形态学的

Spectral–Spatial Morphological Attention Transformer for Hyperspectral Image Classification

Swalpa Kumar Roy, *Student Member, IEEE*, Ankur Deria, Chiranjibi Shah, *Member, IEEE*, Juan M. Haut, *Senior Member, IEEE*, Qian Du, *Fellow, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*— In recent years, convolutional neural networks (CNNs) have drawn significant attention for the classification of hyperspectral images (HSIs). Due to their self-attention mechanism, the vision transformer (ViT) provides promising classification performance compared to CNNs. Many researchers have incorporated ViT for HSI classification purposes. However, its performance can be further improved because the current version does not use spatial–spectral features. In this article, we present a new morphological transformer (morphFormer) that implements a learnable spectral and spatial morphological network, where spectral and spatial morphological convolution operations are used (in conjunction with the attention mechanism) to improve the interaction between the structural and shape information of the HSI token and the `CLS` token. Experiments conducted on widely used HSIs demonstrate the superiority of the proposed morphFormer over the classical CNN models and state-of-the-art transformer models. The source will be made available publicly at https://github.com/mhaut/morphFormer.**

***Index Terms*— Classification, hyperspectral images (HSIs), morphological transformer (morphFormer), spatial–spectral features.**

I. Introduction

HYPERSPECTRAL images (HSIs) contain information in contiguous wavelengths [1], [2], [3]. HSIs have been adopted in many application areas of remote sensing (RS) and Earth observation (EO), such as urban planning, vegetation monitoring, and crop management [4], [5], [6]. HSIs have particularly been used in EO tasks, such as desertification or climate change studies. In addition to land cover classification tasks [2], [7], [8], [9], other areas in which HSIs have been widely exploited include forestry [10], target/object detection, mineral exploration, and mapping [11], [12], environmental monitoring [13], disaster risk management, and biodiversity conservation. The popularity of HSIs is due to rich spectral and spatial information [14].

From the point of view of RS imaging technology, the affinity of spectral and spatial resolution is quite critical [15]. Spatial resolution is often limited by the very high spectral resolution of HSIs, and it may negatively affect land cover classification for complex scenes. For example, hyperspectral (HS) data do not provide proper information about the elevation and size of different structures of interest in particular application domains [14], [16]. Most conventional classifiers often process HSIs depending on spectral information and disregard spatial information among adjacent pixels. To solve this issue, different techniques can be implemented to incorporate both spatial and spectral information. With spatial processing, the size and shape of different objects can be determined resulting in better classification accuracy. In the following, we summarize some of the most relevant methods for exploiting HSI data, outlining their pros and cons.

In HSI classification, conventional classifiers have been widely utilized, even in the presence of limited training samples [3], [17], [18]. In general, these techniques include two stages. First, they reduce the dimensionality of the HSI data and extract some informative features. Then, spectral classifiers are fed with such features for classification purposes [2], [7], [19], [20], [21], [22]. In scenarios with limited training samples, support vector machines (SVMs) with nonlinear kernels have been widely used [23]. Moreover, the extreme learning machine (ELM) has been broadly used to extract features from unbalanced training sets. Li et al. [24] implemented an ELM to classify HSIs by extracting local binary patterns (LBPs) for classification. They demonstrated that ELMs can obtain better classification results than SVMs. The random forest (RF) was also utilized for the classification of HSIs due to its discriminative power [2]. However, the aforementioned classifiers face challenges when the training data are not representative, suffering from data fitting problems. This is because these classifiers consider HSIs as an assembly of measurements in the spectral domain, without considering their arrangement in the spatial domain. Classifiers based on spatial–spectral information significantly enhance the results of spectral-based classifiers with the inclusion of spatial data, such as the size and shape of various objects. In addition,

Manuscript received 12 December 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 3 February 2023; date of current version 23 February 2023. This work was supported in part by the Consejeria de Economia, Ciencia y Agencia Digital de la Junta de Extremadura, and Fondo Europeo de Desarrollo Regional de la Union Europea under Reference Grant GR21040; in part by the Spanish Ministerio de Ciencia e Innovacion under Project PID2019-110315RB-I00 (APRISA); in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 734541 (EOXPOSURE); and in part by the Science and Engineering Research Board (SERB), Government of India, under Project Grant SRG/2022/001390. *(Corresponding author: Antonio Plaza.)*

Swalpa Kumar Roy is with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri 735102, India (e-mail: swalpa@cse.jgec.ac.in).

Ankur Deria is with the Department of Informatics, Technical University of Munich, 85748 Garching bei München, Germany (e-mail: i.am.ankur.deria@tum.de).

Chiranjibi Shah and Qian Du are with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: cs3532@msstate.edu; du@ece.msstate.edu).

Juan M. Haut and Antonio Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain (e-mail: juanmariohaut@unex.es; aplaza@unex.es).

Digital Object Identifier 10.1109/TGRS.2023.3242346

spectral-based classifiers are more sensitive to noise compared to their spatial–spectral counterparts [2], [25].

Deep learning (DL) methods have attracted significant attention for multimodal data integration [26] in RS data classification [27]. A wide variety of fragmented datasets can be intelligently analyzed with DL methods. More recently, a unified and general DL framework was developed by Hong et al. [28] for classification of RS imagery. 1-D convolutional neural networks (CNNs) (CNN1Ds) [29], 2-D CNNs (CNN2Ds) [30], and 3-D CNNs (CNN3Ds) [31] have demonstrated success in the classification of HSI data.

Residual networks (ResNets) were introduced by He et al. [32]. These models have a minimum loss of information after each operation of the convolutional layers to reduce the gradient vanishing problem [32]. Zhong et al. [33] introduced a spatial–spectral ResNet (SSRN) for utilizing both spatial and spectral information to obtain enhanced classification performance. Roy et al. adopted a lightweight paradigm with the extraction of spatial and spectral features via the squeeze-and-excitation ResNet that can be added with a bag-of-features learning mechanism to accurately obtain the final classification results [34], [35]. Zhu et al. [36] incorporated other channel and spatial attention layers inside the SSRN architecture for extracting discriminative features. To take full advantage of ResNets, they can be extended to form even more complex models, such as the inclusion of adaptive kernels [17], lightweight spatial–spectral attention based on squeeze-and-excitation [35], and pyramidal ResNets [37]. Rotation-equivariant CNNs [38], gradient centralized convolutions [1], [39], and lightweight heterogeneous kernel convolutions [40] also enable efficient classification and feature extraction. Generative adversarial networks (GANs), on the other hand, may help with mitigating the class-imbalance problem in HSI classification [41], [42].

Despite their apparent ability to extract contextual information in the spatial domain, CNNs cannot easily sequentially incorporate attributes, in particular, long- and middle-term dependencies. As a consequence, their performance in HSI classification may be affected by the presence of classes with similar spectral signatures, making it difficult to extract diagnostic spectral attributes. The spectral signatures in HSIs can also be modeled using recursive neural networks (RNNs), which accumulate them in a band-by-band fashion. This is important to learn long-term dependencies, as the gradient vanishing problem may further complicate the interpretation of spectrally salient changes [43]. However, RNNs are not suitable for the simultaneous training of models because HSIs generally contain many samples, which limits classifier performance. Our work addresses the aforementioned limitations by rethinking HSI data classification using transformers.

As cutting-edge backbone networks, transformers utilize self-attention techniques to process and analyze sequential data more efficiently [44]. In recent years, several new transformer models have been developed including SpectralFormer [45], which is capable of learning spectral information by creating a transformer encoder module and utilizing adjacent bands. Transformers excel at characterizing spectral signatures, yet they are not able to model local semantic elements or utilize spatial information effectively. He et al. [46] proposed a bidirectional encoder representation for a transformer that incorporates flexible and dynamic input regions for pixel-based classification of HSIs. Zhong et al. [47] proposed a factorized architecture search (FAS) framework, which enables a stable and fast spectral–spatial transformer architecture search subject to find out the optimal architecture settings for the HSI classification task. To further improve the classification performance of HSIs, Sun et al. [48] introduced spatial and spectral tokenization of feature representations in the encoder, which helps to extract local spatial information and establish long-range relations between neighboring sequences. Yang et al. [49] utilize an adaptive 3-D convolution projection module to incorporate spatial–spectral information in an HSI transformer classification network. The above transformer models are designed based on HSI data and utilize spectral–spatial feature representation mechanisms. Roy et al. [50] recently developed a multimodal fusion transformer (MFT) to extract features from HSIs and fuse them with a `CLS` token derived from light detection and ranging (LiDAR) data to enhance the joint classification performance.

Mathematical morphology (MM) is a theory to analyze geometrical structures, based on topology, lattice theory, set theory, and random functions. Researchers have utilized MM-based techniques such as attribute profiles (APs) and extended morphological profiles (EPs) to extract spatial features and classify HSI data more accurately [16], [51], [52]. Rasti et al. [53] applied total variation component analysis for feature fusion to improve the joint extraction of EPs. Merentis et al. [54] used an RF classifier to classify HSI data with an automated fusion approach. By exploiting APs and EPs, MM has been successfully applied to extract features from RS data [55], [56], [57], [58]. In EPs and APs, several handcrafted characteristics are collected by sequentially performing dilation and erosion operations using an extensive set of structuring elements (SEs). There are a few limitations common to both EPs and APs, however. Specifically, the shape of the SE is fixed. In addition, the SEs can only obtain information about the size of existing objects but are unable to collect information about the shape of arbitrary item boundaries in complicated environments. To circumvent these restrictions, Roy et al. [3] introduced a spectral–spatial CNN based on morphological erosion and dilation operations for HSI classification. In this work, a spatial and spectral morphological block was created for extracting discriminative and robust spatial and spectral information from HSIs using its own trainable SEs in the erosion and dilation layers.

Although MM has been successfully applied in RS for extracting the spatial information based on techniques such as EPs or APs, the SEs are nontrainable [55], [56], [57], [58] and unable to capture dynamic features. If the EPs or APs are replaced with learnable MM operations, the resulting networks can be more capable of learning subtle features. Conventional transformer models use self-attention to highlight the most important features. If MM operations are combined with the transformer, the model may be able to learn intrinsic shape information and use this information in

the self-attention block for better feature extraction, leading to higher classification accuracies.

With the aforementioned rationale in mind, a new morphological fusion transformer encoder is introduced in this work, where the input patch is passed through two different morphological blocks simultaneously. The results provided by these blocks are concatenated, and the CLS token is added to the concatenated patch. The objective of our morphological transformer (morphFormer) model is to learn the spectral–spatial information from the patch embeddings of the HSI inputs, as well as to enrich the description of the abstract provided by the CLS token without adding significant computational complexity.

The main contributions of this work can be summarized as follows.

1) We provide a new learnable classification network based on a spectral–spatial morphFormer that conducts spatial and spectral morphological convolutions via dilation and erosion operators.
2) We introduce a new attention mechanism for efficiently fusing the existing CLS tokens and information obtained from HSI patch tokens into a new token that carries out morphological feature fusion.
3) We conduct experiments on four public HSI datasets by comparing the proposed network with other state-of-the-art approaches. The obtained results reveal the effectiveness of the proposed approach.

The remainder of this article is organized as follows. Section II describes the proposed method in detail. Section III discusses our experimental results. Section IV concludes this article.

II. Proposed Method

A. Convolutional Networks for Feature Learning

CNNs exhibit promising performance in HSI classification due to their ability to automatically extract contextual features. Since HSIs have numerous spectral bands, it is possible to take advantage of CNNs for controlling the depth of the output feature maps. CNNs have already been proved to be effective in capturing high-level features independently of the data source modality. Our proposed model uses CNNs for extracting high-level abstract features to be used by the transformer. The spectral dimensions of the HSI are reduced by the CNN.

Our proposed model utilizes sequential layers of Conv3D and HetConv2D for extracting robust and discriminative features from HSIs. The original data are arranged in subcubes $\mathbf{X}_{\text{HSI}}$ (with dimensionality $11 \times 11 \times B$) that are reshaped into $(1 \times 11 \times 11 \times B)$ and used as input to a Conv3D layer with kernel size $(3 \times 3 \times 9)$ and padding $(1 \times 1 \times 0)$. Padding is used so that the spatial size of the output image is the same as that of the input image. The HetConv2D block follows the Conv3D layer and consists of two Conv2D layers working in parallel. One of the Conv2D layers is used for groupwise convolution, and the other one is used for pointwise convolution. HetConv2D utilizes two kernels of different sizes to extract multiscale information. The outputs obtained from these two convolutions are combined in an elementwise fashion ($\oplus$) and returned as output

$$\begin{aligned} X_{\text{in}} &= \text{Reshape}(\text{Conv3D}(\text{Reshape}(X_{\text{HSI}}))) \\ X_{\text{out}} &= \text{Conv2D}(X_{\text{in}}, k1, g1, p1) \oplus \text{Conv2D}\,(X_{\text{in}}, k2, g2, p2) \end{aligned} \tag{1}$$

where $k1 = 3$, $g1 = 4$, $p1 = 1$, $k2 = 1$, $g2 = 1$, and $p2 = 0$. The output shape of the Conv3D layer is $(8 \times 11 \times 11 \times (B-8))$, and that of the HetConv2D block is $(11 \times 11 \times 64)$. Batch normalization (BN) [59] and ReLU activation layers are used after the Conv3D layer and the HetConv2D block. If only a few limited training samples are available, the overfitting phenomenon may arise. To address this issue and accelerate the training performance, we use a BN. ReLU also helps in smoothing the back-propagation of the loss by introducing nonlinearity.

B. Image Tokenization and Position Embedding

HSIs contain spatial and spectral features which can provide highly discriminative information that can lead to higher classification accuracies. Patch tokens of shape (1×64) each are obtained by flattening HSI subcubes of shape $[(11 \times 11) \times 64]$ as follows:

$$X_{\text{flat}} = T(\text{Flatten}(X_{\text{out}})) \tag{2}$$

where $T(\cdot)$ is a transpose function and $X_{\text{flat}} \in \mathcal{R}^{121\times 64}$. The tokenization [48] operation is used to select n from 121 patches as follows:

$$\begin{aligned} X_{W_a} &= \text{softmax}(T(X_{\text{flat}}.W_{\text{aH}})) \\ X_{W_b} &= X_{\text{flat}}.W_{\text{bH}} \end{aligned} \tag{3}$$

where $W_{\text{aH}} \in \mathcal{R}^{64\times n}$, $W_{\text{bH}} \in \mathcal{R}^{64\times 64}$, $X_{W_a} \in \mathcal{R}^{n\times 121}$, and $X_{W_b} \in \mathcal{R}^{121\times 64}$. The tokenization operation uses two learnable weights to extract the key features

$$X_{\text{patch}} = X_{W_a}.X_{W_b} \tag{4}$$

where $X_{\text{patch}} \in \mathcal{R}^{n\times 64}$. A total of $(n+1)$ patches are obtained as described in (5) by concatenating ($\odot$) the CLS token to the HSI patch tokens. The CLS token (X_{cls}) is a learnable tensor, which is randomly initialized. To simplify the calculation of head dimensions, a size of 64 is used

$$\widehat{X} = \left[X_{\text{cls}} \odot \widehat{X}_{\text{patch}}\right]. \tag{5}$$

The semantic textural information in the image patch tokens can be preserved by adding trainable position embeddings to the patch embeddings. Hence, a trainable position embedding is added to the created HSI patch tokens. Fig. 1 graphically illustrates the addition of position embeddings (in elementwise fashion) to the patches (1 to $n+1$). A dropout layer is used after this operation to reduce the effect of the vanishing gradient. The above procedure can be expressed as

$$X = \mathcal{DP}\left(\widehat{X} \oplus \mathcal{PE}\right) \tag{6}$$

where $\mathcal{DP}$ denotes a dropout layer with value of 0.1 and $\mathcal{PE}$ represents a learnable position embedding.

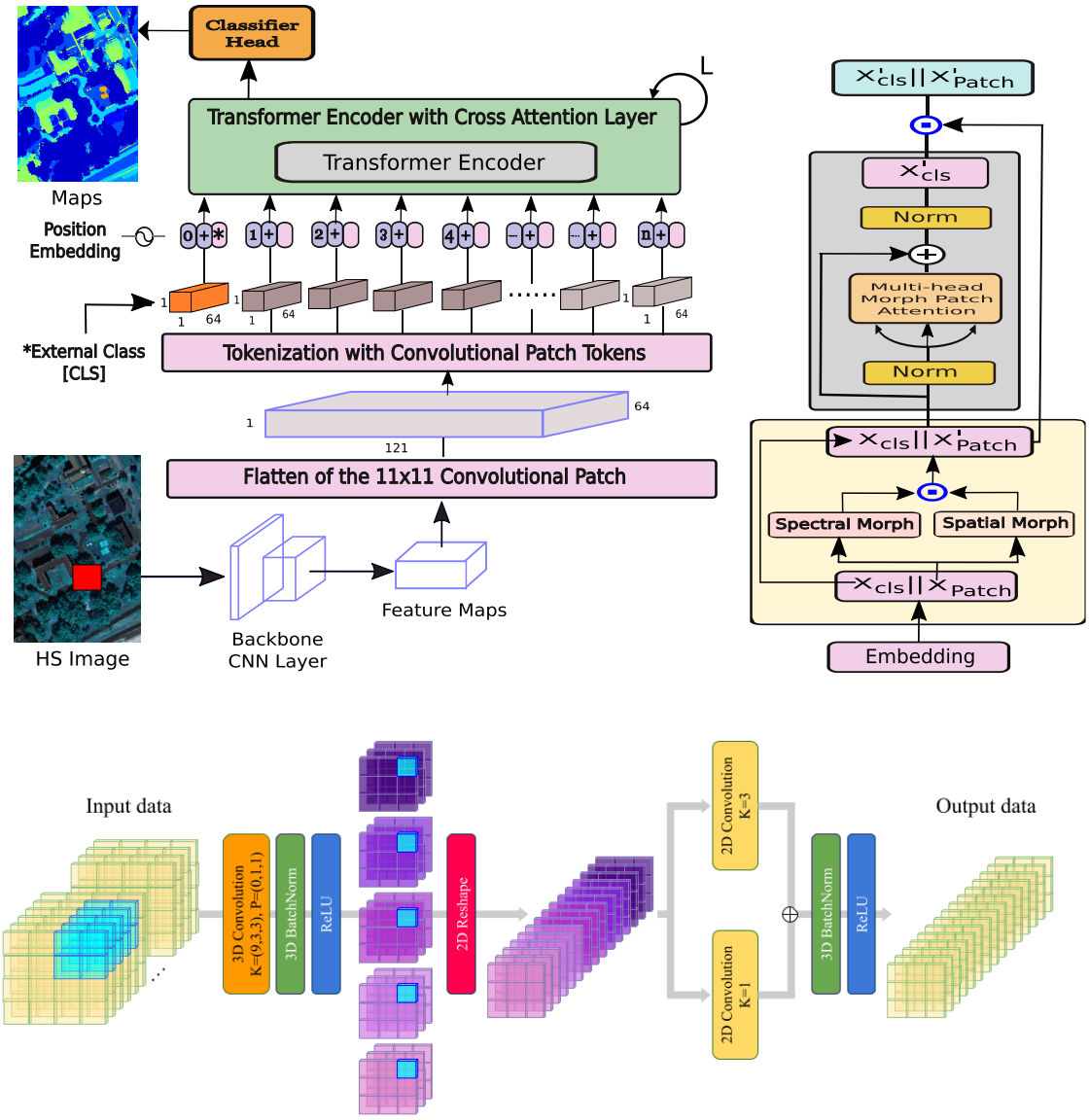


Fig. 1. In the upper row, we show the proposed HSI classification network such that the classification map of the proposed method contains (left) less noise than existing methods and (right) transformer encoder with a multihead patch attention mechanism. In the bottom row, we show the backbone of the proposed method.

C. Spectral and Spatial Morphological Convolutions

MM is a powerful technique for characterizing the intrinsic shape, structure, and size of objects in an image. The spectral and spatial morphological network presented here is designed based on dilation and erosion operations with SEs of size $(s \times s)$.

A dilated image is produced by combining the input HSI patch tokens with SEs, selecting the pixel with the maximum value in the local neighborhood. As a result of the dilation procedure, the boundaries of the foreground objects of the HSI input patch token are broadened. In other words, the size of the kernel affects the size of the texture for various regions of an HSI patch token. The dilation process is represented by $\boxplus$ and can be denoted by the following equation:

$$\left(X_{\text{patch}} \boxplus W_d\right)(x, y) = \max_{(i,j)\in\psi} \left(X_{\text{patch}}(x+i, y+j) + W_d(i, j)\right) \quad (7)$$

where $\psi = \{(i, j) \mid i \in \{1, 2, 3, \ldots, s\};\ j \in \{1, 2, 3, \ldots, s\}\}$ represents the elements of the kernel and W_d denotes the SEs used for the dilation operation.

Regarding the erosion operation, the output of the convolution with the SE selects the pixel with minimum value in the local neighborhood. This operation reduces the shape of the background object in the HSI patch token (as opposed to the dilation). Erosions can eliminate minor details and enlarge holes, making them distinguishable from each other in different texture regions. Let $X_{\text{patch}} \in R^{k\times k}$ be an input HSI patch token of spatial size $k \times k$, and let $\boxminus$ represent the morphological erosion operation. The erosion operation can be defined as

$$\left(X_{\text{patch}} \boxminus W_e\right)(x, y) = \min_{(i,j)\in\psi} \left(X_{\text{patch}}(x+i, y+j) - W_e(i, j)\right) \quad (8)$$

where $\psi = \{(i, j) \mid i \in \{1, 2, 3, \ldots, s\};\ j \in \{1, 2, 3, \ldots, s\}\}$ represents the elements of the kernel and W_e denotes the SEs

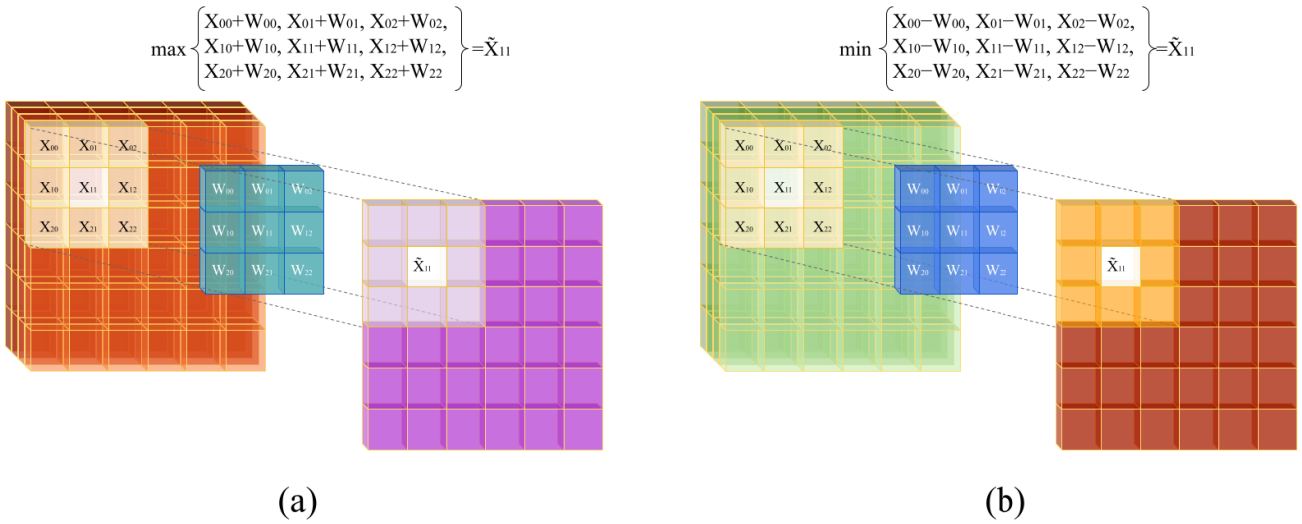


Fig. 2. Graphical visualization of (a) dilation and (b) erosion operations for an input image patch of size (7 × 7), dilated, and eroded with an SE of size (3 × 3). The resulting outputs keep the same size using a padding technique.

used for the erosion operation. It can be understood from the operations defined in (7) and (8) that the HSI patch tokens are shifted by $i \times j$, as in the convolutional operation. The padding function is used to keep the input and output shapes of the objects. After applying the operations given in (7) and (8) to the HSI patch tokens, dilation and erosion maps can be, respectively, obtained.

A graphical visualization of the input and output images after the dilation and erosion operations (using an SE of size 3 × 3) is shown in Fig. 2(a) and (b). To obtain the morphological shape feature from the HSI patch tokens, a spatial morphological (`SpatialMorph`) block with primitive operations (e.g., dilation and erosion) is used. The `SpatialMorph` block comprises parallel branches of dilation and erosion, followed by their respective convolutional operations, and finally, the results from both branches are combined in an elementwise fashion. As morphological operations are nonlinear, they can generate a discrepancy in the learned features. In order to normalize those learned features, convolutional operations are used. The entire `SpatialMorph` block can be described as

$$\mathcal{F}_{\text{SpatMorph}}\left(X_{\text{patch}}\right) = F_{2\text{D}}\left(\left(X_{\text{patch}} \boxplus W_d\right)\right) \oplus F_{2\text{D}}\left(\left(X_{\text{patch}} \boxminus W_e\right)\right) \quad (9)$$

where W_d and W_e are the weights of the (3 × 3) kernel and $F_{2\text{D}}$ is the function that represents the linear combination between the dilation and erosion feature maps obtained utilizing the 2-D convolution. To obtain the morphological spectral feature from the HSI patch tokens, a spectral morphological (`SpectralMorph`) block using the primitive operations is used. This block can be described using the following equation:

$$\mathcal{F}_{\text{SpecMorph}}\left(X_{\text{patch}}\right) = F_{2\text{D}}\left(\left(X_{\text{patch}} \boxplus W_d\right)\right) \oplus F_{2\text{D}}\left(\left(X_{\text{patch}} \boxminus W_e\right)\right) \quad (10)$$

where W_d and W_e are the weights of the (1 × 1) kernel and $F_{2\text{D}}$ is the function that represents the linear combination between the dilation and erosion feature maps obtained utilizing the 2-D convolution.

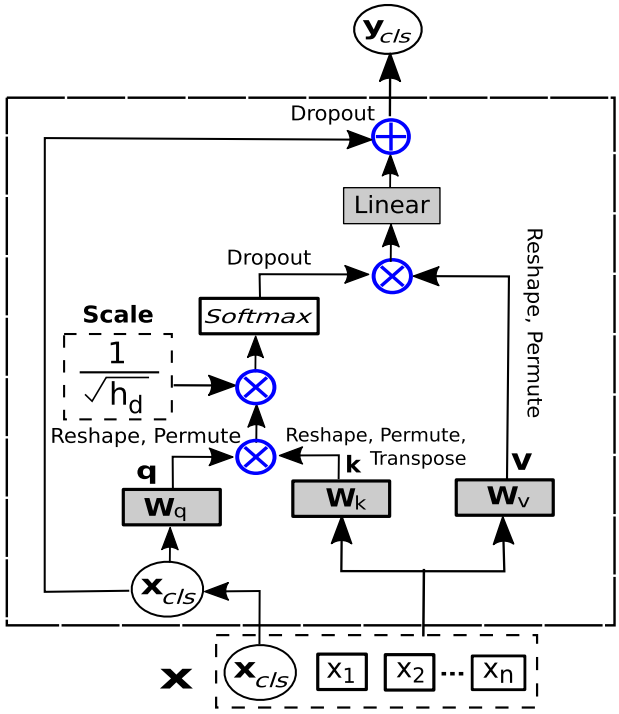


Fig. 3. Multihead patch attention module, where a query value interacts with all the other HSI patch tokens through the attention mechanism.

D. Patch Attention Using Morphological Feature Fusion

As shown in Fig. 3, the `CLS` (X_{cls}) token uses the HSI patch tokens for exchanging information between each other to provide an abstract representation of the whole HSI patch. This entire operation is executed in blocks of the transformer encoder, where each transformer block consists of a spectral and spatial morphological feature extraction block and a residual multihead cross attention block. The spectral and spatial morphological feature extraction block consists of a spectral morphological layer and a spatial morphological layer, both of which take X_{patch} as input. The spectral morphological layer is used to extract morphological spectral features from

the HSI data, whereas the spatial morphological layer is used to extract morphological spatial features using two primitive morphological operations: dilation and erosion. The spatial and spectral morphological features allow for better attention between the intrinsic spatial and spectral characteristics of the image. The outputs from both layers are then concatenated in channelwise form (X'_{patch}) along with X_{cls} to generate the final output of the entire morphological block, as shown in Fig. 1. The output channel from both the spectral and spatial morphological blocks is half of the input X_{patch} so that, after concatenating both of them, the number of channels becomes equal to that of X_{patch}. The entire morphological block can be summarized as follows:

$$\begin{aligned} X'_{\text{patch}} &= \mathcal{F}_{\text{SpatMorph}}(X_{\text{patch}}) \odot \mathcal{F}_{\text{SpecMorph}}(X_{\text{patch}}) \\ X' &= X_{\text{cls}} \odot X'_{\text{patch}}. \end{aligned} \tag{11}$$

On the other hand, a layer normalization (LN) operation is used in the residual attention block. It takes the output from the morphological block as input. A self-attention layer is used after the LN operation, whose output y_{cls} is added in elementwise fashion ($\oplus$) to the input of the LN (as described in Fig. 1).

In the morphological patch attention module (`MorphPAT`) between X_{cls} and X'_{patch}, three linear weights, i.e., **Q**, **K**, and **V**, are used. They are multiplied inside the morphological attention block and can be represented as

$$Z = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{h_d}}\right)\mathbf{V} \tag{12}$$

where $Z \in \mathcal{R}^{1\times 64}$, h_d is the embedding dimension/number of heads, **Q** is the query (which equals $\mathbf{X}_{\text{cls}}\mathbf{W}_q$ where $W_q \in \mathcal{R}^{64\times 64}$), **K** is the key (which equals $\mathbf{X}'_{\text{patch}}\mathbf{W}_k$ with $W_k \in \mathcal{R}^{64\times 64}$), and **V** is the value (which equals $\mathbf{X}'_{\text{patch}}\mathbf{W}_v$ with $W_v \in \mathcal{R}^{64\times 64}$). A dropout layer ($\mathcal{DP}$) with a value of 0.1 is used, followed by a linear projection layer ($\mathbf{W}_l \in \mathcal{R}^{64\times 64}$) that is applied to the final output of the `qkv` operation. A self-attention module with a number of heads greater than one becomes a multihead self-attention module. Similarly, the `MorphAT` module (upon using multiple heads) becomes a multihead morphological attention module and can be represented as `MMorphAT`. Mathematically, the morphological attention module can be formulated as

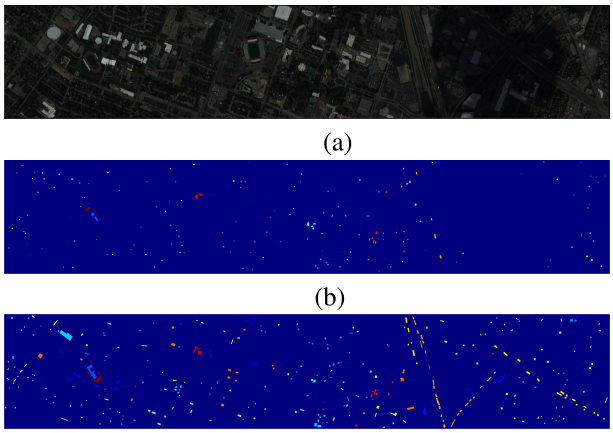
$$\text{MMorphAT}(X') = \mathcal{DP}(\mathbf{W}_l\mathbf{Z}). \tag{13}$$

The output X'_{cls} of the `MMorphAT` module for a given X'_{k-1}, where k is the kth transformer encoder block, can be expressed as

$$\begin{aligned} y_{\text{cls}} &= \text{MMorphAT}\big(\text{LN}\left(X'_{k-1}\right)\big) \\ X'_{\text{cls}} &= \text{LN}\big(y_{\text{cls}} \oplus X'_{k-1\text{cls}}\big) \end{aligned} \tag{14}$$

where $X'_{\text{cls}} \in \mathcal{R}^{1\times 64}$. This output X'_{cls} is then concatenated with X'_{patch} to yield the final output of that particular transformer encoder block, as shown in Fig. 1, and can be defined as

$$X'_k = X'_{\text{cls}} \odot X'_{\text{patch}}. \tag{15}$$



Color	Class	Train	Test	Color	Class	Train	Test
	Background	662013	652648		Grass-healthy	198	1053
	Grass-stressed	190	1064		Grass-synthetic	192	505
	Tree	188	1056		Soil	186	1056
	Water	182	143		Residential	196	1072
	Commercial	191	1053		Road	193	1059
	Highway	191	1036		Railway	181	1054
	Parking-lot1	192	1041		Parking-lot2	184	285
	Tennis-court	181	247		Running-track	187	473

Fig. 4. UH data. (a) Pseudocolor image using bands 64, 43, and 22. (b) Disjoint train samples. (c) Disjoint testing samples. The table shows land-cover types for each class along with the number of disjoint train and test samples.

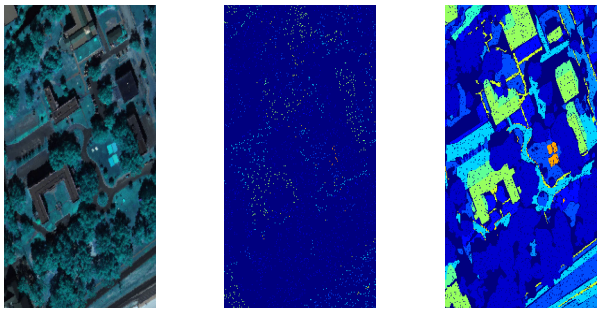
The proposed model uses eight heads. Finally, the `CLS` token is extracted from the output of the transformer encoder blocks (X_k), and the final classification results are obtained from the `CLS` token via a classifier head.

III. Experiments

For evaluating the classification performance of the proposed morphFormer, we have considered four different datasets and compared our approach with other state-of-the-art techniques. The datasets utilized in experiments were collected from the University of Houston (UH), the University of Southern Mississippi Gulfpark (MUUFL), and the cities of Trento and Augsburg.

A. Image Datasets

1) In the experiments, four HSI datasets, i.e., UH, MUUFL, Trento, and Augsburg are used. The UH data were gathered by the Compact Airborne Spectrographic Imager (CASI) in 2013 and published by the IEEE Geoscience and Remote Sensing Society. The image consists of 340 × 1905 pixels and 144 different spectral bands. Its wavelength range is 0.38–1.05 μm with a spatial resolution of 2.5 meters per pixel (MPP). Its ground truth consists of 15 distinct classes. Total samples are separated into 15 distinct classes by disjoint train and test samples. Fig. 4 lists the disjoint train and test samples for each of the 15 distinct classes of land cover.
2) The MUUFL data were obtained in November 2010 around the region of the University of Southern Mississippi Gulf Park, Long Beach, MS, USA, by using the Reflective Optics System Imaging Spectrometer



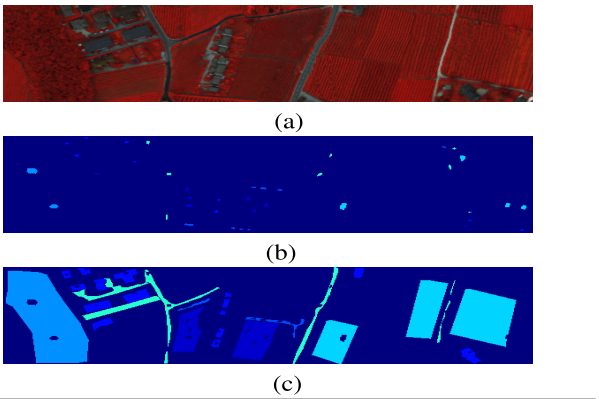
(a) (b) (c)

Color	Class	Train	Test	Color	Class	Train	Test
	Background	68817	20496		Trees	1162	22084
	Grass-Pure	214	4056		Grass-Groundsurface	344	6538
	Dirt-And-Sand	91	1735		Road-Materials	334	6353
	Water	23	443		Buildings'-Shadow	112	2121
	Buildings	312	5928		Sidewalk	69	1316
	Yellow-Curb	9	174		ClothPanels	13	256

Fig. 5. MUUFL data. (a) Pseudocolor image using bands 40, 20, and 10. (b) Disjoint train samples. (c) Disjoint testing samples. The table shows land-cover types for each class along with the number of disjoint train and test samples, where the train samples represent 5% of the available ground truth, and the test samples represent the remaining 95% of the ground truth.

(ROSIS) sensor [60], [61]. It is made up of 325 × 220 pixels along with 72 spectral bands, and LiDAR data are also available and made up of elevation data from two rasters. The first and last eight bands are deleted owing to noise, resulting in 64 bands in total. There are 53 687 ground-truth pixels with 11 different types of classes for urban land cover. 5% of the samples are randomly selected for training from each of the 11 classes, as shown in Fig. 5.

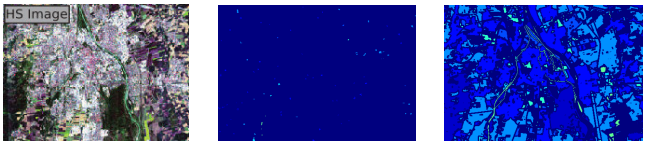
3) The Trento data were collected around rural areas in the south of Trento, Italy, by utilizing the AISA eagle sensor. The corresponding LiDAR data were obtained by the Optech ALTM 3100EA sensor. The HS image comprises 63 different spectral channels with wavelengths ranging from 0.42 to 0.99 μm, whereas the LiDAR data have two rasters with elevation data. The HSI data include 600 × 166 pixels with 6 mutually exclusive land-cover classes of vegetation, with a spatial resolution of 1 MPP and spectral resolution of 9.2 nm. Furthermore, the total samples are separated into six groups of disjoint train and testing samples. The information regarding the number of samples per class is given in Fig. 6.
4) The Augsburg scene includes three distinct data sources, including an HSI, a dual-Pol synthetic aperture radar (SAR) image, and a digital surface model (DSM) [62]. The HSI and DSM data were acquired by DLR, whereas the SAR data were collected from the Sentinel-1 platform over the city of Augsburg, Germany. The information was gathered with the HySpex sensor [63], the Sentinel-1 sensor, and the DLR-3 K system [64], respectively. For proper multimodal fusion, the spatial resolutions of all datasets were downsampled to a uniform spatial resolution of 30-m ground sampling distance (GSD). It has 332 × 485 pixels in the HSI, with 180 spectral channels that ranges from 0.4 to 2.5 μm. In the ground truth, there are 15 distinct classes of land cover. The train and testing sets are demonstrated in detail in Fig. 7.



(a) (b) (c)

Color	Class	Train	Test	Color	Class	Train	Test
	Background	98781	70205		Apples	129	3905
	Buildings	125	2778		Ground	105	374
	Woods	154	8969		Vineyard	184	10317
	Roads	122	3052				

Fig. 6. Trento data. (a) Pseudocolor image using bands 40, 20, and 10. (b) Disjoint train samples. (c) Disjoint testing samples. The table shows land-cover types for each class along with the number of disjoint train and test samples.



(a) (b) (c)

Color	Class	Train	Test	Color	Class	Train	Test
	Background	68816	20497		Forest	1162	22084
	Residential-Area	344	6538		Industrial-Area	91	1735
	Low-Plants	334	6353		Allotment	23	443
	Commercial-Area	112	2121		Water	312	5928

Fig. 7. Augsburg data. (a) Pseudocolor image using bands 40, 20, and 10. (b) Disjoint train samples. (c) Disjoint testing samples. The table shows land-cover types for each class along with the number of disjoint train and test samples.

B. Experimental Setting

Extensive experiments have been performed using the proposed morphFormer model, and the results have been compared with that of traditional and state-of-the-art models to assess the performance of the proposed model.

The compared techniques include traditional classifiers, such as RF [2], K-nearest neighbors KNN [65], and SVM [23], in addition to classical CNN methods, such as CNN1D [66], CNN2D [30], CNN3D [31], and RNN [67]. We also included state-of-the-art transformer-based techniques, such as vision transformer (ViT) [68] and SpectralFormer [45].

For testing, a CPU with a Red Hat Enterprise Server (Release 7.6) has been used that consists of the ppc64le architecture, 40 cores consisting of four threads in each core, and 377 GB of memory. The GPU utilized is a single Nvidia Tesla V100 having VRAM of 32 510 MB.

During our experiments, the number of HSI patch tokens (n) obtained from the tokenization process is taken as four. During training and testing, batch sizes of 64 and 500 were, respectively, utilized. Patches with a size of 11 × 11 × B are taken from the HSI and used as input to the model. Aside

TABLE I
CLASSIFICATION PERFORMANCE (IN %) ON THE UH HSI DATASET

Class No.	Conventional Classifiers			Classical Convolutional Networks				Transformer Networks		
	KNN	RF	SVM	CNN1D	CNN2D	CNN3D	RNN	ViT	SpectralFormer	morphFormer
1	77.87	82.81 ± 00.08	79.77	81.10 ± 00.43	80.53 ± 00.23	81.70 ± 00.38	80.22 ± 02.80	82.40 ± 00.39	**82.49 ± 00.25**	82.37 ± 00.63
2	77.44	82.86 ± 00.36	82.42	80.23 ± 01.51	83.90 ± 00.09	80.55 ± 01.84	78.51 ± 00.19	80.29 ± 00.79	**89.13 ± 06.36**	85.03 ± 00.12
3	96.83	63.10 ± 01.71	59.41	53.73 ± 00.09	57.49 ± 02.10	96.57 ± 00.09	52.94 ± 17.72	97.43 ± 01.32	69.77 ± 12.29	**98.09 ± 00.57**
4	75.28	91.95 ± 00.15	83.81	83.74 ± 00.71	89.46 ± 00.24	78.54 ± 02.82	83.81 ± 07.71	90.40 ± 00.25	91.73 ± 03.42	**95.23 ± 01.82**
5	90.72	99.78 ± 00.12	95.27	87.06 ± 01.51	92.36 ± 01.20	98.48 ± 00.28	85.61 ± 09.73	**99.24 ± 00.13**	96.78 ± 00.76	99.15 ± 00.95
6	66.43	96.97 ± 00.33	67.13	52.45 ± 01.14	64.10 ± 02.38	73.89 ± 00.66	70.16 ± 07.18	91.38 ± 00.87	85.31 ± 07.99	**98.14 ± 01.74**
7	76.96	85.23 ± 00.50	83.21	71.42 ± 01.46	71.39 ± 04.04	82.77 ± 01.45	73.01 ± 01.31	86.10 ± 01.54	80.25 ± 02.23	**89.61 ± 01.73**
8	30.96	42.58 ± 00.36	29.53	41.12 ± 01.70	44.95 ± 06.09	38.30 ± 01.09	43.84 ± 08.75	**73.95 ± 00.29**	62.74 ± 11.68	72.33 ± 00.45
9	69.50	85.36 ± 00.13	75.45	60.25 ± 00.08	62.45 ± 02.04	65.94 ± 02.86	68.84 ± 02.54	85.33 ± 02.24	70.57 ± 01.62	**88.86 ± 03.21**
10	42.95	35.81 ± 00.70	46.62	39.12 ± 02.03	49.94 ± 01.66	43.28 ± 07.46	37.52 ± 01.15	50.42 ± 05.59	48.17 ± 05.04	**61.97 ± 02.91**
11	56.17	63.03 ± 00.39	45.07	42.06 ± 00.78	44.53 ± 00.78	33.59 ± 02.72	49.65 ± 09.61	80.80 ± 03.06	62.75 ± 04.12	**96.24 ± 02.09**
12	75.79	66.63 ± 00.50	70.03	62.98 ± 04.58	53.92 ± 06.25	67.85 ± 02.10	64.07 ± 00.28	81.91 ± 01.77	79.09 ± 02.16	**94.46 ± 02.36**
13	60.35	87.60 ± 00.17	68.42	42.11 ± 01.52	47.13 ± 02.33	77.54 ± 01.98	53.92 ± 08.72	**89.47 ± 01.74**	63.63 ± 05.56	87.02 ± 02.91
14	76.92	99.73 ± 00.19	75.30	83.94 ± 00.38	82.46 ± 03.07	92.58 ± 01.06	81.38 ± 12.63	99.33 ± 00.95	93.66 ± 04.30	**99.73 ± 00.19**
15	88.37	85.62 ± 01.38	49.89	34.46 ± 02.04	42.92 ± 05.11	93.52 ± 01.55	44.12 ± 13.23	**99.72 ± 00.26**	77.24 ± 07.34	96.69 ± 04.24
OA	69.48	74.87 ± 00.09	68.13	63.04 ± 00.08	65.85 ± 00.09	70.26 ± 00.14	65.20 ± 02.74	83.23 ± 00.47	76.35 ± 01.97	**87.85 ± 00.20**
AA	70.84	77.94 ± 00.14	67.42	61.05 ± 00.10	64.50 ± 00.11	73.67 ± 00.17	64.51 ± 02.77	85.88 ± 00.33	76.89 ± 02.31	**89.66 ± 00.39**
$\kappa(\times 100)$	67.08	72.93 ± 00.09	65.56	60.01 ± 00.08	63.04 ± 00.09	67.91 ± 00.16	62.43 ± 02.87	81.88 ± 00.51	74.42 ± 02.14	**86.81 ± 00.22**

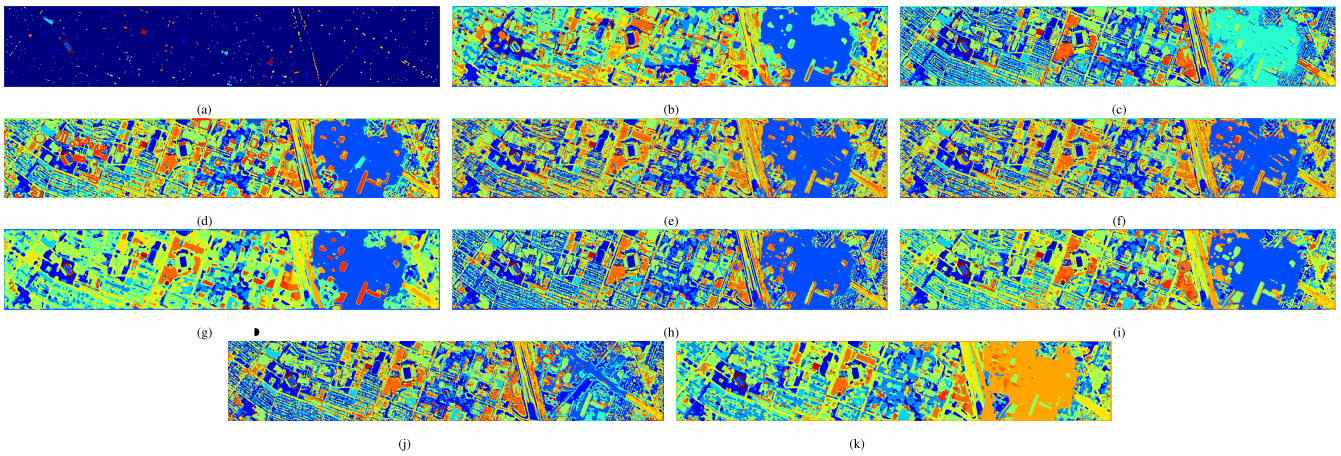


Fig. 8. Classification maps for the Houston (UH) HSI dataset. (a) Ground truth. (b) KNN (69.48%). (c) RF (74.87%). (d) SVM (68.13%). (e) CNN1D (63.04%). (f) CNN2D (65.85%). (g) CNN3D (70.26%). (h) RNN (65.20%). (i) ViT (83.23%). (j) SpectralFormer (76.35%). (k) morphFormer (87.85%).

from KNN, RF, SVM, and RNN, the Adam optimizer [69], [70] has been used to train the models, with a weight decay of $5e^{-3}$ and learning rate of $5e^{-4}$. In addition, these methods (considering also the RNN) used a step scheduler with a gamma of 0.9, steps of size 50, and trained during 500 epochs. The average and standard deviation of each experiment have been calculated based on three repetitions. Python 3.7.7 and PyTorch 1.5.0 were used to implement the coding of the proposed morphFormer.

Different widely utilized quantitative measurement methods, such as the overall accuracy (OA), average accuracy (AA), and kappa coefficients ($k$appa), are utilized for assessing the performance. The experiments have been performed on spectrally and spatially disjoint sets of train and testing samples [71] such that there is no interaction between the respective samples. In addition, varying percentages or train samples have been considered for validating the performance of the considered techniques.

C. Performance Analysis With Disjoint Train/Test Samples

A quantitative assessment of classification performance is presented in Tables I–IV. The best classification values are displayed in bold. The results show that the proposed approach is superior to all other techniques in terms of OA, AA, and ks, and exhibits better performance in most cases in terms of classwise accuracy.

It is worth noting that conventional classifiers, such as KNN, RF, or SVM, exhibit similar performance. An exception is the KNN with the MUFFL and Trento datasets, which provides inferior accuracies than those provided by RF and SVM. In addition, the performance of DL-based classifiers, such as CNN1D, CNN2D, CNN3D, and RNN, is generally superior to that of conventional classifiers, except for RF in UH and MUFFL datasets (which is better than CNN2D and CNN3D). Transformer methods, such as ViT and SpectralFormer, provide better performance due to the incorporation of the sequential mechanism. However, the incorporation of the spatial–spectral information in the proposed morphFormer leads to better classification performance in terms of OA, AA, and k in all considered datasets.

Table I shows that the RF provides better performance in the UH dataset in comparison to other conventional classifiers, but it cannot provide better performance than transformer methods. The proposed technique exhibits a performance that

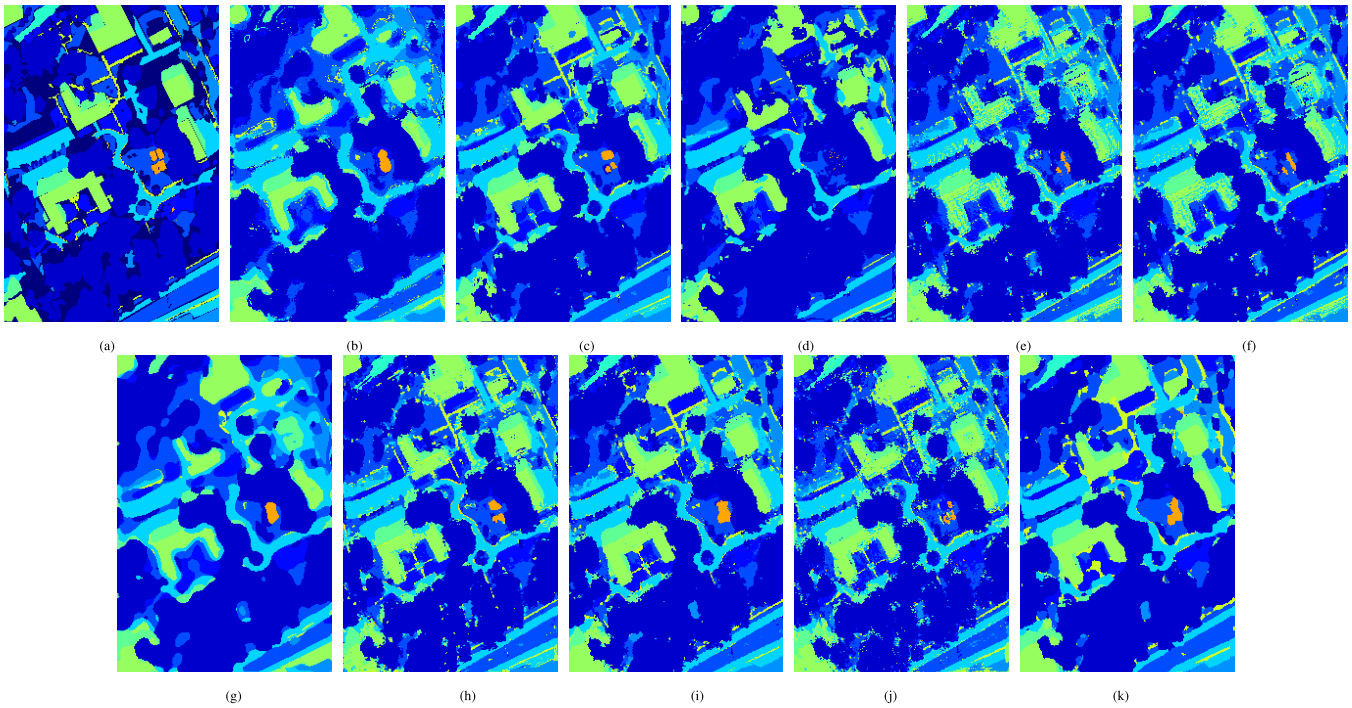


Fig. 9. Classification maps for the MUUFL HSI dataset. (a) Ground truth. (b) KNN (75.80%). (c) RF (89.85%). (d) SVM (84.30%). (e) CNN1D (81.17%). (f) CNN2D (82.95%). (g) CNN3D (77.59%). (h) RNN (88.60%). (i) ViT (91.99%). (j) SpectralFormer (86.68%). (k) morphFormer (93.84%).

TABLE II

CLASSIFICATION PERFORMANCE (IN %) ON THE MUUFL HSI DATASET

Class No.	Conventional Classifiers			Classical Convolutional Networks				Transformer Networks		
	KNN	RF	SVM	CNN1D	CNN2D	CNN3D	RNN	ViT	SpectralFormer	morphFormer
1	92.74	97.97 ± 00.07	96.64	95.21 ± 00.05	95.94 ± 00.20	94.72 ± 00.65	96.06 ± 00.19	97.40 ± 00.25	94.63 ± 00.89	**97.76 ± 00.25**
2	47.46	77.86 ± 00.19	59.37	69.31 ± 01.06	70.87 ± 01.50	62.98 ± 03.76	80.34 ± 02.85	77.34 ± 01.62	77.12 ± 03.77	**90.60 ± 00.71**
3	69.10	84.33 ± 00.23	81.49	74.31 ± 00.60	78.65 ± 00.78	71.15 ± 03.46	80.65 ± 00.87	86.10 ± 01.49	74.26 ± 02.81	**90.99 ± 00.08**
4	53.43	86.09 ± 00.37	74.35	78.71 ± 01.27	82.96 ± 01.17	64.86 ± 02.15	86.09 ± 01.95	**93.35 ± 00.84**	84.59 ± 01.18	91.89 ± 00.19
5	83.68	92.10 ± 00.08	83.80	75.84 ± 01.93	78.76 ± 01.09	77.95 ± 02.15	90.15 ± 00.39	**94.82 ± 00.77**	89.38 ± 02.06	94.54 ± 00.71
6	01.13	68.77 ± 00.65	15.35	46.35 ± 00.74	48.53 ± 02.60	75.24 ± 90.92	53.42 ± 04.33	82.69 ± 01.31	60.42 ± 09.08	**84.88 ± 01.87**
7	43.33	80.47 ± 00.42	77.09	78.45 ± 00.58	78.96 ± 00.30	48.44 ± 01.87	80.39 ± 02.75	87.44 ± 00.97	83.77 ± 03.58	**91.01 ± 00.48**
8	73.48	91.57 ± 00.17	87.11	68.28 ± 02.36	68.86 ± 00.37	69.48 ± 01.91	88.67 ± 02.04	**97.41 ± 00.75**	88.91 ± 02.78	96.67 ± 00.07
9	04.94	45.87 ± 00.31	21.50	39.97 ± 00.27	45.64 ± 00.48	04.79 ± 03.39	58.94 ± 00.73	58.92 ± 07.76	55.02 ± 01.43	**61.12 ± 02.05**
10	00.00	04.98 ± 03.05	00.00	08.81 ± 00.72	13.03 ± 00.27	00.00 ± 00.00	22.80 ± 04.51	**32.76 ± 07.01**	17.62 ± 02.12	16.67 ± 01.24
11	58.59	48.31 ± 02.12	61.33	23.83 ± 03.24	24.74 ± 00.97	61.20 ± 02.17	**80.99 ± 04.01**	66.67 ± 04.94	45.44 ± 07.29	69.92 ± 02.78
OA	75.80	89.85 ± 00.03	84.30	81.17 ± 00.09	82.95 ± 00.12	77.59 ± 00.05	88.60 ± 00.69	91.99 ± 00.35	86.68 ± 00.93	**93.84 ± 00.10**
AA	47.99	70.76 ± 00.29	59.82	59.92 ± 00.52	62.45 ± 00.22	50.58 ± 00.34	74.41 ± 01.40	79.54 ± 01.93	70.11 ± 01.52	**80.55 ± 00.27**
$\kappa(\times 100)$	67.34	86.44 ± 00.04	78.89	74.95 ± 00.12	77.30 ± 00.16	69.81 ± 00.10	84.90 ± 00.92	89.37 ± 00.46	82.43 ± 01.19	**91.84 ± 00.13**

TABLE III

CLASSIFICATION PERFORMANCE (IN %) ON THE TRENTO HSI DATASET

Class No.	Conventional Classifiers			Classical Convolutional Networks				Transformer Networks		
	KNN	RF	SVM	CNN1D	CNN2D	CNN3D	RNN	ViT	SpectralFormer	morphFormer
1	89.07	96.65 ± 00.36	97.21	97.29 ± 00.20	96.76 ± 00.24	92.22 ± 00.66	88.06 ± 04.36	89.07 ± 00.86	94.01 ± 03.77	**98.51 ± 00.84**
2	88.59	89.38 ± 00.74	94.17	86.73 ± 01.60	84.17 ± 01.05	**91.42 ± 00.69**	79.11 ± 06.15	84.74 ± 06.22	91.26 ± 01.23	90.28 ± 08.19
3	87.17	73.71 ± 01.61	56.15	50.62 ± 00.13	50.53 ± 02.84	**96.88 ± 00.91**	39.22 ± 19.36	92.25 ± 01.65	46.52 ± 05.46	90.55 ± 05.13
4	80.24	99.92 ± 00.02	84.78	99.17 ± 00.14	96.12 ± 00.24	99.62 ± 00.10	83.31 ± 04.10	**99.63 ± 00.21**	85.48 ± 11.89	98.97 ± 00.73
5	95.15	99.96 ± 00.01	98.13	99.18 ± 00.02	98.76 ± 00.16	98.99 ± 00.12	97.96 ± 00.64	98.23 ± 00.10	97.60 ± 01.45	**99.86 ± 00.05**
6	69.59	66.73 ± 01.52	55.05	59.62 ± 01.47	66.19 ± 00.90	**85.47 ± 00.62**	70.87 ± 15.09	84.04 ± 06.23	61.42 ± 01.79	83.89 ± 06.89
OA	86.42	94.73 ± 00.14	88.55	93.02 ± 00.06	92.31 ± 00.05	96.14 ± 00.02	86.83 ± 01.91	94.62 ± 00.21	88.42 ± 03.45	**96.73 ± 00.58**
AA	84.97	87.73 ± 00.43	80.91	82.10 ± 00.08	82.09 ± 00.46	**94.10 ± 00.14**	76.42 ± 04.46	91.33 ± 00.22	79.38 ± 00.96	93.68 ± 01.28
$\kappa(\times 100)$	82.21	92.92 ± 00.18	84.83	90.65 ± 00.09	89.69 ± 00.07	94.83 ± 00.02	82.49 ± 02.53	92.81 ± 00.28	84.68 ± 04.41	**95.62 ± 00.77**

is superior to that of all compared methods due to its capacity to learn spatial and spectral information. The morphFormer shows mean OA, AA, and k of 87.85%, 89.66%, and 86.81% having a standard deviation of 0.20%, 0.39%, and 0.22%, respectively.

Table II shows the generalization ability of the MUUFL dataset for disjoint train and test samples. Both RNN and RF exhibit comparable accuracies and outperform the remaining conventional classifiers. The morphFormer shows better accuracy than that of all other techniques, including transformer-based approaches, with OA, AA, and k of 93.84 ± 0.10%, 80.55 ± 0.27%, and 91.84 ± 0.13%, respectively.

Table III lists the classification results on the Trento dataset. RF outperforms other conventional classifiers, and

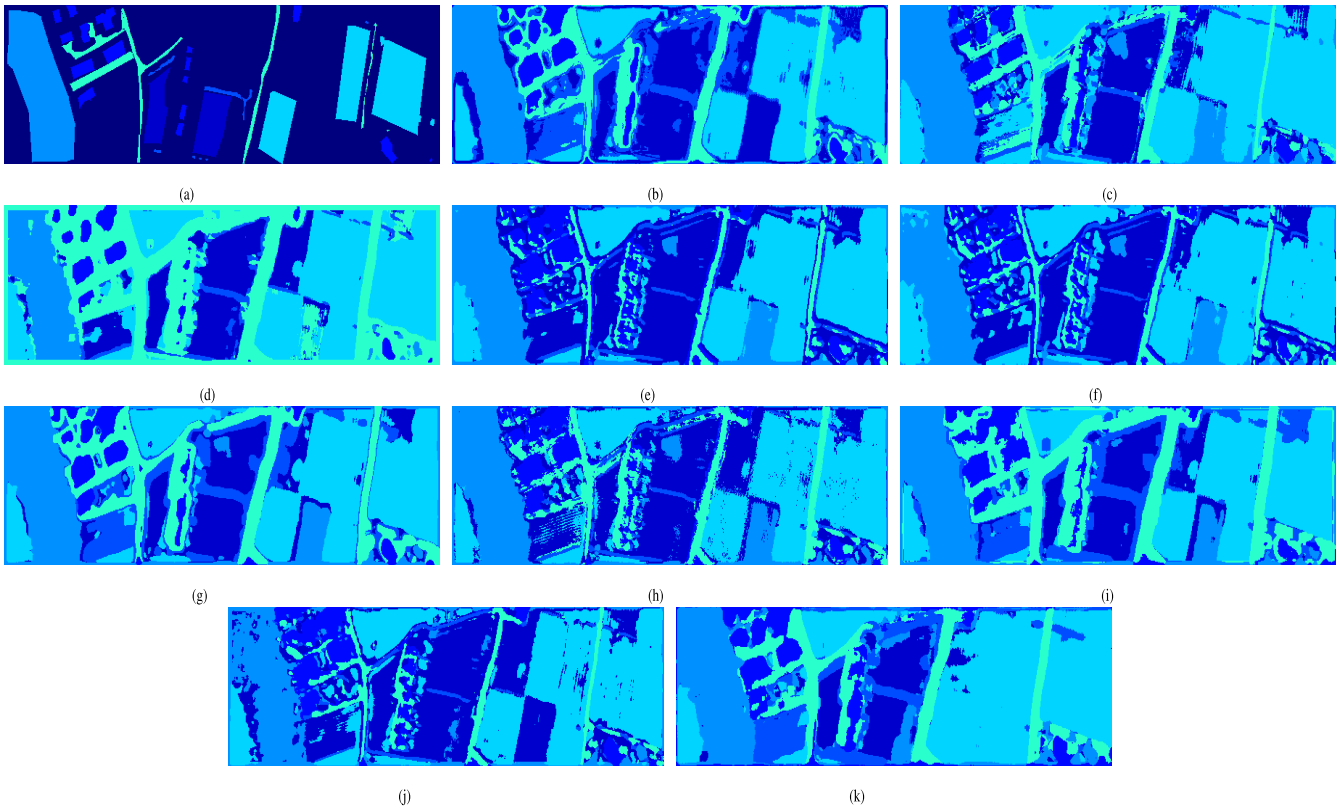


Fig. 10. Classification maps for the Trento HSI dataset. (a) Ground truth. (b) KNN (86.42%). (c) RF (94.73%). (d) SVM (88.55%). (e) CNN1D (93.02%). (f) CNN2D (92.31%). (g) CNN3D (96.14%). (h) RNN (86.83%). (i) ViT (94.62%). (j) SpectralFormer (88.42%). (k) morphFormer (96.73%).

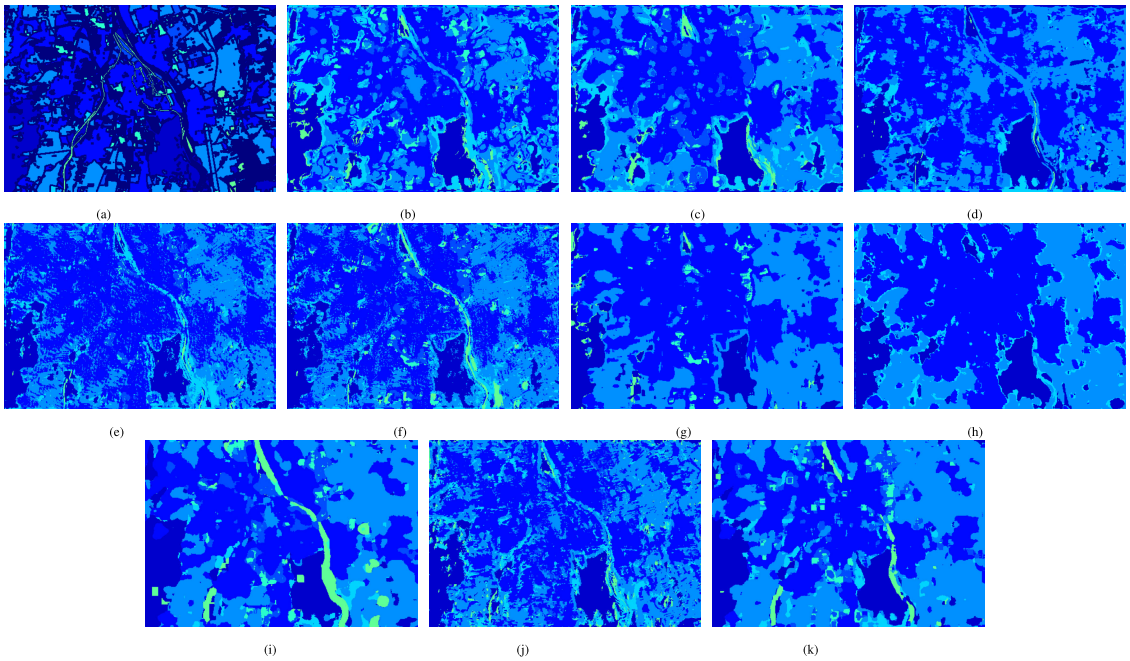


Fig. 11. Classification maps for the Augsburg HSI dataset. (a) Ground truth. (b) KNN (67.27%). (c) RF (79.96%). (d) SVM 71.60 (%). (e) CNN1D (72.00%). (f) CNN2D (73.59%). (g) CNN3D (82.89%). (h) RNN (40.26%). (i) ViT (85.90%). (j) SpectralFormer (70.81%). (k) morphFormer (88.68%).

CNN3D shows better accuracy than other DL-based methods. The morphFormer shows better classification accuracy than all other methods with OA, AA, and k of 96.73 $\pm$ 0.58%, 93.68 $\pm$ 1.28%, and 95.62 $\pm$ 0.77%, respectively.

Table IV shows the classification results on the Augsburg dataset. RNN exhibits lower accuracies than other conventional classifiers, while RF is the best conventional classifier, and CNN3D outperforms other DL-based approaches. The transformer ViT method outperforms our approach in terms

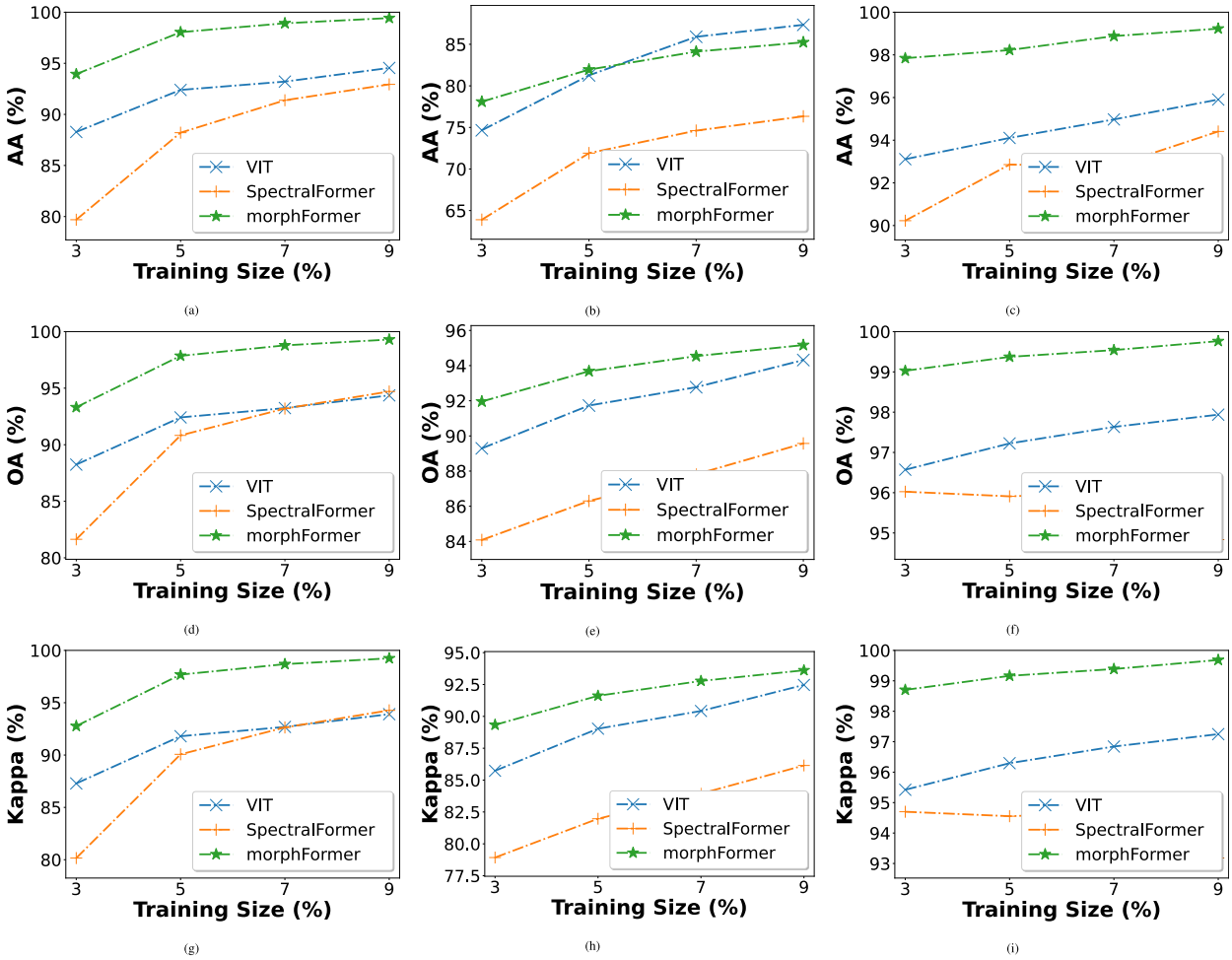


Fig. 12. Classification accuracies in terms of AA, OA, and kappa (κ) obtained by various techniques with different percentages of training samples randomly selected from (a), (d), (g), UH (b), (e), (h) MUUFL, and (c), (f), (i) Trento datasets.

of AA. However, the morphFormer outperforms all other methods in terms of OA and k.

D. Visual Comparison

Figs. 8–11 show the obtained classification maps. Our goal is to perform a qualitative evaluation of the compared methods. Conventional classifiers, such as KNN, RF, and SVM, provide classification maps with salt and pepper noise around the boundary areas because they only exploit spectral information. In addition, the DL methods produce better classification noise in comparison to conventional classifiers. Specifically, the maps produced by CNN1D, CNN2D, and CNN3D are smoother because the boundaries between land-use and land-cover classes can be separated in a better way. ViT can extract more abstract information in sequential representation, so it provides better classification maps. Compared to ViT and SpectraFormer, the proposed morphFormer exhibits better classification maps. In other words, our newly proposed morphFormer can enhance classification performance by considering spatial-contextual information and positional information across different layers. As a result, it characterizes texture and edge details better than other transformer-based techniques.

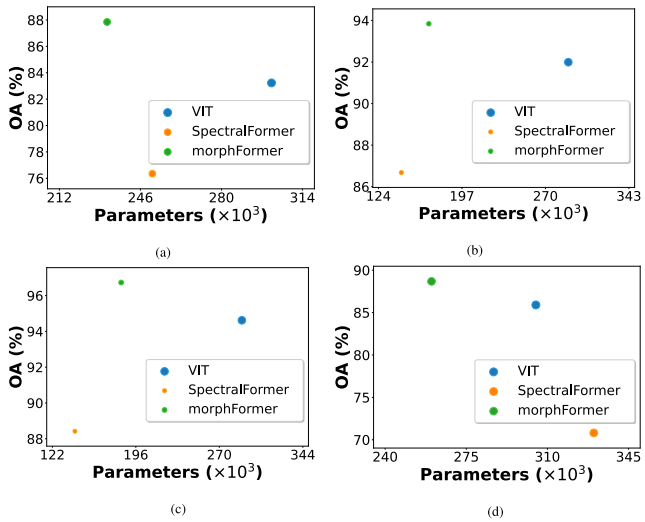


Fig. 13. Comparing the performance of transformer methods in terms of OA, network parameters, and FLOPs, shown by the radii of circles considered from (a) UH, (b) MUUFL, (c) Trento, and (d) Augsburg.

E. Performance Over Different Train Sample Sizes

Fig. 12(a)–(i) shows the classification performance of transformer models with different percentages of training samples

TABLE IV
CLASSIFICATION PERFORMANCE (IN %) ON THE AUGSBURG HSI DATASET

Class No.	Conventional Classifiers			Classical Convolutional Networks				Transformer Networks		
	KNN	RF	SVM	CNN1D	CNN2D	CNN3D	RNN	ViT	SpectralFormer	morphFormer
1	83.27	90.18 ± 00.34	91.51	77.79 ± 04.77	82.77 ± 02.64	78.66 ± 01.26	21.21 ± 29.99	88.62 ± 03.36	83.23 ± 04.20	**93.30 ± 01.41**
2	86.38	97.41 ± 00.14	86.18	87.40 ± 01.55	86.19 ± 01.41	96.10 ± 00.98	**98.99 ± 01.39**	94.90 ± 00.84	86.63 ± 02.81	96.08 ± 01.02
3	36.79	04.59 ± 00.33	10.34	30.30 ± 08.46	56.08 ± 00.99	49.25 ± 23.59	00.00 ± 00.00	**68.15 ± 03.93**	27.75 ± 05.25	60.63 ± 10.84
4	49.72	76.16 ± 00.99	62.75	65.25 ± 01.80	64.95 ± 02.05	84.96 ± 01.88	05.35 ± 07.56	84.40 ± 02.26	60.96 ± 04.83	**91.95 ± 01.53**
5	38.62	18.16 ± 00.62	23.14	15.74 ± 04.72	10.26 ± 02.98	10.13 ± 09.16	00.00 ± 00.00	34.29 ± 02.60	37.54 ± 07.18	**42.51 ± 14.44**
6	08.30	00.12 ± 00.09	00.00	**20.35 ± 13.47**	14.12 ± 08.17	08.45 ± 03.66	00.00 ± 00.00	17.83 ± 05.12	03.24 ± 00.90	10.36 ± 07.33
7	05.37	08.03 ± 00.99	10.68	14.36 ± 02.30	24.62 ± 02.92	11.70 ± 04.50	00.00 ± 00.00	**45.90 ± 00.92**	13.49 ± 08.87	14.64 ± 02.93
OA	67.27	79.96 ± 00.23	71.60	72.00 ± 01.70	73.59 ± 00.59	82.89 ± 00.78	40.26 ± 02.03	85.90 ± 00.26	70.81 ± 01.47	**88.68 ± 01.01**
AA	44.07	42.09 ± 00.21	40.66	44.45 ± 01.41	48.43 ± 01.99	48.46 ± 04.17	14.94 ± 00.86	**62.01 ± 01.18**	44.69 ± 02.31	58.50 ± 01.57
$\kappa(\times 100)$	53.86	70.04 ± 00.33	58.57	59.96 ± 02.40	62.16 ± 01.10	75.08 ± 01.35	02.65 ± 03.60	79.88 ± 00.27	58.05 ± 02.01	**83.62 ± 01.40**

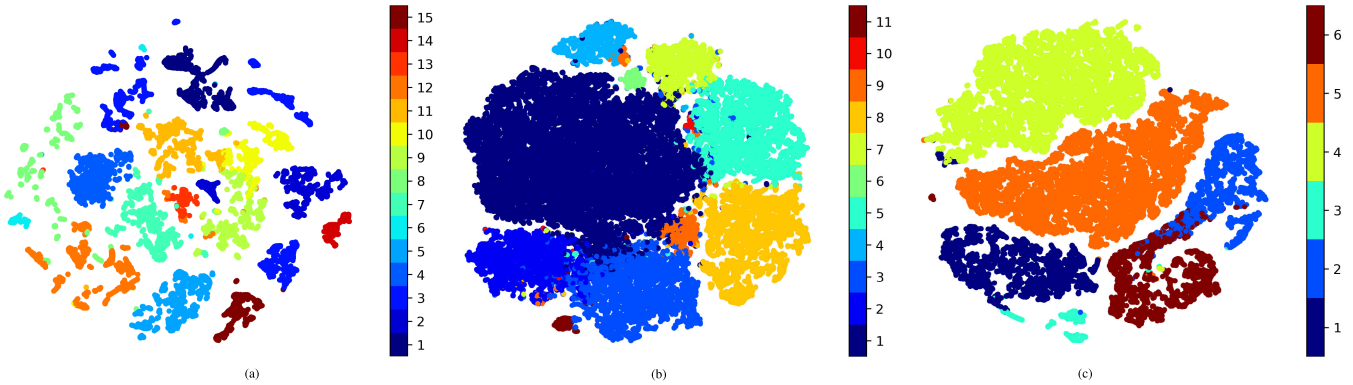


Fig. 14. 2-D graphical visualization of the features extracted by the proposed morphFormer through t-SNE. (a) Houston. (b) MUUFL. (c) Trento.

on three HSI datasets of Houston, MUUFL, and Trento. The training samples on these three datasets are randomly selected as 3%, 5%, 7%, and 9%.

In the Houston dataset, the proposed morphFormer outperforms the second-best-performing transformer model (ViT) by a margin of approximately 4% in terms of OA, AA, and k for all considered percentages of randomly selected samples. Although the margin is smaller for larger training sizes, the proposed morphFormer exhibits superior classification performance for all sample sizes in the other two datasets (MUUFL and Trento). It can be concluded that the proposed morphFormer exhibits significantly better classification performance than the other transformer networks, even with a limited number of training samples.

F. Hyperparameter Sensitivity Analysis

In terms of computing complexity, the proposed model is not only effective but also rather efficient. In Fig. 13(a)–(d), the parameters and calculations of the proposed method are compared to those of various transformer networks. Specifically, we show the OA, the number of parameters, and the number of calculations (FLOPs) for the UH, Trento, MUUFL, and Augsburg datasets. The calculations are shown by the radii of circles. The efficiency of morphFormer is clear in Houston and Augsburg datasets, where it needs the fewest parameters and FLOPS. Although the parameters and FLOPS needed by morphFormer are higher than those required by SpectralFormer in certain cases, the gain in performance compensates for that. As can be seen with the UH data, morphFormer offers an outstanding gain in OA (4.62%) over the next best model (ViT). In this case, the parameter tradeoff is justified by the significant increase in classification accuracy.

Furthermore, 2-D graphical plots depicting the features extracted by the proposed morphFormer are presented in Fig. 14(a)–(c) for Houston, MUUFL, and Trento datasets, respectively. Using the t-SNE approach [72], the features extracted by morphFormer can be analyzed. It can be observed that samples of similar categories gather together, and intraclass variance is minimized in all three datasets.

IV. CONCLUSION

We present a novel morphFormer network for HSI data classification, which is based on spectral and spatial morphological convolutions. Although fusing attention and morphological characteristics are not straightforward, our approach can successfully merge attention mechanisms with morphological operations and provide superior classification performance compared to standard convolutional models and the recently developed transformer models. Our morphFormer has the potential to excel in many different classification tasks in EO and RS. It is because of its ability to apply learnable morphological operations in addition to multihead self-attention mechanisms. A general adversarial network (GAN)-based method will be investigated with the morphFormer in our future work. Moreover, the LiDAR processing problem will also be solved using a morphFormer-based approach.

REFERENCES

[1] S. K. Roy, P. Kar, D. Hong, X. Wu, A. Plaza, and J. Chanussot, "Revisiting deep hyperspectral feature extraction networks via gradient centralized convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2021.

[2] M. Ahmad et al., "Hyperspectral image classification-traditional to deep models: A survey for future prospects," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 968–999, 2022.

[3] 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.

[4] B. Lu, Y. He, and P. D. Dao, "Comparing the performance of multispectral and hyperspectral images for estimating vegetation properties," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1784–1797, Jun. 2019.

[5] C. Chen, J. Yan, L. Wang, D. Liang, and W. Zhang, "Classification of urban functional areas from remote sensing images and time-series user behavior data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1207–1221, 2020.

[6] J. Yuan, S. Wang, C. Wu, and Y. Xu, "Fine-grained classification of urban functional zones and landscape pattern analysis using hyperspectral satellite imagery: A case study of Wuhan," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3972–3991, 2022.

[7] C. Shah and Q. Du, "Spatial-aware collaboration–competition preserving graph embedding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, May 2022, Art. no. 5506005.

[8] E. Bartholomé and A. S. Belward, "GLC2000: A new approach to global land cover mapping from Earth observation data," *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1959–1977, Feb. 2005.

[9] J. Senthilnath, S. N. Omkar, V. Mani, N. Karnwal, and S. P. B., "Crop stage classification of hyperspectral data using unsupervised techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 861–866, Apr. 2013.

[10] B. Koetz, F. Morsdorf, S. van der Linden, T. Curt, and B. Allgöwer, "Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data," *Forest Ecology Manage.*, vol. 256, no. 3, pp. 263–271, Jul. 2008.

[11] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, "Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, Feb. 2020.

[12] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, Jul. 2019.

[13] S. L. Ustin, *Manual of Remote Sensing, Remote Sensing for Natural Resource Management and Environmental Monitoring*, vol. 4. Hoboken, NJ, USA: Wiley, 2004.

[14] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[15] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.

[16] P. Ghamisi, J. A. Benediktsson, and S. Phinn, "Land-cover classification using both hyperspectral and LiDAR data," *Int. J. Image Data Fusion*, vol. 6, no. 3, pp. 189–215, 2015.

[17] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive Spectral–Spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.

[18] M. E. Paoletti, S. Moreno-Álvarez, and J. M. Haut, "Multiple attention-guided capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, 2022.

[19] M. Paoletti, X. Tao, J. Haut, S. Moreno-Álvarez, and A. Plaza, "Deep mixed precision for hyperspectral image classification," *J. Supercomput.*, vol. 77, pp. 9190–9201, Feb. 2021.

[20] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Jun. 2020.

[21] C. Shah and Q. Du, "Collaborative and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5248–5259, 2021.

[22] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral–multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, Jun. 2021.

[23] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[24] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

[25] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sens.*, vol. 10, no. 3, p. 482, Mar. 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/3/482

[26] S. K. Roy, P. Kar, M. E. Paoletti, J. M. Haut, R. Pastor-Vargas, and A. Robles-Gomez, "SiCoDeF^{2}net: Siamese convolution deconvolution feature fusion network for one-shot classification," *IEEE Access*, vol. 9, pp. 118419–118434, 2021.

[27] X. Wang, Y. Feng, R. Song, Z. Mu, and C. Song, "Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data," *Inf. Fusion*, vol. 82, pp. 1–18, Jun. 2022.

[28] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[29] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2020.

[30] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.

[31] 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.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[33] 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, Aug. 2018.

[34] S. K. Roy, S. R. Dubey, S. Chatterjee, and B. B. Chaudhuri, "FuSENet: Fused squeeze- and-excitation network for spectral-spatial hyperspectral image classification," *IET Image Process.*, vol. 14, no. 8, pp. 1653–1661, 2020.

[35] S. K. Roy, S. Chatterjee, S. Bhattacharyya, B. B. Chaudhuri, and J. Platoš, "Lightweight spectral–spatial squeeze-and-excitation residual bag-of-features learning for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5277–5290, Aug. 2020.

[36] 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 2020.

[37] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2018.

[38] M. E. Paoletti, J. M. Haut, S. K. Roy, and E. M. T. Hendrix, "Rotation equivariant convolutional neural networks for hyperspectral image classification," *IEEE Access*, vol. 8, pp. 179575–179591, 2020.

[39] S. K. Roy, M. E. Paoletti, J. M. Haut, E. M. T. Hendrix, and A. Plaza, "A new max-min convolutional network for hyperspectral image classification," in *Proc. 11th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, 2021, pp. 1–5.

[40] S. K. Roy, D. Hong, P. Kar, X. Wu, X. Liu, and D. Zhao, "Lightweight heterogeneous kernel convolution for hyperspectral image classification with noisy labels," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 2022, Art. no. 5509705.

[41] 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.

[42] S. K. Roy, J. M. Haut, M. E. Paoletti, S. R. Dubey, and A. Plaza, "Generative adversarial minority oversampling for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.

[43] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[44] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, pp. 1–41, Jan. 2022, doi: 10.1145%2F3505244.

[45] D. Hong et al., "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.

[46] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 165–178, Sep. 2020.

[47] Z. Zhong, Y. Li, L. Ma, J. Li, and W.-S. Zheng, "Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.

[48] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5522214.

[49] X. Yang, W. Cao, Y. Lu, and Y. Zhou, "Hyperspectral image transformer classification networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5528715.

[50] S. K. Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, "Multimodal fusion transformer for remote sensing image classification," 2022, *arXiv:2203.16952*.

[51] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Graph-based feature fusion of hyperspectral and lidar remote sensing data using morphological features," in *Proc. IGARSS*, 2013, pp. 4942–4945.

[52] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[53] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[54] A. Merentitis, C. Debes, R. Heremans, and N. Frangiadakis, "Automatic fusion and classification of hyperspectral and LiDAR data using random forests," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 1245–1248.

[55] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.

[56] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[57] S. K. Roy, B. Chanda, B. B. Chaudhuri, D. K. Ghosh, and S. R. Dubey, "Local morphological pattern: A scale space shape descriptor for texture classification," *Digit. Signal Process.*, vol. 82, pp. 152–165, Nov. 2018.

[58] 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.

[59] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015, pp. 448–456.

[60] X. Du and A. Zare, "Scene label ground truth map for MUUFL gulfport data set," Dept. Elect. Comput. Eng., Univ. Florida, Gainesville, FL, USA, Tech. Rep, 2017.

[61] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL gulfport hyperspectral and LiDAR airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013–570, 2013.

[62] D. Hong, J. Hu, J. Yao, J. Chanussot, and X. X. Zhu, "Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 68–80, Aug. 2021.

[63] A. Baumgartner, P. Gege, C. Köhler, K. Lenhard, and T. Schwarzmaier, "Characterisation methods for the hyperspectral sensor HySpex at DLR's calibration home base," *Proc. SPIE*, vol. 8533, Nov. 2012, Art. no. 85331H.

[64] F. Kurz, D. Rosenbaum, J. Leitloff, O. Meynberg, and P. Reinartz, "Real time camera system for disaster and traffic monitoring," in *Proc. Int. Conf. SMPR*, 2011, pp. 1–6.

[65] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Apr. 2020.

[66] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[67] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*.

[68] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[69] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "DiffGrad: An optimization method for convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4500–4511, Nov. 2019.

[70] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

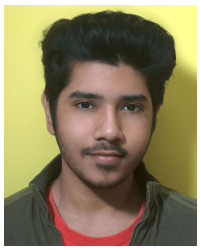
[71] E. M. T. Hendrix, M. Paoletti, and J. M. Haut, *On Training Set Selection in Spatial Deep Learning*. Cham, Switzerland: Springer, 2022, pp. 327–339, doi: 10.1007/978-3-031-00832-0_9.

[72] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014. [Online]. Available: http://jmlr.org/papers/v15/vandermaaten14a.html

Swalpa Kumar Roy (Student Member, IEEE) received the bachelor's degree in computer science and engineering from the West Bengal University of Technology, Kolkata, India, in 2012, the master's degree in computer science and engineering from the Indian Institute of Engineering Science and Technology, Shibpur (IIEST Shibpur), Howrah, India, in 2015, and the Ph.D. degree in computer science and engineering from the University of Calcutta, Kolkata, in 2021.

From July 2015 to March 2016, he was a Project Linked Person with the Optical Character Recognition (OCR) Laboratory, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, India. His research interests include computer vision, deep learning, and remote sensing.

Dr. Roy was nominated for the Indian National Academy of Engineering (INAE) Engineering Teachers Mentoring Fellowship Program by INAE Fellows in 2021. He was a recipient of the Outstanding Paper Award in the second Hyperspectral Sensing Meets Machine Learning and Pattern Analysis (Hyper-MLPA) at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He serves as an Associate Editor for the journal *Computer Science* (Springer Nature) (SNCS) and an Editor for the *Frontiers Journal of Advanced Machine Learning Techniques for Remote Sensing Intelligent Interpretation*. He has served as a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



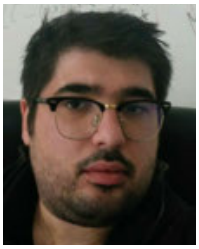
Ankur Deria received the bachelor's degree in computer science and engineering from the Jalpaiguri Government Engineering College, Jalpaiguri, India, in 2022. He is currently pursuing the M.Sc. degree with the Department of Informatics, Technical University of Munich, Garching bei München, Germany.

His research interests include computer vision and deep learning.

Mr. Deria was nominated for the Indian National Academy of Engineering (INAE) Engineering Students Mentoring Fellowship by INAE fellows in academic tenure 2022–2023.

Chiranjibi Shah (Member, IEEE) received the B.E. degree in electronics and communication from Pokhara University, Pokhara, Nepal, in 2012, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in May 2022.

His research interests include applying different machine learning and deep learning techniques for the classification of hyperspectral imagery, image recognition, dimensionality reduction, and object detection.



Juan M. Haut (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering and the Ph.D. degree in information technology, supported by an University Teacher Training Programme from the Spanish Ministry of Education, from the University of Extremadura, Cáceres, Spain, in 2011, 2014, and 2019, respectively.

He is currently a Professor with the Department of Computers and Communications, University of Extremadura. He is also a member of the Hyperspectral Computing Laboratory (HyperComp), Department of Technology of Computers and Communications, University of Extremadura. Some of his contributions have been recognized as hot-topic publications for their impact on the scientific community. His research interests include remote sensing data processing and high-dimensional data analysis, applying machine (deep) learning and cloud computing approaches. In this sense, he has authored/coauthored more than 50 *Journal Citation Reports* (JCR) journal articles (more than 30 in IEEE journals) and more than 30 peer-reviewed conference proceeding papers.

Dr. Haut was a recipient of the Outstanding Ph.D. Award at the University of Extremadura in 2019. He was a recipient of the Outstanding Paper Award in the 2019 and 2021 IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) conferences. He has been awarded the Best Reviewer Recognition of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2018 and 2020, respectively. From his experience as a Reviewer, it is worth mentioning his active collaboration in more than ten scientific journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. Furthermore, he has guest-edited three special issues on hyperspectral remote sensing for different journals. He is also an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a fellow of the SPIE-International Society for Optics and Photonics. She is a member of the IEEE TAB Periodicals Review and Advisory Committee (PRAC) and the SPIE Publications Committee. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013 and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. From 2016 to 2020, she was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 800 publications, including 393 JCR journal papers, 25 book chapters, and 330 peer-reviewed conference proceeding papers. He has guest edited 17 special issues on hyperspectral remote sensing for different journals. His main research interests comprise hyperspectral data processing and parallel computing of remote sensing data.

Prof. Plaza was a member of the Editorial Board of the IEEE Geoscience and Remote Sensing Newsletter from 2011 to 2012 and the *IEEE Geoscience and Remote Sensing Magazine* in 2013. He was also a member of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is a Fellow of IEEE "for contributions to hyperspectral data processing and parallel computing of Earth observation data." He was a recipient of the Best Column Award of the *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award of the JSTARS journal, and the Most Highly Cited Paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He received the Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He was a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), for which he served as Associate Editor from 2007 to 2012. He was recognized as a Highly Cited Researcher by Clarivate Analytics from 2018 to 2022. He is also an Associate Editor of IEEE ACCESS (receiving recognition as an Outstanding Associate Editor of the journal in 2017). He has served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President of the Spanish Chapter of IEEE GRSS from 2012 to 2016. He has reviewed more than 500 manuscripts for over 50 different journals. He has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. Additional information is available at http://sites.google.com/view/antonioplaza.