Doctoral Research Fellowship 2023, and the Prime Minister's Research Fellowship (PMRF) conferred by the Ministry of Education, Government of India.

Krishna Mohan Buddhiraju (Member, IEEE) received the Ph.D. degree in electrical engineering from the Indian Institute of Technology Bombay (IITB), Mumbai, India, in 1991.

He is currently a Professor with the Centre of Studies in Resources Engineering (CSRE), IITB. He supervised more than 60 master's and Ph.D. theses. He has authored or coauthored more than 150 papers published in international conferences and journals. His research interests include multispectral and hyperspectral image processing and analysis, and machine learning.

Dr. Buddhiraju was the recipient of M.N. Saha Memorial Gold Medal for best application paper published in the *IETE Journal of Research* in September 2000 and the Indian Society of Remote Sensing National Geospatial Award for Excellence for 2012.

Anil Kumar received his B.Tech. degree in civil engineering from IET, his M.E. degree, and his Ph.D. in soft computing from the Indian Institute of Technology, Roorkee, India.

He is currently a Scientist/Engineer "SG" and the Head of the Photogrammetry and Remote Sensing Department, Indian Institute of Remote Sensing (IIRS), ISRO, Dehradun, India. He has guided several dissertations of Ph.D., M.Tech., M.Sc., B.Tech., and postgraduate diploma courses. He is the author of the two books "*Fuzzy Machine Learning Algorithms for Remote Sensing Image Classification*" and "*Multi-Sensor and Multi-Temporal Remote Sensing- Specific Single Class Mapping*" (CRC Press). His current research interests include soft-computing-based machine learning, deep learning for single date and temporal, multi-sensor remote-sensing data for specific-class identification, mapping through the in-house development of the SMIC tool, digital photogrammetry, GPS/GNSS, and LiDAR.

Dr. Kumar was the recipient of the Pisharoth Rama Pisharoty Award for contributing state-of-the-art fuzzy-based algorithms for Earth-observation data.