



Multiscale Discriminative Feature Adaptive Learning Network for Hyperspectral Image Classification

Linfeng Liu , Shuanggen Jin , Senior Member, IEEE, Abhishek Banerjee, and Chengcai Zhang

Abstract—Deep learning has advanced hyperspectral image (HSI) classification by efficiently extracting spectral and spatial features. However, its performance is often limited when labeled data are scarce and class distributions are imbalanced. Moreover, current methods frequently fail to effectively capture multiscale discriminative representations that bridge local spatial details with global contextual information. To address these challenges, in this work, we present a multiscale discriminative feature adaptive learning (MDFAL) network for HSI classification. MDFAL employs a cascaded architecture that integrates shallow, intermediate, and deep sub-networks, jointly optimized through multiscale feature outputs and adaptive weighting of multitask losses. To capture discriminative representations more effectively, the framework incorporates three complementary modules: 1) a spectral–spatial attention module that strengthens cross-dimensional interactions while suppresses redundancy; 2) a convLSTM-based contextual learning module that models both short- and long-range spectral dependencies together with spatial context; and 3) a dilated convolution block that expands the receptive field, thereby enhancing multiscale representation across network layers. Experimental results demonstrate that the proposed MDFAL model surpasses existing methods by effectively capturing multiscale discriminative features with yielding strong intraclass compactness and interclass separability. Under identical conditions, MDFAL improves overall accuracy by 2.31%, 3.12%, and 3.10% on the Indian Pines, Pavia University, and Tea datasets, respectively. These improvements arise from the synergistic interaction between its cascaded shallow, intermediate, deep architecture, and adaptive multiscale loss weighting. The findings confirm MDFAL as a robust framework for advancing HSI classification network design.

Index Terms—Adaptive learning, attention mechanism, convLSTM, discriminative feature, hyperspectral image (HSI) classification.

I. INTRODUCTION

LAND-USE classification using remote sensing imagery has become a key priority in Earth observation, driven by advances in aerospace imaging, the integration of multisource datasets, and continuous improvements in imaging methodologies [1], [2], [3]. Hyperspectral image (HSI) captures surface reflectance across hundreds of narrow bands spanning the visible, near-infrared, shortwave infrared, and thermal infrared regions. Each spectral range provides unique diagnostic information but also faces inherent limitations. Visible and near-infrared bands are effective for vegetation and land-cover classification, though they are hindered by atmospheric scattering and water vapor absorption. Shortwave infrared bands are highly sensitive to mineral and soil properties yet require stringent atmospheric correction. Thermal infrared bands uniquely enable energy and temperature-related analyses, but their spectral resolution is comparatively lower. Harnessing the complementary strengths of these spectral domains is central to achieving fine-grained material discrimination in Earth observation. Each image captures spatial variation in two dimensions and spectral variation in the third, encoding distinctive radiometric signatures for various land-cover types. As a result, HSI has seen rapidly adopted diverse applications, for instance, urban land-use mapping [4], [5], mineral and marine exploration [1], crop yield and pest analysis [6], disaster assessment [7], and environmental protection [8]. This widespread adoption is largely due to the unique strengths of HSI data, including high information content, fine spatial and spectral resolution, and strong interoperability with cartographic frameworks.

To assess HSI analysis, researchers have developed a wide range of classical classification algorithms tailored to the unique properties of HSI data [9], [10], [11]. These efforts highlight the central role that HSI classification across numerous applications. Previous studies have engrossedly focused on machine learning methods such as support vector machines (SVM) [12], k-nearest neighbor (KNN), naive Bayes, radial basis function networks, decision trees [13], and extreme learning machine [14]. These approaches primarily relied on spectral features and often failed to capture the complex spatial structures present in HSI data, resulting in spatial inconsistencies often referred to as the “pretzel” effect. Among these methods, SVM introduced kernel-based transformations to create discriminative

Received 29 July 2025; revised 1 October 2025 and 16 October 2025; accepted 22 October 2025. Date of publication 27 October 2025; date of current version 17 November 2025. This work was supported in part by the “Double First-Class” Discipline Construction Project of the School of Surveying and Geoinformation Engineering, Henan Province under Grant BSJJ202303, in part by the Fundamental Research Funds for the Universities of Henan Province under Grant NSFRF2502031, in part by the Jiangsu Marine Science and Technology Innovation Project under Grant JSZRHYKJ202202, in part by the National Natural Science Foundation of China under Grant 42471130 and Grant 51739009, and in part by the PhD Fund of Henan Polytechnic University under Grant B2024-20. (Corresponding author: Shuanggen Jin.)

Linfeng Liu is with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China, and also with the School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China (e-mail: liulinfeng@hpu.edu.cn).

Shuanggen Jin is with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China, and also with the School of Artificial Intelligence, Anhui University, Hefei 230601, China (e-mail: sgjin@hpu.edu.cn).

Abhishek Banerjee is with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China (e-mail: abhishek@hpu.edu.cn).

Chengcai Zhang is with the School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China (e-mail: liulinfeng0408@126.com).

Digital Object Identifier 10.1109/JSTARS.2025.3625972

hyperplanes, while KNN applied a simple Euclidean metric to evaluate sample similarity. Recognizing the importance of spatial information, subsequent research introduced numerous techniques that combine spatial and spectral domains, in turn, improving classification performance [15]. These include joint sparse representation models [16], superpixel-based segmentation [17], and morphology-driven approaches [18]. Nonetheless, challenges remain due to their strong nonlinearities in HSI data, spectral redundancy [19], interband variation, and limited labeled samples [20], all of which constrain the effectiveness of traditional machine learning methods.

In response, a variety of deep learning approaches have emerged for hyperspectral feature extraction [21], including autoencoders, deep belief networks [22], and recurrent neural networks (RNNs) [23]. The rise of deep learning in computer vision has motivated its application in HSI classification. Deep networks are increasingly employed for their capability to automatically extract abstract and robust features from raw data, minimizing the need for manual feature engineering [24]. Among them, convolutional neural networks (CNNs) have achieved notable accomplishments by leveraging local connectivity, weight sharing, and parameter efficiency [25]. Both two-dimensional (2-D) CNNs [26] and three-dimensional (3-D) CNNs [27] have proven effective in extracting spatial and spectral features, respectively [28]. Deeper architectures such as VGG and GoogLeNet have also been adapted for HSI classification; however, their performance often suffers from gradient vanishing issues as depth increases. Residual networks (ResNets) [29] mitigate this limitation by introducing identity shortcut connections that stabilize training and enhance accuracy. Based on CNN and ResNet backbones, numerous HSI classification models (e.g., SSRN [27], DFRN [30], RSSAN [31]) have been developed for multiscale spatial–spectral feature extraction [32], achieving superior performance compared to traditional methods [33].

RNNs have been widely used in HSI classification due to their ability to extract sequential and semantic features from spectral data [34]. To overcome vanishing and exploding issues common in conventional RNNs, long short-term memory (LSTM) networks were introduced. LSTMs effectively capture long-range dependencies in continuous data, offering improved stability and performance over conventional RNNs [35]. However, a key limitation of LSTMs is the need to flatten input data into one-dimensional sequences, which disrupts the inherent spatial structure of HSI data. To overcome this drawback, Shi et al. [36] proposed the convolutional LSTM (ConvLSTM), which incorporates convolutional operations within the LSTM framework. Thereby, preserving spatial dimensions during the temporal modeling process improves the ability to capture spatiotemporal dependencies [37]. Building on this concept, several advanced architectures have been developed for HSI classification, including Spectrum–Space LSTM [38], Bi-ConvLSTM [39], and ConvLSTM 3-D [40]. Notably, models such as A3CLNN [41], RS-AMCNN [42], and ASSMN [43] have demonstrated state-of-the-art classification performance with higher accuracy.

Attention mechanisms, inspired by human visual perception, have also been integrated into deep learning models to focus

selectively on salient information while filtering out irrelevant information [44], [45]. These mechanisms have significantly enhanced model performance across various fields [46], [47], including classification [48], [49]. In remote sensing, attention modules enable adaptive weighting of spatial and channel-wise features, leading to improved feature representation and classification accuracy. In the context of HSI classification, numerous attention-based deep learning models have been developed [50], [51]. For example, Li et al. [41] introduced A3CLNN, a dual-channel ConvLSTM classifier that incorporates attention for enhanced spatial and spectral modeling. Similarly, Jie et al. [42] proposed RS-AMCNN, an attention-guided multibranch framework incorporating adaptive region selection to boost classification performance in complex HSI environments.

With continued innovation, advanced architectures such as transformers [10], diffusion models [52], and Mamba networks [53], [54] have been increasingly applied to HSI classification. Transformer-based models emphasize the extraction of global contextual information, diffusion models progressively capture the spatial–spectral distribution of high-dimensional data through iterative diffusion processes, while Mamba networks leverage global receptive fields and dynamic weighting mechanisms to overcome modeling constraints, showing strong performance in long-sequence tasks. Given the scarcity of labeled HSI samples, various data-efficient strategies have been introduced, including semi-supervised learning [55], active learning [56], few-shot learning [57], and generative adversarial networks [20]. More recently, multiscale feature learning that integrates local representations from CNNs with global context captured by transformers has also gained growing attention [10], [52].

To extract multiscale discriminative features for HSI classification, we propose an adaptive learning framework that integrates a deep ConvLSTM module with a spectral–spatial attention mechanism. The network consists of three parallel submodules, each including a spectral–spatial attention module (SSAM), a convLSTM-based contextual learning module (CCLM), a residual connection, and a dilated convolutional unit. The CCLM is designed to capture both short- and long-term spectral dependencies while preserving spatial context embedded in HSI data. Simultaneously, the SSAM dynamically learns spectral and spatial features by attending to location-specific patterns within the HSI cube. To capture spatial features at multiple scales, we employ dilated convolutions, while residual connections facilitate gradient flow and efficient parameter propagation. Multiscale discriminative features extracted from the shallow, intermediate, and deep layers are further refined through a coregulation network with multiple loss branches. To enhance layer-wise feature optimization, we incorporate an adaptive weight learning strategy, enabling more accurate and robust classification. The aims of this study are summarized as follows.

- 1) To capture multiscale and highly discriminative spectral–spatial features, this study proposes an adaptive deep learning framework for HSI classification. The framework, termed multiscale discriminative feature adaptive learning (MDFAL), employs a hierarchical architecture with shallow, intermediate, and deep layers to jointly

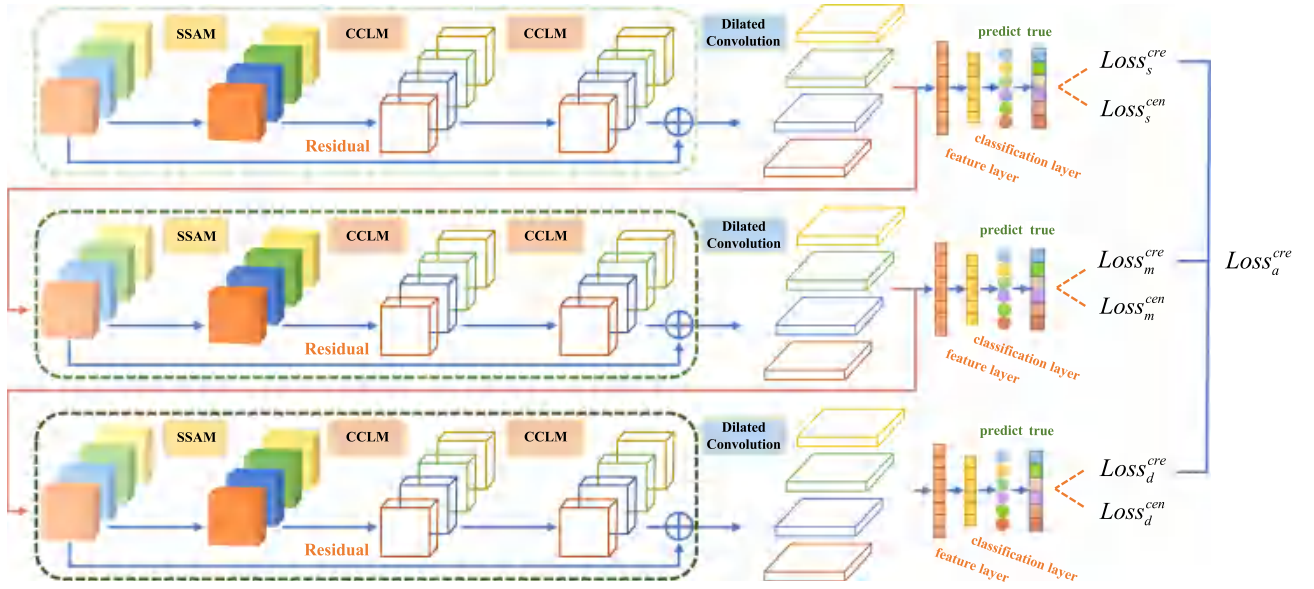


Fig. 1. Overall architecture diagram.

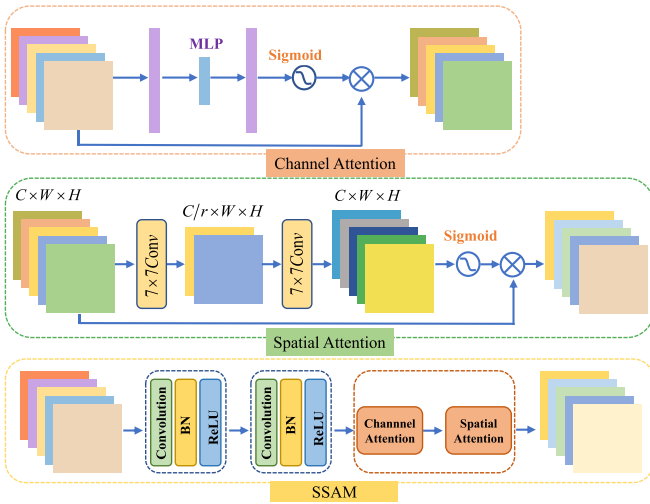


Fig. 2. Configuration specifics of the SSAM.

represent spectral–spatial information across scales. It further incorporates a multitask adaptive weighting strategy grounded in the principle of homoscedastic uncertainty to balance multiscale losses, while jointly optimizing cross-entropy and center losses for more robust training.

- 2) To effectively capture global similarity features and both short- and long-range dependencies among hyperspectral pixels, we developed an integrated multitype feature extraction module. The framework incorporates an SSAM to selectively emphasize class-relevant regions, a CCLM to model cross-band spectral dependencies and spectral–spatial context relationships, and a dilated convolution module to expand the receptive field. Collectively, these components enable shallow, intermediate, and deep layers to learn complementary features across multiple receptive field scales.

- 3) To enhance model convergence and mitigate overfitting, MDFAL integrates residual connections, batch normalization (BN), and dropout as regularization techniques. Extensive experiments conducted on three public benchmark hyperspectral datasets demonstrate that the proposed cascaded MDFAL network not only effectively extracts multiscale discriminative features with strong intraclass compactness and interclass separability, but also exhibits superior generalization ability and stability.

The rest of this article is structured as follows. Section II introduces the neural networks, with emphasis on the convolutional operations embedded in the proposed MDFAL model. Section III details the MDFAL architecture and parameter settings. Section IV reports results, evaluation, and comparisons with state-of-the-art classification methods. Finally, Section V concludes this article with the main findings and discussions on future potential directions.

II. MULTISCALE DISCRIMINATIVE FEATURE ADAPTIVE LEARNING

A. Architecture Diagram

Fig. 1 presents the architecture proposed for HSI classification, which integrates attention mechanisms, residual connections, and ConvLSTM into a hierarchical learning framework. This design enables continuous extraction of both spectral and spatial features across multiple levels of abstraction. The network is structured into three levels: shallow, intermediate, and deep. Each level is designed to capture increasingly complex semantic representations. Features from each level undergo independent representation learning and are then fused via a cross-scale integration module that aggregates complementary information across layers. Within each tier, spatial, and spectral feature learning modules produce discriminative outputs that strengthen the model's ability to identify class-specific patterns.

To optimize training, the model jointly employs cross-entropy and center loss functions [58], reducing prediction error while enhancing intraclass compactness and interclass separability. This optimization is embedded within a multitask learning framework that adaptively tunes the contribution of each loss component during training. Through this mechanism, the network effectively consolidates multiscale discriminative features, thereby advancing classification accuracy under complex HSI conditions.

Each processing block begins by dividing the input, which may be the original hyperspectral data cube or the output from the preceding layer, into multiple spectral segments across a temporal sequence referred to as T-period slices. A global attention mechanism is then applied to each spectral subcube to most irrelevant features for classification while reducing the influence of redundant or inappropriate components. This process effectively minimizes spectral redundancy and captures the spatial distribution of ground objects across temporally partitioned data. The refined features are then passed through two stacked ConvLSTM layers, which jointly model local spatial structures and long-range spectral dependencies across the T-period sequence. To further increase the receptive field while maintaining spatial resolution, 2-D dilated convolutions are used, enabling multiscale spatial feature extraction. Each stage includes BN and ReLU activation to stabilize training, facilitate gradient flow, and accelerate convergence. Residual connections are integrated throughout the network to support efficient feature propagation and mitigate the risk of vanishing or exploding gradients, which enhances the robustness and learning efficiency of the model.

B. Spectral-Spatial Attention Module

Recent advances in HSI classification have facilitated the development of a variety of attention mechanisms, including triplet attention modules and transformer-based architectures [38], [39]. These approaches enhance classification performance by selectively enriching informative features. Early approaches, such as SENet, focused primarily on single-dimensional attention, whereas later methods, including coordinate attention, incorporated both channel and spatial information. Hybrid architectures such as CBAM and BAM further extended this concept by integrating channel and spatial attention either sequentially or in parallel. Despite these improvements, most existing methods still struggle to capture the joint dependencies across the spectral, spatial, and channel dimensions. To overcome this limitation, we extended previous research [59], [60] and propose a novel attention mechanism that explicitly models multidimensional interactions. This design enhances classification accuracy by alleviating the suppression of salient features and improving the global representation of spectral and spatial relationships in HSI data.

Given an input feature map of size $E_i = R^{H \times W \times C}$, let E_c and E_s denote the outputs of the channel and spatial attention mechanisms, respectively, and E_o denote the final output obtained through residual attention

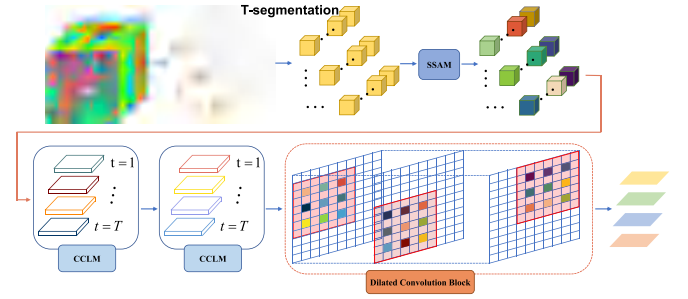


Fig. 3. Structural diagram of individual components.

$$E_c = A_c(E_i) \otimes E_i$$

$$E_s = A_s(E_c) \otimes E_c$$

$$E_o = E_s + E_i \quad (1)$$

where A_c and A_s correspond to the channel and spatial attention components, respectively, while \otimes means elemental-wise multiplication.

Fig. 2 illustrates the function of the channel-focus submodule, which uses a multilayer perceptron to compress the feature map along the spatial dimension. This process captures the global distribution responses across channels, enhancing the network's capacity to aggregate information over the entire receptive field and to model complex nonlinear interactions between spatial and channel dimensions. Similarly, the spatial attention submodule utilizes two successive 2-D convolutional layers to extract and integrate spatially salient features, while preserving information by avoiding pooling operations.

C. Configuration Specifics for a Single Module

To effectively capture multiscale discriminative features and complex local spatial contextual dependencies critical for HSI classification, the proposed network incorporates a feature learning module with residual connections. This module includes an SSAM and a CCLM to enhance spectral-spatial feature representation. In addition, the network supports adaptive weight optimization across multiple modules, tasks, and loss functions. Fig. 3 presents a detailed illustration of the internal structure of a representative component.

In this context, the original HSI data cube is denoted as $X = R^{H \times W \times D}$, with its height, width, and number of spectral bands represented as H, W , and D , respectively. To extract informative spatial context from neighboring pixels, a key pre-processing step was implemented. A neighborhood block of size $s \times s \times d$ was defined, centered on a target pixel x , and all surrounding pixels within this block, hereafter referred to as $X^G \in R^{s \times s \times D}$, were included to construct the local spatial context. The resulting three-dimensional representation, denoted as X^G , was subsequently decomposed into its constituent components and reshaped into a 1-D sequence of length T , as described in (4). This transformation enabled seamless integration into the ConvLSTM module, facilitating the modeling of long-range

dependencies in the spectral domain. In summary, this preprocessing pipeline systematically prepares the HSI data for the proposed MDFAL model by extracting localized spatial context and structuring the data to support temporal modeling across spectral bands.

The target pixel served as the central reference point. Subsequently, the preprocessed data were prepared for input into the proposed MDFAL model. Notably, all pixels in the surrounding region were encompassed within a neighborhood block, hereafter referred to as $X^G \in R^{s \times s \times D}$. The resulting 3-D representation, denoted as X^G , was then decomposed into its constituent components and reshaped into a 1-D sequence of length T , as described in (4). This transformation enabled subsequent integration into the ConvLSTM module, facilitating the modeling of long-term dependencies within the spectral domain. In summary, the processing of HSI data involves a systematic sequence of steps to extract local spatial context and to restructure the data for effective spectral-temporal modeling within the MDFAL framework

$$X^G = \{X_1^G, X_2^G, \dots, X_t^G, \dots, X_T^G\} \quad (2)$$

where X_t^G denotes the t th 3-D sequence of size $s \times s \times d$, $d = D/T$, and $t \in \{1, 2, \dots, T\}$. The variable T represents the number of divisions applied to the spectral axis within the ConvLSTM layer. To effectively capture spectral-geographical dependencies within each time series, denoted as t , the segmented sequence is passed through an attention mechanism. The SSAM comprises two successive 2-D convolutional layers, followed by a BN layer, a nonlinear activation function, a channel attention submodule, a spectral attention submodule, and a residual connection. The convolution is performed with a kernel size of 3×3 , a stride of 1, and zero-padding to maintain consistent spatial dimensions. BN stabilizes feature distributions, while the rectified linear unit (ReLU) activation promotes sparsity and reduces overfitting. The corresponding formulation is given in the following equation:

$$\begin{aligned} B_t^G &= f_{\text{relu}}(f_{\text{bn}}(f_{\text{cv}2D_3}(X_t^G))) \\ C_t^G &= f_{\text{bn}}(f_{\text{cv}2D_3}(B_t^G)). \end{aligned} \quad (3)$$

The spectral features are initially compressed using a 2-D convolution with a kernel size of 1×1 , a stride of 1, and zero-padding to preserve spatial dimensions. This operation yields a compression ratio, denoted as γ , resulting in a reduced spectral dimension of $s \times s \times (d/\gamma)$. During online learning, spatial structures are preserved while redundant spectral information is effectively suppressed. Following compression, the spectral features are upsampled using the same convolutional configuration to restore their original spectral resolution. This reconstruction enhances spectral representation learning, which is critical for accurate classification. Notably, the output dimensionality of the spectral attention submodule ($s \times s \times d$) remains identical to its input. This relationship is formally defined by the following equation:

$$\begin{aligned} Z_{c,t} &= f_{\text{cv}2D_1^d}(f_{\text{relu}}(f_{\text{cv}2D_1^{d/\gamma}}(C_t^G))) \\ E_{c,t} &= Z_{c,t} \odot X_t^G. \end{aligned} \quad (4)$$

The sequence characteristics generated by the spectral attention mechanism are denoted as $\{E_{c,1}, E_{c,2}, \dots, E_{c,t}, \dots, E_{c,T}\}$. Here, $E_{c,t}$ represents the output of the t th attention unit, $f_{\text{cv}2D_1^d}$ and $f_{\text{cv}2D_1^{d/\gamma}}$ denotes 2-D convolutional operations with d and d/γ filters, respectively, and 1×1 indicates the convolutional kernel size. The nonlinear activation function is represented by f_{relu} , while \otimes denotes the dot product operation. The variable $Z_{c,t}$ is element-wise multiplied with the input to yield X_t^G , which serves as the spectrally enhanced feature representation. All symbols maintain the same definitions as previously introduced.

The spectral attention output is subsequently input to the spatial attention submodule. Within this module, two convolutional kernels of size 7×7 and d/γ filters are employed, together with a convolutional layer utilizing zero-padding to preserve spatial dimensions. The resulting feature map is then passed through a nonlinear activation function followed by BN, mirroring the operations performed in the spectral attention submodule. This consistency in parameter sharing facilitates seamless feature transfer and accelerates model convergence. The corresponding operation is formally expressed as follows:

$$\begin{aligned} E_{s,t} &= \text{sigmoid} \\ &\quad \left(f_{\text{bn}} \left(f_{\text{cv}2D_7^d} \left(f_{\text{relu}} \left(f_{\text{bn}} \left(f_{\text{cv}2D_7^{d/\gamma}}(E_{c,t}) \right) \right) \right) \right) \right) \quad (5) \\ Z_{s,t} &= f_{\text{relu}} \left(f_{\text{bn}} \left(f_{\text{cv}2D_7^{d/\gamma}}(E_{c,t}) \right) \right) \\ Y_{s,t} &= \text{sigmoid}(f_{\text{bn}}(f_{\text{cv}2D_7^d}(Z_{s,t}))) \\ E_{s,t} &= Y_{s,t} \odot E_{c,t}. \end{aligned} \quad (6)$$

The feature sequence produced by the spatial attention module is denoted as $\hat{E}_S = \{E_{s,1}, E_{s,2}, \dots, E_{s,t}, \dots, E_{s,T}\}$. The intermediate feature vectors are represented by $Z_{s,t}$ and $Y_{s,t}$ let $E_{s,t}$ signify the output of the t th spatial attention unit. The operations $f_{\text{cv}2D_7^d}$ $f_{\text{cv}2D_7^{d/\gamma}}$ correspond to 2-D convolutional functions with d and d/γ filters, respectively, and 7×7 denote the convolutional kernel size. The symbol f_{bn} refers to BN. To generate a spatially enhanced feature representation, the spatial attention submodule output is element-wise multiplied with the nonlinear input, yielding RR as the normalized informative feature map.

Following processing by the attention module, a sequence of length T and feature dimension $s \times s \times d$ is systematically derived from each feature map cube. These sequences are seamlessly integrated to conform to the input dimensionality required by the ConvLSTM module, which subsequently processes them through two consecutive ConvLSTM blocks. Each block comprises a ConvLSTM unit with a hidden state size of 3×3 , followed by a BN layer and a nonlinear activation function. These ConvLSTM modules are specifically designed to model the global temporal dynamics of the spectral sequence, thereby capturing long-range spectral dependencies and producing a normalized attention map. The element-wise multiplication of the input feature map with this attention map yields a more robust and informative spectral representation. For clarity, the mathematical expressions of each processing module are given as follows:

$$U_{\text{cvlm}} = f_{\text{relu}} \left(f_{\text{bn}} \left(f_{\text{cvlm}3}(\hat{E}_S) \right) \right)$$

TABLE I
EXPERIMENT DATASET INFORMATION COMPARISON

Dataset	Sensor	size	bands	Spatial resolution	Spectral resolution	Class	wavelength
IN	AVIRIS	145*145	200	20 m	10 nm	16	0.4–2.5 μm
PU	ROSIS	610*340	103	1.3 m	4 nm	9	0.43–0.86 μm
Tea	PHI	512*348	80	2.25 m	8nm	10	0.42–0.86 μm

$$\hat{U}_{\text{cvlm}} = U_{\text{cvlm}} \odot \hat{E}_s. \quad (7)$$

For each ConvLSTM module, the output is denoted as \hat{U}_{cvlm} .

To this end, a dilated convolution strategy is employed, which, unlike conventional pooling techniques, preserves informative features while expanding the receptive field. This design enables the extraction of multiscale contextual representations without compromising spatial resolution. Each of the T feature sequences, derived from the final output of the ConvLSTM module, is processed by two parallel 2-D dilated convolutional units with a kernel size of 3×3 , a dilation rate of 2, and a stride of 1. Specifically, the filter number in the initial convolutional layer is set to d and remains constant until it increases to $2d$ in the subsequent dilated convolution layer, and so forth.

D. Adaptive Optimization With Multiple Losses

The motivation for adopting an adaptive multitask loss weighting strategy arises from the need to dynamically balance heterogeneous learning objectives. In multitask learning, manually assigning fixed loss weights often leads to suboptimal convergence because individual tasks differ in learning complexity, gradient magnitude, and convergence behavior. To address this imbalance, the proposed MDFAL method incorporates a self-adaptive weighting mechanism that automatically adjusts task contributions according to task-specific uncertainty and loss magnitude.

Theoretically, this approach is grounded in the principle of homoscedastic uncertainty weighting [61], which formulates each task loss as a probabilistic likelihood function. By assigning weights inversely proportional to task uncertainty, the model emphasizes tasks with more reliable gradients while suppressing those with higher variance, thereby ensuring a more stable and balanced optimization process.

Within the proposed MDFAL framework, this adaptive mechanism facilitates joint optimization across shallow, intermediate, and deep discriminative feature learning branches, preventing any single branch from dominating the training. As a result, the model achieves improved convergence stability and stronger complementarity among multiscale spectral-spatial representations, ultimately enhancing classification accuracy and robustness

$$\begin{aligned} \text{Loss}_{\text{total}} = & \frac{1}{\sigma_1^2} \text{Loss}_s^{\text{cre}} + \frac{1}{\sigma_2^2} \text{Loss}_s^{\text{cen}} + \frac{1}{\sigma_3^2} \text{Loss}_m^{\text{cre}} + \frac{1}{\sigma_4^2} \text{Loss}_m^{\text{cen}} \\ & + \frac{1}{\sigma_5^2} \text{Loss}_d^{\text{cre}} + \frac{1}{\sigma_6^2} \text{Loss}_d^{\text{cen}} + \frac{1}{\sigma_7^2} \text{Loss}_a^{\text{cre}} + \log \sigma_1 \\ & + \log \sigma_2 + \log \sigma_3 + \log \sigma_4 + \log \sigma_5 + \log \sigma_6 \\ & + \log \sigma_7 \end{aligned} \quad (8)$$

Algorithm 1: Training Steps for the Proposed MDFAL.

1. **Input:** HSI Data with ground truth is divided into training and test samples, labels, and time step, batch size, quantity of filters, and epochs E .
 2. Initialize the network weights and biases. The training samples are fed into the MDFAL network.
 3. for E epochs
 4. for each batch size training sample
 5. SGD updates network parameters, including the loss weight coefficient, to minimise total loss $\text{Loss}_{\text{total}}$. The test samples were transferred to the intermediate model, and the most accurate model was stored.
 6. end for
 7. end for
 8. According to the trained MDFAL model, the whole HSI image was predicted.
 9. END
-

where s , m , and d represent the shallow, medium, and deep losses, respectively. a symbolizes the average loss over all three layers; cre and cen denotes the cross-entropy loss and center loss, accordingly. $\text{Loss}_{\text{total}}$ is the total loss, and σ_i reflects the appropriate weighting factor $i \in (1, 2, \dots, 7)$.

III. RESULTS AND ANALYSIS

This section introduces three benchmark HSI datasets adopted in this study. It further investigates critical factors affecting the performance of the proposed MDFAL model and presents an ablation analysis to quantify the contribution of each component. A comprehensive comparative analysis of experimental results is then presented. Finally, the proposed framework is benchmarked against advanced HSI classification approaches employing modern machine learning techniques under consistent experimental settings.

A. Dataset and Experiment

This study selected three publicly available hyperspectral datasets to evaluate the classification accuracy, robustness, and generalization capability of the proposed MDFAL model. These datasets were intentionally chosen due to their variations in image dimensions, spatial and spectral resolutions, ground object categories, and sample sizes, facilitating a comprehensive assessment of the model's adaptability. Table I provides an overview of the datasets utilized in this study.

- 1) *Indian Pines (IN)*: Acquired in 1992 by the AVIRIS sensor over northwestern Indiana, it contains 224 spectral bands at 20 m spatial resolution and an image size of 145×145 pixels. To minimize atmospheric interference, 20 bands degraded by water vapor absorption and 4 bands containing only zero values were discarded, resulting in 200 valid bands retained for analysis. It includes 16 land-cover classes spanning various agricultural and natural vegetation types.
- 2) *Pavia University (PU)*: Collected in 2001 by the ROSIS sensor over Pavia, Italy, it comprises 115 spectral bands

with a spatial resolution of 1.3 m, totaling 610×340 pixels. After removing 12 bands affected by noise and water vapor absorption, 103 bands were retained for classification analysis. It comprises nine land-cover classes covering a variety of urban and natural surface types.

- 3) *Tea*: Hyperspectral imagery was collected via an aerial survey over a tea plantation in the Changzhou region of Jiangsu Province, including Fanglu Village within Changzhou City. The dataset offers 2.25 m spatial resolution and an image dimension of 348×512 pixels. It includes 80 spectral bands, covering wavelengths ranging from 417 to 855 nm, as referenced in [62].

All HSI datasets were standardized to unit variance, and labeled samples were divided into training and validation sets to mitigate biases from intrinsic data variability. Model parameters were optimized through iterative training, with hyperparameter adjustments informed by interim performance on the test set, particularly classification accuracy and loss. The model yielding the highest test accuracy, along with its associated weights, was retained for further analysis. The training process employed stochastic gradient descent (SGD) with a learning rate of 0.001, a momentum factor of 0.95, and a batch size of 32. Performance evaluation was based on overall accuracy (OA), average accuracy (AA), and the kappa coefficient (κ). To ensure statistical robustness, each experiment was independently conducted ten times, and the mean along with the standard deviation of each metric was reported.

All experiments were conducted on a workstation configured with 2.80 GHz Intel(R) Core i7-7700HQ CPUs, 8 GB RAM, and an NVIDIA GeForce GTX 1050 Ti GPU, running Windows 10 (64-bit). Most models were developed in Python 3.8 using TensorFlow 2.4.1, except for ASSMN and RSSAN, which were developed with PyTorch 1.11.

The effect of key hyperparameters on the classification performance of the proposed MDFAL model is systematically explored, including the time step T , the number of convolutional kernels, spatial input dimensions, and the ratio of training samples. To ensure evaluation robustness, the model achieving the highest validation performance was selected as a reference for comparative analysis across different configurations. For training efficiency, the dataset was divided into 32 mini-batches, and each configuration was trained for 300 epochs.

- 1) *Network architecture*: Table II presents the network topology of the proposed MDFAL model, providing a detailed breakdown of layer-wise parameters within the modular structure. This includes specifications such as the reduction ratio and expansion rate used in the spectral attention module. Importantly, these architectural parameters are held constant across all three experimental datasets to ensure consistency and enable precise comparative analysis.
- 2) *Time step (T)*: The time step plays a critical role in capturing spectral continuity in HSI, as it governs sequential relationships along the spectral dimension. This study investigates the effect of varying time steps on model performance, as summarized in Table II(d). Distinct time step settings were explored across the three datasets: for IN, values of 2, 4, 5, 8, 10, 20, and 40 were used; for PU,

TABLE II
PARAMETERIZATION OF THE MDFAL MODEL'S NETWORK LAYER

MDFAL (d=8)		
Block 1	conv	3*3 conv, 8
		Spectral attention, 1*1, 2, r = 4, relu
	SSAM	1*1, 8, sigmoid
		Spatial attention, 7*7, 2, r = 4, relu
	ConvLSTM	7*7, 8, sigmoid
	Reset block	3*3, 8, drop = 0.25
Block 2	Dila conv	3*3,
		3*3, 16, d r = 2
	conv	3*3 conv, 16
		Spectral attention, 1*1, 4, r = 4, relu
	SSAM	1*1, 16, sigmoid
		Spatial attention, 7*7, 4, rate = 4, relu
Block 3	ConvLSTM	7*7, 16, sigmoid
	Reset block	3*3, 16, drop=0.25
	Dila conv	3*3
		3*3, 32, d r = 2
	conv	3*3 conv, 32
		Spectral attention, 1*1, 8, r = 4, relu
Block 3	Attention block	1*1, 32, sigmoid
		Spatial attention, 7*7, 8, r = 4, relu
	ConvLSTM	7*7, 32, sigmoid
	Reset block	3*3, 32, drop=0.25
	Dila conv	3*3,
		3*3, 64, d r = 2

TABLE III
INFLUENCE OF FILTER COUNT ON MODEL PERFORMANCE

	Filter num	OA	AA	Kappa
IN	(4,8,16,32)	98.86	98.07	98.27
	(8,16,32,64)	99.04	98.66	98.91
	(16,32,64,128)	99.54	99.32	99.33
PU	(4,8,16,32)	97.99	96.42	97.34
	(8,16,32,64)	98.67	97.51	98.24
	(16,32,64,128)	98.51	97.21	98.03
Tea	(4,8,16,32)	98.43	96.34	97.68
	(8,16,32,64)	98.72	96.83	98.10
	(16,32,64,128)	99.16	97.44	97.79

2, 4, 5, 10, and 20; and for Tea, 2, 4, 5, 8, 16, and 20. The experimental results indicate that variations in temporal step size significantly affect the overall classification performance.

Fig. 4 demonstrates that classification accuracy across the three experimental datasets is variably affected by the choice of time step. The IN dataset shows lower sensitivity to changes in T compared with PU and Tea. Moreover, empirical trends indicate that the optimal T tends to increase with longer training durations. Specifically, the optimal time steps are 8 for IN, and 5 for both PU and Tea, underscoring the importance of temporal resolution in determining classification performance.

- 1) *Number of filters*: The capacity of the model to extract representative features and maintain computational efficiency is strongly influenced by the number of convolutional filters. To determine the optimal configuration, we evaluated filter settings of $d = 4, 8$, and 16. As illustrated in Table III, the IN and Tea datasets exhibited consistent gains in accuracy with increasing filter numbers, followed

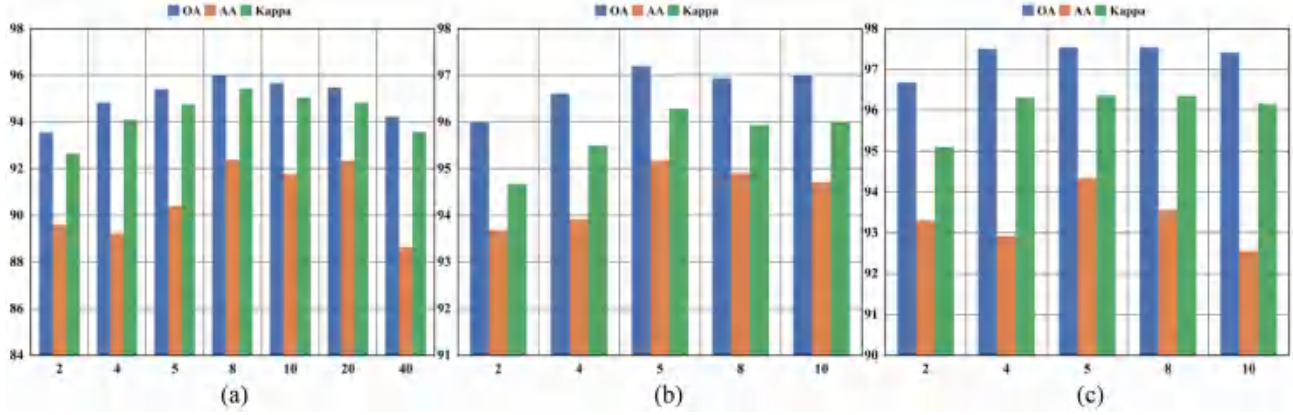


Fig. 4. Effect of time step on a model's performance. (a) IN. (b) PU. (c) Tea.

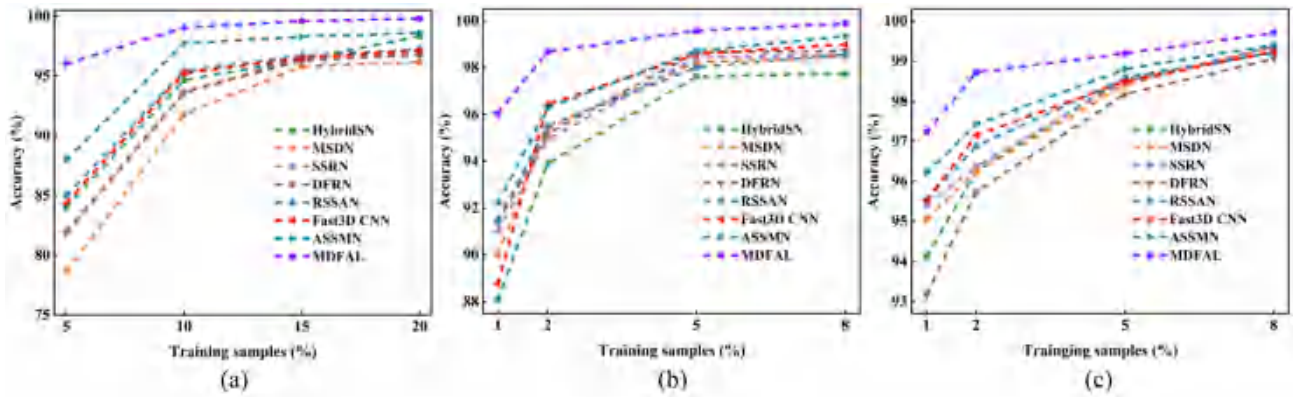


Fig. 5. Variation in OA across the three datasets with varying sample sizes. (a) IN. (b) PU. (c) Tea.

by a saturation point. In contrast, the PU dataset showed a slight decline in performance beyond a certain threshold, although a modest overall improvement was still observed. The initial d was set to 8 for all subsequent comparative experiments across the three datasets.

- 2 *Training sample proportion*: Model performance is positively correlated with the proportion of labeled training samples. To examine this relationship, we systematically evaluated the classification performance of the proposed MDFAL model across varying training ratios using three HSI datasets. Training subsets were selected using fixed random seeds, with proportions of 5%, 10%, 15%, and 20% for the IN dataset, and 1%, 2%, 5%, and 8% for the PU and Tea datasets. As shown in Fig. 5, OA consistently improved with increasing training sample size across all datasets. Under identical sampling conditions, the MDFAL model outperformed seven state-of-the-art methods, achieving the highest OA in every case. In the subsequent comparative experiments, the training sample proportions were fixed at 10%, 2%, and 2% for the IN, PU, and Tea datasets, respectively.

B. Comparison With the State-of-the-Art

In this study, we evaluate the classification performance of the proposed MDFAL model against seven state-of-the-art

deep learning approaches for HSI classification under consistent experimental conditions. The comparative models include the HybridSN [63], MSDN [32], SSRN [27], DFRN [30], RSSAN [31], Fast 3D CNN [64], and ASSMN [43]. MSDN extracts multiscale features using 3-D convolutions applied at varying depths, whereas HybridSN integrates 2-D and 3-D convolutions to extract spectral and spatial representations. Both DFRN and SSRN employ 3-D convolutions in conjunction with residual connections to enrich feature encoding. ASSMN further incorporates a ConvLSTM architecture to model spectral and spatial dependencies across multiple scales. For all experiments conducted on the IN, PU, and Tea datasets, the neighborhood window sizes were set to 9, 7, and 7, respectively, while all other parameters remained consistent with those described earlier.

Fig. 6(a)–(h) illustrates the classification outcomes on the IN dataset produced by several advanced comparison methods. Panels (b), (d), (f), and (g) reveal varying degrees of classification noise, with panel (b) exhibiting the most severe artifacts and the lowest accuracy. Both MSDN and DFRN based on 3-D CNN architectures exploit multiscale hierarchical features for classification, however, their outcomes differ markedly. Notably, DFRN, which incorporates residual learning, achieves superior noise suppression compared to MSDN, as seen in panel (b). Furthermore, combining residual learning with SSRN and RSSAN produces classification maps with smoother boundaries and reduced noise levels, outperforming the remaining approaches.

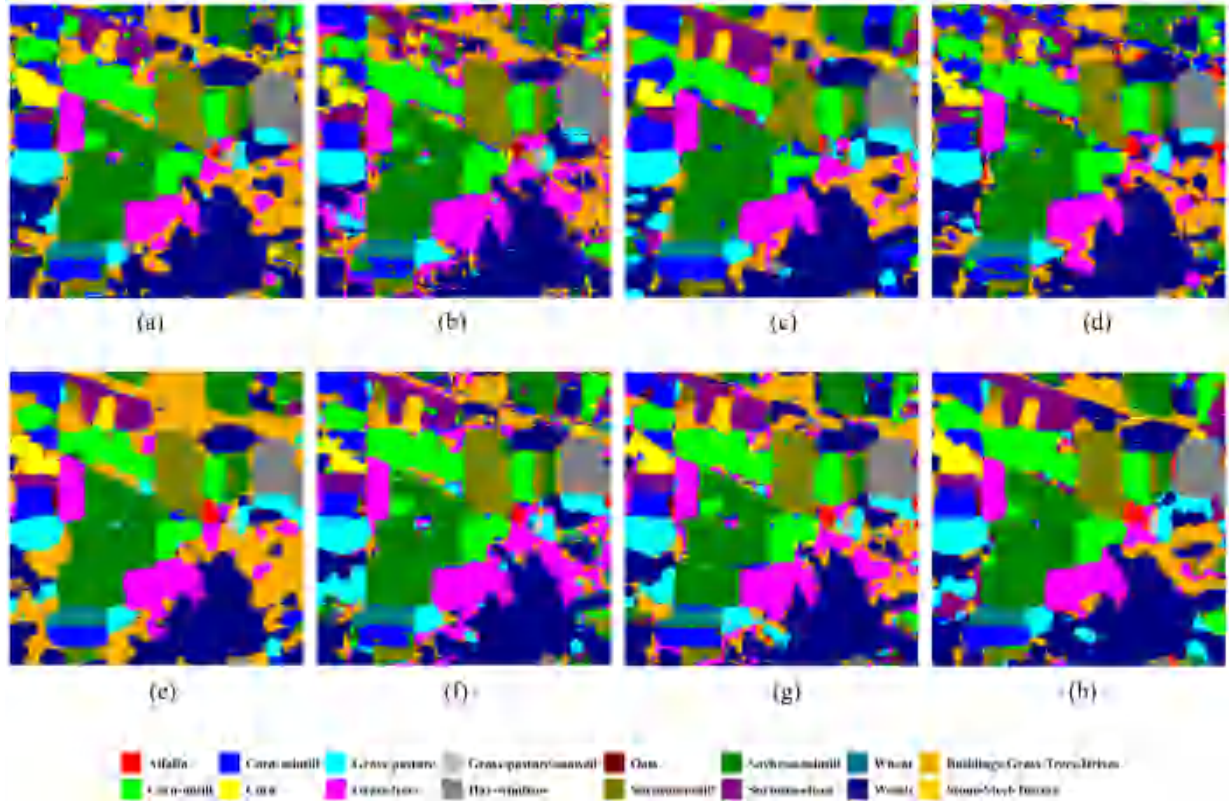


Fig. 6. IN dataset classification via different strategies. (a)–(h) Outcomes from the HybridSN, MSDN, SSRN, DFRN, RSSAN, Fast 3-D CNN, ASSMN, and MDFAL.

These results highlight the effectiveness of residual connections in enhancing spatial consistency and improving overall classification accuracy.

Table IV presents a comparative evaluation of the IN dataset. Among existing baseline models, ASSMN featuring a ConvLSTM backbone that demonstrates competitive performance, achieving an OA of 97.74%, AA of 96.46%, and a κ of 97.42%. Although panels (e) and (g) yield the highest AA individually, RSSAN exhibits substantially lower OA and Kappa values compared to ASSMN, highlighting the advantage of temporal modeling via ConvLSTM in capturing spectral dependencies inherent in hyperspectral data. By comparison, the proposed MDFAL model, while showing weaker performance in a few classes such as Stone-Steel-Towers, outperforms all benchmark methods across three quantitative metrics. It achieves an OA of 99.08%, an AA of 98.69%, and a Kappa coefficient of 98.97%, representing gains of 1.34, 2.23, and 1.54 percentage points, respectively, over the strongest baseline, ASSMN. Moreover, MDFAL exhibits the lowest variance among competing algorithms, underscoring its robustness and stability. These gains underscore the efficacy of MDFAL's architectural innovations, which integrate shallow, intermediate, and deep multiscale receptive fields while concurrently optimizing global attention mechanisms and adaptive loss weighting across diverse modules and tasks.

Fig. 7(a)–(h) shows the classification outcomes on the PU dataset using a suite of representative state-of-the-art models.

Compared to panels (a), (b), and (e), the classification maps in panels (c), (d), (f), and (g) demonstrate strongly improved delineation with reduced background noise. The *Bare Soil* category is most effectively segmented by SSRN, ASSMN, and MDFAL, attributable to their enhanced spatial coherence and superior suppression of irrelevant features. Among all models, the proposed MDFAL framework delivers the most accurate delineation of the *Bricks* class, producing sharply defined and spatially consistent boundaries. Particularly, the interface between *Trees* and *Bare Soil* is resolved with the highest classification accuracy, highlighting MDFAL's efficacy in capturing nuanced spectral and spatial transitions. The spatial continuity observed in smooth-surface regions closely resembles that in Fig. 8(a), further validating the robustness of the proposed approach.

Table V summarizes the comparative results on the PU dataset. The proposed MDFAL model achieves notable gains over competing methods, improving OA by 2.25%, AA by 2.49%, and the κ by 2.66%. The MDFAL model achieves superior classification accuracy across most categories, including Gravel, Trees, and Bitumen. Although it performs slightly inferior to the best methods for the Metal Sheets and Bricks classes, its overall performance consistently exceeds that of the majority of benchmark approaches. Furthermore, MDFAL exhibits the lowest variance among all compared algorithms, underscoring its robustness and stability. These results, combined with superior visual fidelity, underscore MDFAL's effectiveness in estimating fine-scale spatial structures and delineating complex boundary

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS ON THE IN DATASET

Class	HybridSN	MSDN	SSRN	DFRN	RSSAN	Fast 3D CNN	ASSMN	MDFAL
1	34.15	65.85	45.45	97.56	100.00	82.93	86.11	95.12
2	95.25	90.58	85.56	92.92	83.42	96.11	98.94	98.29
3	91.42	88.06	99.49	86.48	96.92	90.35	97.86	100
4	100	68.07	66.67	81.69	96.71	80.75	95.19	96.24
5	92.40	93.54	92.59	94.47	96.55	95.85	98.16	100
6	99.09	99.39	99.42	99.54	99.85	100	99.48	99.70
7	92.00	92.00	7.41	80.00	92.00	92.00	81.82	100
8	100	99.77	100	100	100	100	100	100
9	94.12	82.35	0	100	100	64.71	100	100
10	93.36	98.39	87.32	91.88	90.51	95.77	98.04	99.08
11	95.88	91.67	97.90	95.47	99.55	95.29	95.72	98.96
12	82.18	75.80	90.94	72.42	95.51	87.43	97.01	97.94
13	99.46	100	100	100	100	100	99.38	100
14	99.47	97.89	99.67	100	96.22	99.30	99.60	99.82
15	86.46	88.18	97.28	98.56	100	92.51	96.05	98.27
16	97.59	59.04	96.59	97.59	100	98.80	100	95.18
OA	94.64	91.80	93.68	93.61	95.43	95.23	97.74	99.08
	± 0.81	± 0.79	± 0.75	± 0.64	± 0.59	± 0.41	± 0.38	± 0.24
AA	90.80	86.91	79.14	93.04	96.30	91.99	96.46	98.69
	± 1.02	± 0.96	± 0.67	± 0.59	± 0.65	± 0.55	± 0.45	± 0.25
κ	93.88	90.64	92.78	92.69	94.79	94.57	97.42	98.97
	± 0.78	± 0.81	± 0.82	± 0.77	± 0.64	± 0.57	± 0.49	± 0.37

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS ON THE PU DATASET

Class	HybridSN	MSDN	SSRN	DFRN	RSSAN	Fast 3-D CNN	ASSMN	MDFAL
1	95.41	95.38	95.41	94.51	95.95	96.68	98.89	98.45
2	98.01	99.27	99.11	99.45	98.38	99.35	98.16	99.98
3	78.61	72.29	69.68	85.51	73.05	90.23	92.32	95.33
4	91.51	98.17	90.38	95.90	93.66	93.04	87.80	98.47
5	100	99.85	100	97.12	99.77	100	100	99.92
6	90.48	94.32	99.72	92.35	94.68	98.78	99.54	99.72
7	76.36	81.04	46.63	73.60	84.01	91.17	86.22	91.25
8	89.63	90.99	98.78	90.60	90.74	91.05	88.24	95.15
9	94.83	99.68	99.88	95.26	96.98	98.89	99.89	99.35
OA	93.90	95.43	94.92	95.18	94.90	95.42	96.28	98.67
AA	± 1.01	± 0.93	± 0.88	± 0.79	± 0.82	± 0.76	± 0.64	± 0.36
	90.54	92.33	88.84	91.59	91.91	93.02	94.56	97.51
	± 0.94	± 0.87	± 0.75	± 0.65	± 0.77	± 0.65	± 0.55	± 0.35
κ	91.89	93.93	93.24	93.60	93.22	93.68	95.07	98.24
	± 0.97	± 0.82	± 0.77	± 0.57	± 0.69	± 0.57	± 0.46	± 0.24

transitions. This performance is predominantly attributed to its integration of hierarchical features across shallow, intermediate, and deep levels, enabled by expanded receptive fields and refined attention mechanisms.

Fig. 8 compares classification outcomes on the Tea dataset, where SSRN, ASSMN, and MDFAL achieve smoother, less noisy, and more coherent maps than other methods, possibly due

to their multiscale feature learning modules. Notably, ASSMN and MDFAL deliver the highest connectivity and completeness in the Building and Weed categories, while models such as HybridSN, MSDN, DFRN, and Fast 3-D CNN exhibit greater fragmentation.

Table VI summarizes the quantitative comparison, where the proposed MDFAL model surpasses all baselines by 1.34%,

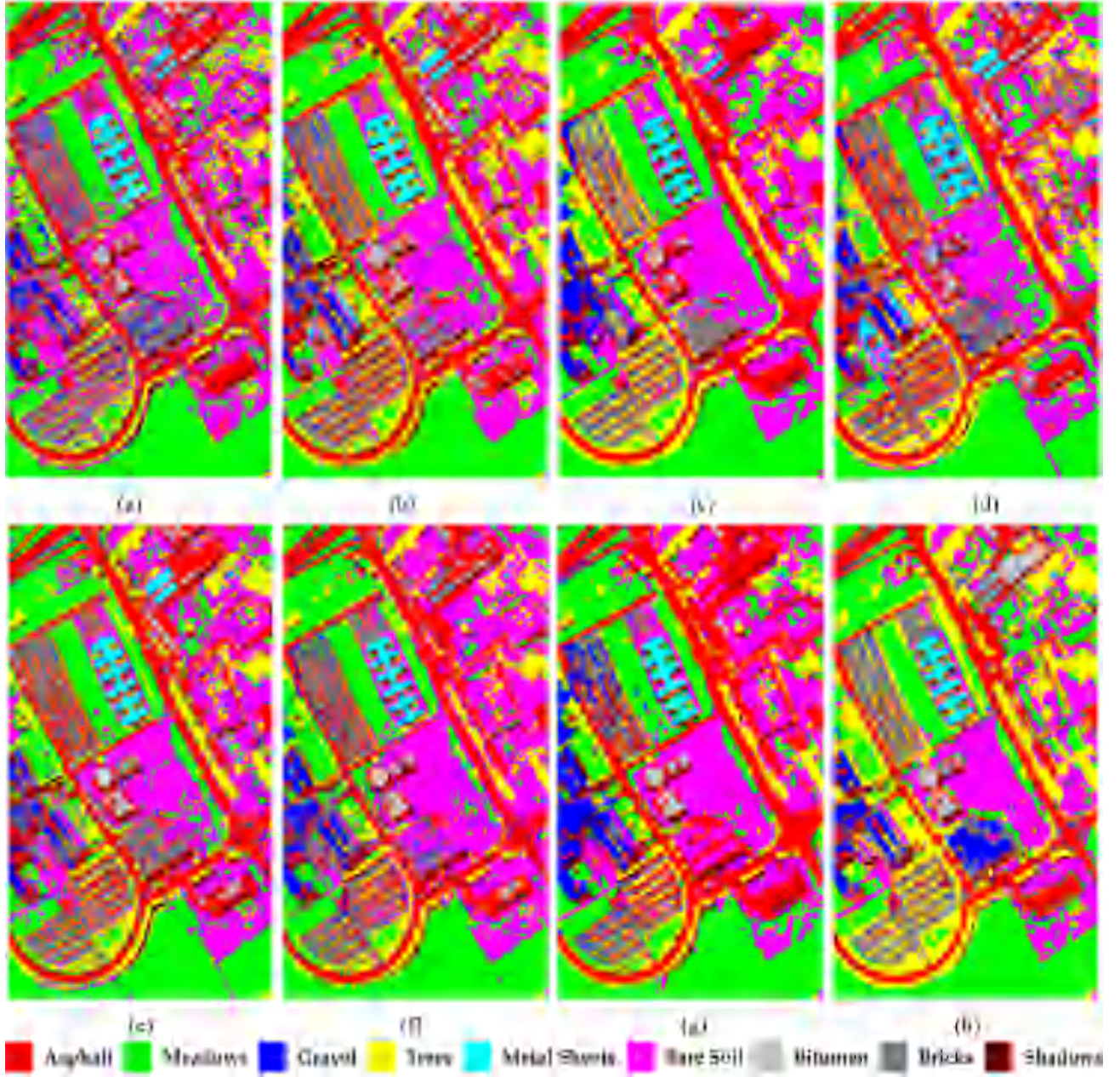


Fig. 7. PU dataset classification via different strategies. (a)–(h) Outcomes from the HybridSN, MSDN, SSRN, DFRN, RSSAN, Fast 3-D CNN, ASSMN, and MDFAL.

2.67%, and 1.92% in OA, AA, and κ , respectively. While MDFAL performs slightly below the top method in the *Building* category, it achieves leading accuracy in classes with sparse annotations, such as *Reed* and *Weed*. These results demonstrate the model's robustness in low-resource scenarios and its ability to generalize across heterogeneous spatial features. Collectively, MDFAL strikes an effective balance between statistical accuracy and visual fidelity, underscoring its suitability for real-world hyperspectral classification tasks.

Across all experimental comparisons, the MDFAL model consistently produced cleaner classification maps, characterized by reduced speckle noise and sharper, smoother boundaries. Quantitative evaluations using OA, AA, and the κ further

confirmed its effectiveness. Although MDFAL achieved slightly lower accuracy than certain algorithms in a few specific classes, it outperformed all compared methods on all three-evaluation metrics across the three datasets. Moreover, its variance was the lowest, indicating that MDFAL offers not only superior classification performance but also greater stability.

IV. DISCUSSION

To further assess the effectiveness of the proposed MDFAL model, this section presents a series of experiments, including convergence analysis, ablation studies, runtime comparison, and feature visualization on benchmark datasets. Unless otherwise

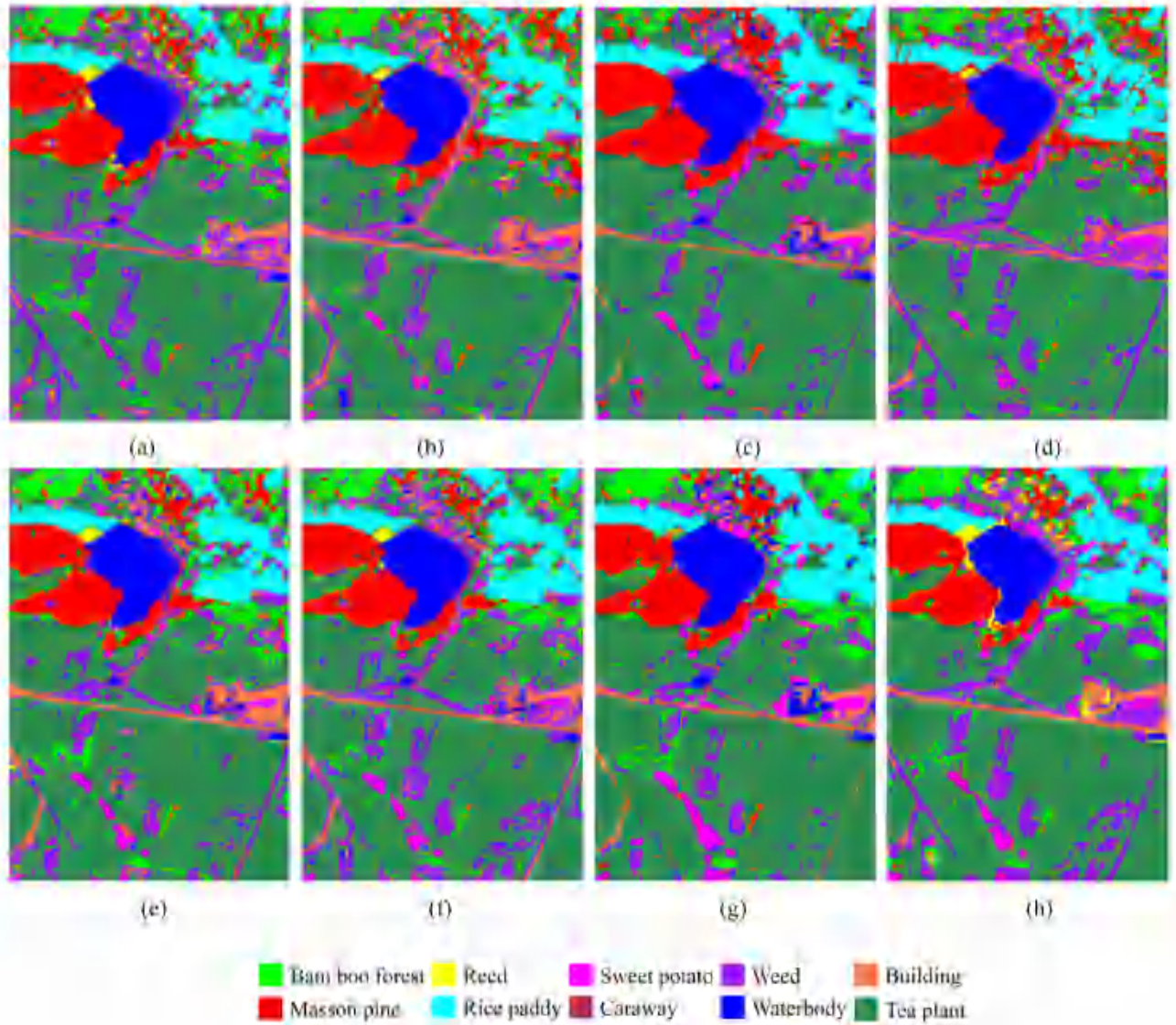


Fig. 8. Tea dataset classification via different strategies. (a)–(h) Outcomes from the HybridSN, MSDN, SSRN, DFRN, RSSAN, Fast 3-D CNN, ASSMN, and MDFAL.

specified, all experimental settings are consistent with those described in Section III.

A. Convergence Analysis

Fig. 9 illustrates the training trajectories of accuracy and loss across the three datasets. Despite noticeable fluctuations over the course of training, a consistent improvement in classification accuracy is observed. The IN and Tea datasets exhibit greater variability in validation loss than the PU dataset. Nevertheless, all models converge reliably, with performance stabilizing within the first 75 epochs.

B. Runtime Comparison

Table VII summarizes the training and testing times of the compared models on the IN, PU, and Tea benchmark datasets, each trained for 300 iterations. DFRN and MDFAL record the longest training times, as both compute losses from multilayer

feature outputs, which increases computational overhead. For MDFAL, the ConvLSTM module further extends runtime by segmenting HSI data along the spectral dimension for sequential feature learning. By contrast, other models use simpler architectures that compute loss only from the final output, thereby reducing complexity and computational cost.

C. Ablation Analysis

Ablation analysis was performed to quantitatively assess the individual contribution of each component within the MDFAL framework to overall classification performance. In the baseline configuration, each of the three modules consists solely of an inflated convolutional layer that was independently connected to the final fully connected layer for evaluation.

As presented in Table VIII, the baseline MDFAL model achieves overall accuracies of 91.17%, 91.10%, and 88.72% on the IN, PU, and Tea datasets, respectively. These results

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS ON THE TEA DATASET

Class	HybridSN	MSDN	SSRN	DFRN	RSSAN	Fast 3-D CNN	ASSMN	MDFAL
1	98.05	95.94	99.95	98.19	98.61	98.82	98.51	98.49
2	74.24	73.67	46.65	65.74	79.96	81.68	96.13	93.22
3	98.10	98.47	99.79	98.28	98.20	98.01	98.96	99.42
4	99.04	85.65	0	77.14	100	96.04	39.05	99.52
5	99.36	99.63	100	99.19	99.37	99.43	100	99.70
6	79.50	92.63	95.88	95.13	92.87	90.25	99.63	93.88
7	80.71	81.67	88.12	63.57	97.86	91.67	98.57	95.24
8	79.05	71.31	81.25	76.04	73.07	83.87	58.55	94.19
9	99.83	99.93	100	99.40	99.92	99.95	100	99.42
10	92.04	98.99	99.44	86.67	99.33	98.21	99.78	94.51
OA	96.24	96.24	96.37	95.74	96.87	96.15	97.37	98.71
	± 0.69	± 0.64	± 0.57	± 0.74	± 0.56	± 0.51	± 0.47	± 0.31
AA	89.99	89.79	81.11	85.94	93.92	92.09	88.92	96.83
	± 0.54	± 0.57	± 0.45	± 0.58	± 0.65	± 0.45	± 0.55	± 0.35
κ	94.43	94.42	94.55	93.66	95.38	94.78	96.16	98.10
	± 0.58	± 0.67	± 0.50	± 0.63	± 0.58	± 0.47	± 0.57	± 0.34

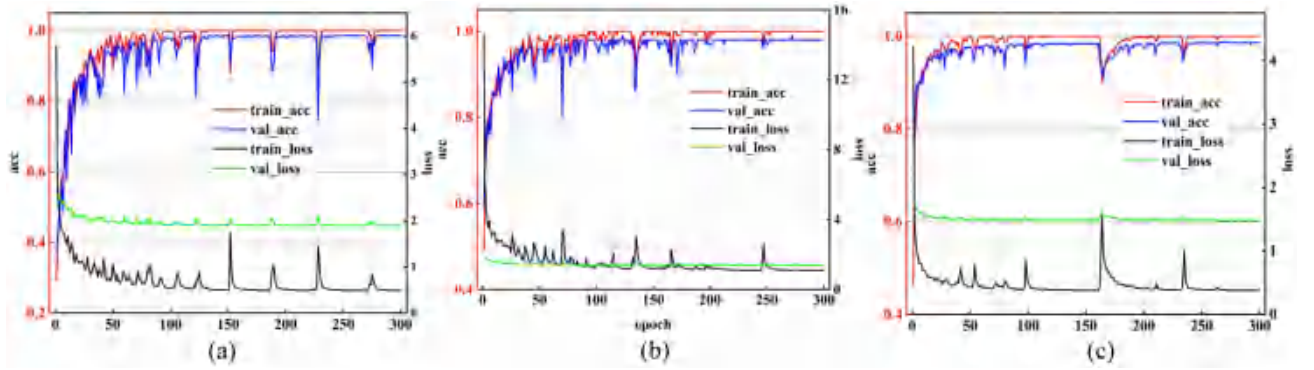


Fig. 9. Accuracy and loss curves while the system runs. (a) IN. (b) PU. (c) Tea.

TABLE VII
DURATION OF VARIOUS MODELS' OPERATION

		HybridSN	MSDN	SSRN	DFRN	RSSAN	Fast 3-D CNN	ASSMN	MDFAL
IN	Train. (m)	89.13	246.25	110.52	753.55	65.67	91.68	85.32	699.01
	Test. (s)	122.75	82.97	81.53	248.78	45.98	44.33	42.16	95.29
PU	Train. (m)	115.39	642.10	277.58	1082.82	55.94	98.15	118.00	939.49
	Test. (s)	106.73	538.80	207.61	119.56	97.50	155.83	186.99	858.01
Tea	Train. (m)	103.62	288.51	146.28	998.08	60.03	93.95	98.63	982.04
	Test. (s)	94.72	168.46	326.23	385.07	56.39	86.95	89.67	63.37

underscore the model's strong foundational performance, which can be attributed to its ability to integrate multiscale features across shallow, intermediate, and deep layers, thereby enhancing loss computation and gradient optimization. On the IN dataset, the addition of the CCLM improves the OA, AA, and κ by 2.83%, 0.95%, and 3.24%, respectively, Line *b* demonstrating its effectiveness in boosting classification performance. Comparable improvements on the PU and Tea datasets further confirm

the CCLM's ability to capture both spectral continuity and sequential context.

As indicated in Table VIII, row (c), augmenting the baseline model with the SSAM markedly improves classification performance, particularly for the Tea dataset. This gain is likely attributable to SSAM's capacity to enhance feature discrimination by emphasizing informative spectral-spatial cues. Furthermore, row (d) shows that combining SSAM with the contextual learning module (CCLM) leads to additional performance gains across all three datasets. These findings suggest that CCLM synergistically complements SSAM by further refining salient features and adaptively adjusting their contributions during training.

D. Visualization of Discriminative Features

Table VIII(e) summarizes the classification outcomes of the MDFAL model following parameter tuning and hierarchical training across shallow, intermediate, and deep layers, along with the corresponding Center loss values for each module.

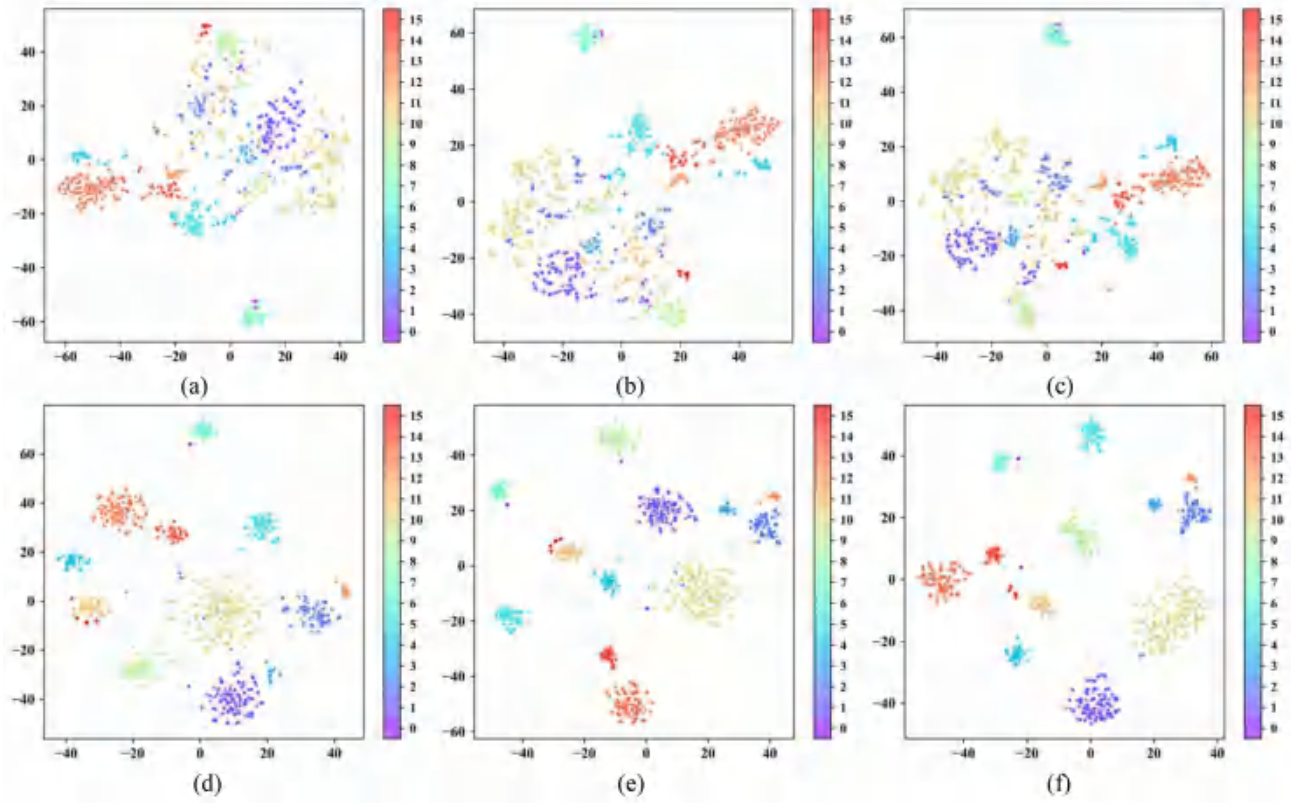


Fig. 10. Feature distance visualization of test samples in the IN dataset. The characteristics of test samples from the first, second, and third module layers are shown in sections (a)–(c) of Table VIII(d), while the corresponding integrated features from the complete model are presented in section (d)–(f) of Table VIII(f).

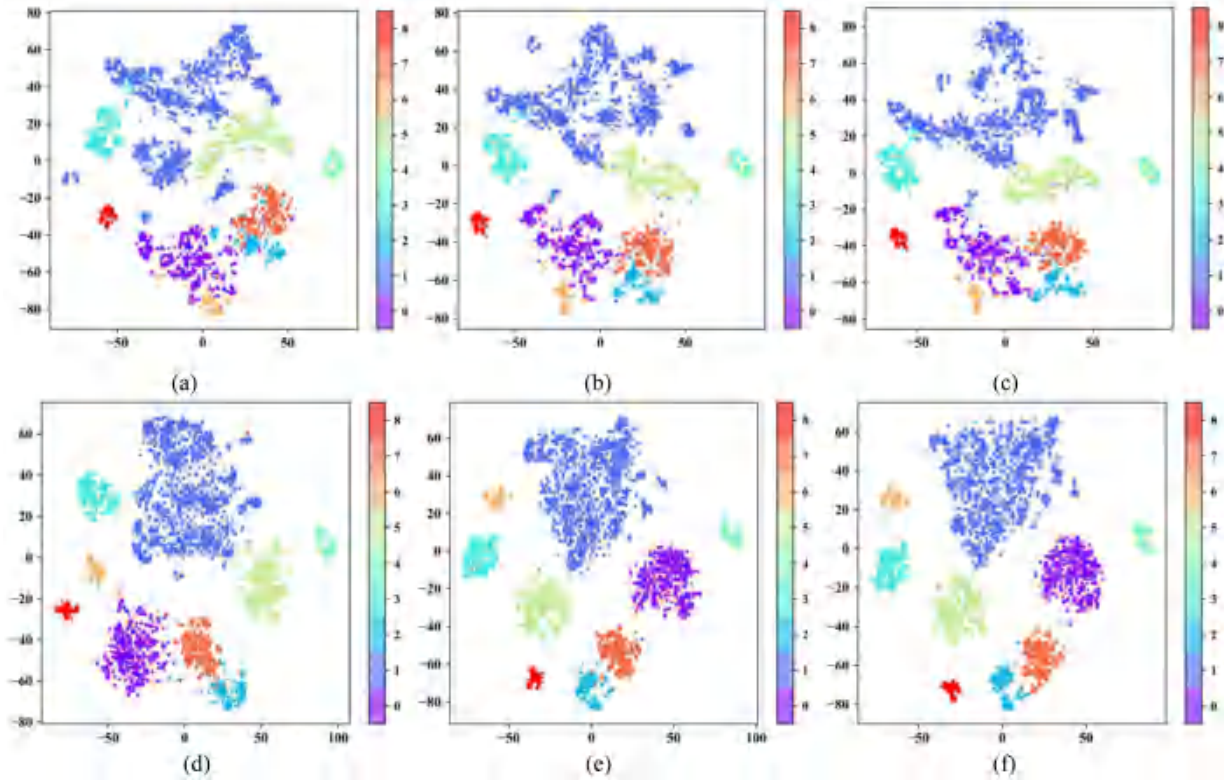


Fig. 11. Feature distance visualization of test samples in PU dataset. The characteristics of test samples from the first, second, and third module layers are detailed in sections (a)–(c) of Table VIII(d), while the corresponding integrated features within the model are shown in section (d)–(f) of Table VIII(f).

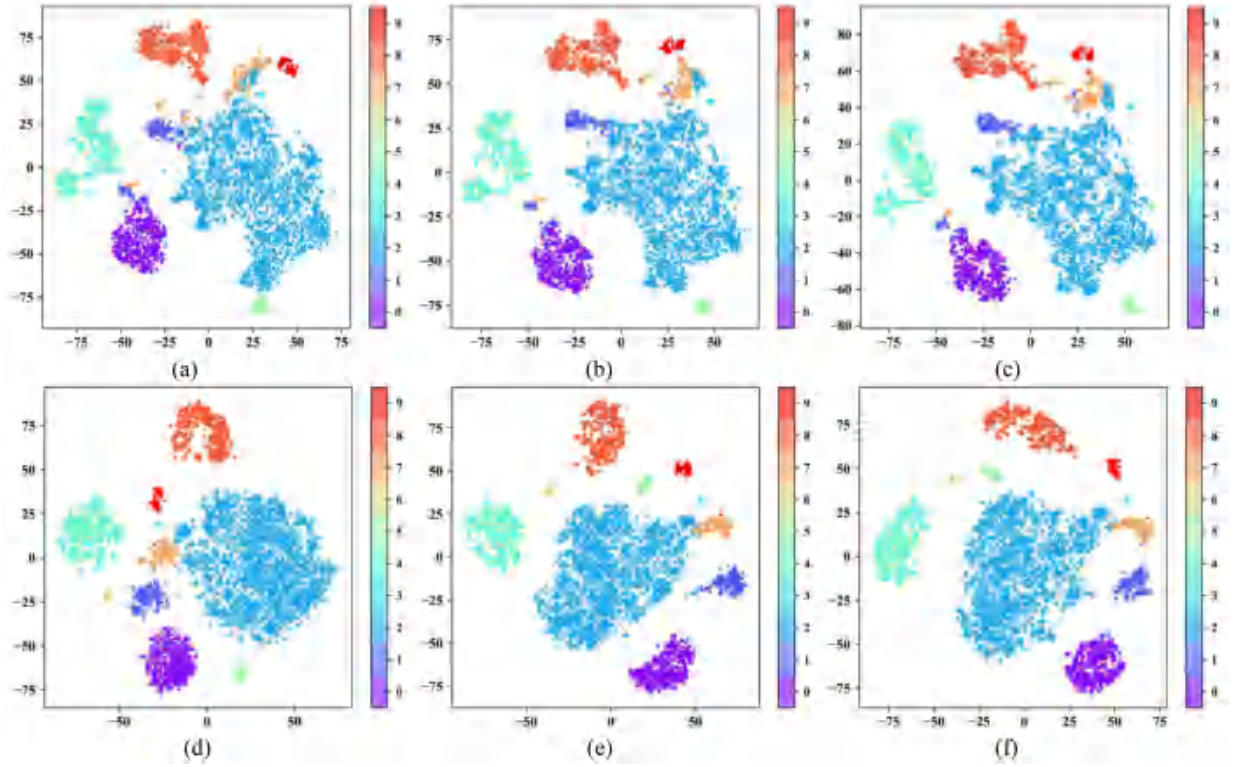


Fig. 12. Feature distance visualization of test samples in the Tea dataset. The characteristics of test samples from the first, second, and third module layers are detailed in sections (a)–(c) of Table VIII(d), while the corresponding integrated features within the model are presented in sections (d)–(f) of Table VIII(f).

TABLE VIII
PRECISION WITH DIFFERENT SETTINGS IN MDFAL

	method	convlstm	attention	Center loss	Adaptive weight	OA	AA	Kappa
IN	(a) baseline	—	—	—	—	91.17	87.97	89.91
	(b) baseline+convlstm	✓	—	—	—	94.00	88.92	93.15
	(c) baseline+attention	—	✓	—	—	95.78	92.41	95.18
	(d) baseline+convlstm+ attention	✓	✓	—	—	95.85	91.44	95.26
	(e) MDFAL w/o adaptive weight	✓	✓	✓	—	98.66	97.05	98.48
	(f) MDFAL	✓	✓	✓	✓	99.08	98.69	98.97
PU	(a) baseline	—	—	—	—	91.10	87.37	88.13
	(b) baseline+convlstm	✓	—	—	—	94.80	91.57	93.11
	(c) baseline+attention	—	✓	—	—	94.71	91.46	92.98
	(d) baseline+convlstm+ attention	✓	✓	—	—	97.20	95.18	96.28
	(e) MDFAL w/o adaptive weight	✓	✓	✓	—	98.15	97.07	98.08
	(f) MDFAL	✓	✓	✓	✓	98.67	97.51	98.24
Tea	(a) baseline	—	—	—	—	88.72	46.16	82.73
	(b) baseline+convlstm	✓	—	—	—	96.80	91.75	95.28
	(c) baseline+attention	—	✓	—	—	97.37	91.75	95.28
	(d) baseline+convlstm+ attention	✓	✓	—	—	97.54	94.35	96.37
	(e) MDFAL w/o adaptive weight	✓	✓	✓	—	98.05	96.12	97.33
	(f) MDFAL	✓	✓	✓	✓	98.72	96.83	98.10

On the IN dataset, substantial improvements were recorded, with OA rising from 95.85% to 98.66%, AA from 91.44% to 97.05%, and the Kappa coefficient from 95.26% to 98.48%. Comparable gains were observed on the PU dataset, with respective increases of 0.95%, 1.89%, and 1.80% in OA, AA, and κ . The Tea dataset demonstrated consistent improvements across all three metrics. These results suggest that coordinated multi-task loss optimization enhances classification performance by simultaneously minimizing loss at each layer and refining the

extraction of discriminative features. Table VIII(f) presents the outcomes of an adaptive weight adjustment strategy incorporated into the multitask loss framework. Across all datasets, this approach consistently yielded further accuracy gains, underscoring its usefulness in dynamically balancing learning contributions across network components.

To evaluate the representational capacity of the proposed MD-FAL model, t-distributed stochastic neighbor embedding was used to visualize the high-dimensional discriminative features

of test samples into a two-dimensional space. The resulting visualizations, corresponding to the configurations in Table VIII(d) and (f), are shown in Figs. 10–12, with each class assigned a distinct color. In the baseline setting [see Table VIII(d)], the feature distributions show weak intraclass compactness and substantial overlap between classes, reflecting limited discriminative power. By contrast, the MDFAL model [see Table VIII(f)] yields feature embeddings with clearly delineated interclass boundaries and dense intraclass clusters. These findings further confirm the effectiveness of the proposed MDFAL model in extracting and encoding abstract spectral–spatial representations. It not only captures discriminative features inherent in hyperspectral data but also significantly enhances class separability and overall classification performance.

These results demonstrate that the proposed MDFAL model not only exhibits consistent and rapid convergence across benchmark datasets but also delivers significant performance improvements through its integrated architecture. The ablation experiments reveal that each feature extraction component, together with the shallow, intermediate, and deep hierarchical discriminative outputs and the adaptive multitask loss weighting strategy, contributes meaningfully to enhancing model performance. Moreover, the visualization of discriminative features further corroborates the effectiveness of the adaptive optimization mechanism in improving the network’s representational capability.

Nevertheless, the MDFAL network also has certain limitations, such as a relatively long computational runtime, which highlights the need for further optimization and lightweight architectural design in future work. As shown in Fig. 5, when trained with 5%, 1%, and 1% of the available samples from the IN, PU, and Tea datasets, respectively, the MDFAL model achieves overall classification accuracy improvements of 9.12%, 4.17%, and 1.05%. These findings indicate that the proposed MDFAL model offers substantial advantages over other state-of-the-art approaches, particularly under conditions of limited and imbalanced training data, and demonstrates superior performance on datasets with complex feature characteristics.

V. CONCLUSION

This study proposes a multiscale discriminative feature adaptive learning framework (MDFAL) for HSI classification, featuring two core innovations. First, MDFAL employs a cascaded architecture of shallow, intermediate, and deep layers that jointly leverage multiscale features under a multiloss training strategy. Second, it integrates the SSAM, CCLM, and dilated convolution modules, combined with a dynamic multitask adaptive weighting mechanism. Extensive experiments on the IN, PU, and Tea datasets demonstrate that, although MDFAL performs slightly worse in a few classes compared with certain methods, it effectively learns multiscale discriminative features with stronger intraclass compactness and interclass separability, while exhibiting superior generalization and stability. However, the method remains computationally demanding, primarily due to the co-output training strategy of the multilayer architecture, the sequential nature of ConvLSTM operations, and the high

spatial redundancy inherent in pixel-block representations. Future research will explore more efficient network architectures and emerging techniques such as fully convolutional networks, Mamba, few-shot learning, semisupervised learning, and multimodal data integration, to further improve both computational efficiency and classification accuracy.

ACKNOWLEDGMENT

The datasets from the Indiana Pines and the University of Pavia may be found on the internet at http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. The Tea datasets may be found online at <http://www.geodai.ac.cn/WebCn/doi.aspx?Id=720>.

Author Contributions: L. Liu: Conceptualization, Methodology, Software, Validation, Data curation, Writing—original draft. S. Jin: Writing—review & editing. A. Banerjee: Conceptualization. C. Zhang: Data curation.

REFERENCES

- [1] M. Amani et al., “Remote Sensing Systems for Ocean: A Review (Part 1: Passive Systems),” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 210–234, 2022.
- [2] Z. Jiang et al., “Cross-domain hyperspectral image classification,” *Pattern Recognit.*, vol. 168, 2025, Art. no. 111836.
- [3] Z. Zheng, S. Du, H. Taubenböck, and X. Zhang, “Remote sensing techniques in the investigation of Aeolian sand dunes: A review of recent advances,” *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112913.
- [4] J. Arndt and D. Lunga, “Large-Scale Classification of Urban Structural Units From Remote Sensing Imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2634–2648, 2021.
- [5] L. Liu, C. Zhang, W. Luo, S. Chen, F. Yang, and J. Liu, “New remote sensing image fusion for exploring spatiotemporal evolution of urban land use and land cover,” *J. Appl. Remote Sens.*, vol. 16, 2022.
- [6] L. Karthikeyan, I. Chawla, and A. K. Mishra, “A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses,” *J. Hydrol.*, vol. 586, 2020, Art. no. 124905.
- [7] M. Kucharczyk and C. H. Hugenholtz, “Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities,” *Remote Sens. Environ.*, vol. 264, 2021, Art. no. 112577.
- [8] H. Wu et al., “A novel remote sensing ecological vulnerability index on large scale: A case study of the China-Pakistan Economic Corridor region,” *Ecological Indicators*, vol. 129, 2021, Art. no. 107955.
- [9] M. E. Paoletti, J. M. 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.
- [10] C. Fu, T. Zhou, T. Guo, Q. Zhu, F. Luo, and B. Du, “CNN-Transformer and Channel-Spatial Attention based network for hyperspectral image classification with few samples,” *Neural Netw.*, vol. 186, 2025, Art. no. 107283.
- [11] M. A. Moharram and D. M. Sundaram, “Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions,” *Neurocomputing (Amsterdam)*, vol. 536, pp. 90–113, 2023.
- [12] 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.
- [13] M. Xu, P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, “Decision tree regression for soft classification of remote sensing data,” *Remote Sens. Environ.*, vol. 97, pp. 322–336, 2005.
- [14] 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.
- [15] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, “Spectral–spatial classification of hyperspectral imagery based on partitioned clustering techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

- [16] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [17] C. Shi and C. Pun, "Superpixel-based 3D deep neural networks for hyperspectral image classification," *Pattern Recognit.*, vol. 74, pp. 600–616, 2018.
- [18] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [19] Z. Li et al., "Cycle self-training with joint adversarial for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5531117.
- [20] Y. Song et al., "Deep learning for hyperspectral image classification: A comprehensive review and future predictions," *Inf. Fusion*, vol. 123, 2025, Art. no. 103285.
- [21] S. Liu, C. Fu, Y. Duan, X. Wang, and F. Luo, "Spatial-spectral enhancement and fusion network for hyperspectral image classification with few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5502414.
- [22] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [23] 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.
- [24] W. Luo, C. Zhang, Y. Li, F. Yang, D. Zhang, and Z. Hong, "Deeply-supervised pseudo learning with small class-imbalanced samples for hyperspectral image classification," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 112, 2022, Art. no. 102949.
- [25] J. Wang, J. Zhou, and X. Liu, "Toward universal representation learning for multidomain hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5508116.
- [26] 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.
- [27] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [28] J. Mahmoodi, D. Abbasi-Moghadam, A. Sharifi, H. Nezamabadi-Pour, M. Esmaeili, and A. Vafaeinejad, "DESSA-net model: Hyperspectral image classification using an entropy filter with spatial and spectral attention modules on DeepNet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14588–14613, 2024.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [30] C. Zhang et al., "Deep feature aggregation network for hyperspectral remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5314–5325, 2020.
- [31] 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, Jan. 2021.
- [32] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9201–9222, Nov. 2019.
- [33] L. Linfeng, "Multisource remote sensing data and deep neural networks research for urban flood risk assessment," Ph.D. dissertation, Zhengzhou Univ., Henan, China, 2023.
- [34] S. Hao, W. Wang, and M. Salzmann, "Geometry-aware deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2448–2460, Mar. 2021.
- [35] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, Oct. 2017.
- [36] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [37] S. M. Mirhoseini Nejad, D. Abbasi-Moghadam, and A. Sharifi, "ConvLSTM-ViT: A deep neural network for crop yield prediction using earth observations and remotely sensed data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 17489–17502, 2024.
- [38] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral-spatial LSTMs," *Neurocomputing*, vol. 328, pp. 39–47, 2019.
- [39] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, 2017, Art. no. 1330.
- [40] W. Hu, H. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.
- [41] H. C. Li, W. S. Hu, W. Li, J. Li, Q. Du, and A. Plaza, "A(3) CLNN: Spatial, spectral and multiscale attention ConvLSTM neural network for multisource remote sensing data classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 747–761, Feb. 2022.
- [42] J. Feng et al., "Attention multibranch convolutional neural network for hyperspectral image classification based on adaptive region search," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5054–5070, Jun. 2021.
- [43] D. Wang, B. Du, L. Zhang, and Y. Xu, "Adaptive spectral-spatial multiscale contextual feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2461–2477, Mar. 2021.
- [44] Z. Zhang et al., "Attention residual hybrid network for unmanned aerial vehicles hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 7662–7681, 2025.
- [45] C. Fu, B. Du, and L. Zhang, "Do we need learnable classifiers? A hyperspectral image classification algorithm based on attention-enhanced ResBlock-in-ResBlock and ETF classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5505013.
- [46] H. Guo, K. Zhu, M. Tang, and J. Wang, "Two-level attention network with multi-grain ranking loss for vehicle re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4328–4338, Sep. 2019.
- [47] A. Sharifi and M. M. Safari, "Enhancing the spatial resolution of sentinel-2 images through super-resolution using transformer-based deep-learning models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 4805–4820, 2025.
- [48] Y. Wang, G. Liu, L. Yang, J. Liu, and L. Wei, "An attention-based feature processing method for cross-domain hyperspectral image classification," *IEEE Signal Process. Lett.*, vol. 32, pp. 196–200, 2025.
- [49] Z. Feng, X. Liu, S. Yang, K. Zhang, and L. Jiao, "Hierarchical feature fusion and selection for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5501205.
- [50] Z. Zhao, D. Hu, H. Wang, and X. Yu, "Center attention network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3415–3425, 2021.
- [51] S. Li, X. Luo, Q. Wang, L. Li, and J. Yin, "H2AN: Hierarchical homogeneity-attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5509816.
- [52] G. Zhang, L. Zhang, Z. Zhang, J. Deng, L. Bian, and C. Yang, "DECT: Diffusion-enhanced CNN-transformer for multisource remote sensing data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 19288–19301, 2024.
- [53] R. Ming, N. Chen, J. Peng, W. Sun, and Z. Ye, "Semantic tokenization-based mamba for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 4227–4241, 2025.
- [54] Z. Pan, C. Li, A. Plaza, J. Chanussot, and D. Hong, "Hyperspectral image classification with Mamba," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5602814.
- [55] X. Geng, L. Li, L. Jiao, X. Liu, F. Liu, and S. Yang, "Knowledge-aware geometric contourlet semantic learning for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 1, pp. 698–712, Jan. 2025.
- [56] F. Tong and Y. Zhang, "Active learning-based spectral-spatial classification for discriminating tree species in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 9403–9414, 2024.
- [57] J. Bai et al., "Few-shot hyperspectral image classification based on adaptive subspaces and feature transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523917.
- [58] W. Yandong, Z. Kaipeng, L. Zhifeng, and Q. Yu, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 9911, pp. 499–515.
- [59] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," in *Proc. Comput. Vis. Pattern Recognit.*, 2021, arXiv:2112.05561.

- [60] F. Xu, G. Zhang, C. Song, H. Wang, and S. Mei, "Multiscale and cross-level attention learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501615.
- [61] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7482–7491.
- [62] Z. Xia, Z. Bing, Z. Lifu, and S. Yanli, "Hyperspectral remote sensing dataset for tea farm," Global Change Research Data Publishing & Repository, 2017.
- [63] 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, Feb. 2020.
- [64] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali, and M. S. Sarfraz, "A fast and compact 3-D CNN for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5502205.



Linfeng Liu was born in Henan, China. He received the M.S. degree in hydraulic engineering from China Three Gorges University, Yichang, China, in 2017, and the Ph.D. degree in hydrology and water resources from Zhengzhou University, Zhengzhou, China, in 2023. He is currently a Lecturer with Henan Polytechnic University, Jiaozuo, China, and is concurrently engaged in postdoctoral research. His primary research interests include land cover classification of hyperspectral imagery integrated with deep learning, as well as cloud computing in remote sensing.



Shuanggen Jin (Senior Member, IEEE) was born in Anhui, China, in September 1974. He received the B.Sc. degree in geodesy from Wuhan University, Wuhan, China, in 1999, and the Ph.D. degree in geodesy from University of Chinese Academy of Sciences, Beijing, China, in 2003. He is currently a Vice-President and Professor with Henan Polytechnic University, Jiaozuo, China, and also a Professor with the School of Artificial Intelligence, Anhui University, Hefei, China. His main research areas include satellite navigation, remote sensing, and

space/planetary exploration. He has authored or coauthored more than 500 papers in peer-reviewed journals and proceedings, 30 patents/software copyrights, and 15 books/monographs with more than 18 000 citations and H-index 73.

Prof. Jin has been President of International Association of Planetary Sciences (IAPS) (2015–2019), President of the International Association of Chinese Professionals in GNSS (CPGNSS) (2016–2017), Chair of IUGG Union Commission on Planetary Sciences (UCPS) (2015–2027), Vice-President of Institute of Data Science and Artificial Intelligence (2025–2028), Editor-in-Chief of *International Journal of Geosciences*, Associate Editor of *IEEE TRANSACTIONS ON GEOSCIENCE & REMOTE SENSING* and *Journal of Navigation*, Editorial Board Member of *GPS Solutions* and *Journal of Geodynamics*. He was recipient of 100-Talent Program of CAS, IUGG Fellow, IETI Fellow, IAG Fellow, AAIS Fellow, EMA Fellow, World Class Professor of Ministry of Education and Cultures, Indonesia, Chief Scientist of National Key R&D Program, China, Member of European Academy of Sciences, Member of Turkish Academy of Sciences, and Member of Academia Europaea.



Abhishek Banerjee received the Ph.D. degree from the East China Normal University, Shanghai, China. He is currently working as a Postdoctoral Research Fellow with Henan Polytechnic University, Jiaozuo, China. He worked as a Postdoctoral Fellow with Northwest Institute of Eco-environment and Resources, Chinese Academy of Sciences.

Dr. Banerjee is also the Steering Committee Member of International Geographical Union, Commission on Climatology from 2022 to recent. He has authored or coauthored more than 40 research articles in globally well-known journals, mostly on climate change, cryospheric dynamics, regional hydrology, and major environmental changes, including anthropogenic activities through remote sensing and GIS applications.



Chengcai Zhang was born in Henan in 1964. He received the M.S. degree in hydrology and water resources from Wuhan University, Wuhan, China, in 1992, and the Ph.D. degree in the same field from Wuhan University in 2000. Since 2002, he has been affiliated with Zhengzhou University, where he holds the position of Professor. His primary research interests focus on surveying and remote sensing technologies and their applications.