



Multipath feature fusion for hyperspectral image classification based on hybrid 3D/2D CNN and squeeze-excitation network

Ali Ari¹

Received: 23 November 2022 / Accepted: 28 December 2022 / Published online: 7 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Hyperspectral Images (HSI) are commonly used for classification thanks to their rich spectral feature information along with their spatial feature information. Convolutional Neural Network (CNN) based deep learning methods are commonly used in HSI classification (HSIC) applications to process the high nonlinearity and high dimensionality of HSI. This study proposes a method consisting of a combination of multipath Hybrid CNN and a Squeeze and Excitation (SE) network for HSIC. Features extracted with different kernel sizes in the multipath method are used together to extract richer feature information from HSI in this proposed method (PM). In the Hybrid CNN used in PM, 3D CNN was used to extract the spectral-spatial features. However, computational complexity increases with 3D CNN. Computational complexity is decreased with the use of Hybrid CNN. In addition, 2D CNN used in Hybrid CNN provides more spatial feature information to be extracted. However, in this study, 2D depthwise separable convolution (DSC) was used instead of 2D CNN. By using 2D DSC instead of standard 2D CNN, computational cost and the number of trainable parameters is significantly decreased. Finally, the PM is combined with the SE network to advance the HSIC accuracies. The SE network is designed to enhance the representation quality of CNN. WHU-Hi-HongHu (WHHH), WHU-Hi-HanChuan (WHHC), and WHU-Hi-LongKou (WHLK) datasets were used to evaluate the classification accuracies of the PM. Using a 5% training sample with WHLK, WHHC and WHHH, OA values of 99.86%, 97.51% and 97.64% were obtained. Furthermore, the PM was compared with the latest technology methods in the literature and outperformed all methods.

Keywords Hyperspectral image classification · Depthwise separable convolution · Squeeze and excitation network · Hybrid 3D/2D CNN

Introduction

Hyperspectral images (HSIs) are 3-dimensional image cubes that are often used in remote sensing, have hundreds of different spectral bands for each pixel, and consist of two spatial and one spectral dimension (Firat et al. 2022b). The rich information hidden in the spectral bands is associated with the physical appearance of each object. Therefore, using this spectral information is more beneficial in distinguishing objects. HSI has recently been used in many areas such as target and object detection, identification of physical

evidence such as blood stains, explosive residues, semen stains and stamp ink in criminal applications, vegetation analysis in precision agriculture applications, geological mapping, land registry and cadastral services, and underground mine detection. Analyzing, classifying and interpreting with learning-based methods has become a popular research area (Lanthier et al. 2008; Li et al. 2018; Firat and Hanbay 2022; Jia et al. 2020; Hörig et al. 2001; Firat et al. 2022a).

The literature shows that in the classification of HSI, different machine learning based methods such as Artificial Neural Networks (ANN), k-Nearest Neighbor (k-NN), Decision Trees (DT), Support Vector Machine (SVM) and Logistic Regression (LR), Maximum Likelihood are used (Ahmad et al. 2019; Melgani and Bruzzone 2004; Blanzieri and Melgani 2008; Ratle et al. 2010). Since there may not be enough training samples for a large number of spectral bands, many traditional spectral classifiers fail to detect some regions in the HSI due

Communicated by: H. Babaie

✉ Ali Ari
ali.ari@inonu.edu.tr

¹ Department of Computer Engineering, Faculty of Engineering, İnönü University, 44280 Malatya, Turkey

to the high correlation ratio between spectral bands and spectral redundancy. This causes an imbalance between the limited number of training samples and the high dimensionality of the spectral bands (Alcolea et al. 2020; Wang et al. 2020).

The high spectral resolutions of HSI can adversely affect the efficiency of data classification and increase the cost of processing the data. Therefore, dimension reduction (DR) is significant in HSI classification (HSIC). It is aimed to separate important spectral bands in HSIs by DR. As a result, both the data dimension and the transaction cost is decreased. DR can be performed in HSIC in two ways: spectral band selection (BS) and band extraction (BE) (Xu et al. 2019). In the spectral BS, analysis is performed on the most useful spectral bands selected from the original HSI spectral band. In spectral BE, the spectral size of HSI data with a large spectral size is decreased. However, the spatial dimension of HSI data with decreasing spectral size does not change. Thus, DR methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are frequently used in the literature in the analysis of HSI (Firat et al. 2022b; Kang et al. 2020).

Deep neural networks are adaptive and efficient learning architectures. Deep learning (DL) architectures offer dynamic and high-level image features and are frequently used in the classification of HSIs (Hanbay 2020; Üzen et al. 2022; Uzen et al. 2021). The training phase of the DL method automatically extracts the features and uses them in the classification step (Firat and Hanbay 2021). As a classical DL method, the stacked auto-encoder can extract spatial-spectral features. It can then combine this information for HSIC.

Firat H., et al. (Firat and Hanbay 2022), aimed to develop a new 3D Convolutional neural network-based method to extract features from spectral-spatial feature information simultaneously and benefit from this information to increase the classification results in HSIC. This method consists of 4 consecutive 3D convolution layers. PCA is implemented to the original HSI as a pre-processing step, and the most useful principal components are extracted. The proposed method was tested on datasets obtained from Indian Pines (IP) and Salinas (SA) agricultural fields. A classification result of 99.93% for IP dataset, 100% for SA agricultural area dataset, 99.99% for University of Pavia (PU) dataset and 99.81% for SA dataset was obtained. Hanbay D., et al. (Firat et al. 2022b), aimed to examine the effects of 7 different DR methods on classification performances as a preprocessing step in HSIC. Along with a hybrid method in which 2D and 3D CNN are used together, 7 different DR methods are used together as a preprocessing step. The proposed method was tested on 4 HSI datasets (IP, HyRank-Loukia (HL), Botswana (BO), PU). DR with PCA generally generated better classification results in all datasets. Hanbay (2020), developed a

new method for HSIC using complex two-dimensional Gabor transform and CNN architecture. For the construction of the feature extraction procedure, the Gabor features of different samples of the given class of CNN architecture were calculated simultaneously. The feature vectors obtained from the Gabor transform and CNN architecture of the same class data were combined and provided as input to the classifier. Comparative experimental studies were conducted on the dataset of PU and the IP dataset. It was seen that the method produces more successful results than the traditional CNN architecture. Chen Y., et al. (Data et al. 2015) used DL architectures for the first time for HSIC. They developed a method for spectral-spatial feature extraction and classification using a combination of PCA, stacked auto-encoder and logistic regression. A DL method, which is a combination of 3D/2D CNN method and 3D/2D Inception module, is suggested by Asker M. E., et al. (Firat et al. 2022b) to extract more features from HSI and acquire better classification results. With the proposed method, 99.91% of classification results were obtained in studies conducted in two agricultural areas (IP, SA) and 98.93% in urban areas (Houston (UH)). Hanbay D., et al. (Firat and Hanbay 2023), also aimed to examine the effect of a 3D CNN-based hybrid approach on classification accuracy, together with the LeNet5, AlexNet, VGG16, GoogleNet and ResNet50 DL architectures, which are among the successful examples of CNN for the HSIC problem. 3D CNN-based deep learning architectures have been developed by using layers in 5 different DL architectures together with 3D CNN. Applications performed on 4 different hyperspectral data sets (IP, SA, BO, HL) showed that better results were obtained with VGG16. Jia-dong H., et al. (Huang et al. 2022) aimed to limit the HSI of 25 kinds of commercial textile fibers. In the study, firstly, HSI pixel-based preprocessing was performed. Classification performances are presented using four conventional machine learning (ML) classification methods, Partial Least Squares Discriminatory Analysis (PLS-DA), Random Forest (RF), SVM, and KNN. It is seen that RF has the highest classification performance, reaching 91.40%. Then, a Backpropagation Neural Network (BPNN) method and a One-Dimensional CNN (1D-CNN) method were created and compared with conventional ML methods. The results indicate that the 1D-CNN models have 97.90% and 98.60% accuracy in training and testing sets, respectively. F1 score, specificity (Sp), sensitivity (Se), and precision (Pr) of the methods were calculated as 98.60%, 99.90%, 98.60%, and 98.70%, respectively. Xinzhi L., et al. (Liu et al. 2022) tried to detect CNN-based green pepper from HSI. In HSI, 3 different datasets were created with gradual improvements based on the amount of light. First, images of green peppers were tagged and given as input to CNN. Throughout all the experiments, the

optimization process in the training process was done with the Adam technique. As a post-process, a 5×5 Gaussian filter was applied to the segmentation results, and OTSU threshold method was used instead of a fixed threshold to obtain a binary classification. The highest segmentation accuracy was obtained at 93.45%. Mohan A., et al. (Mohan and Venkatesan 2020) suggested a hybrid CNN method consisting of a multi-scale spatial-spectral feature based on 2D CNN and 3D CNN to classify HSIs. To decrease spectral band excess and ensure optimum band extraction, a hybrid DR method consisting of linear Gaussian random projection and Non-Linear Kernel PCA was used. Studies were conducted on SA, PU, IP, and UH datasets to test the proposed method's classification accuracy. The overall accuracy results obtained were 99.80% for IP, 99.99% for PU, 100% for SA, and 99.12% for UH dataset, respectively. Tao L., et al. (Mughees and Tao 2017) proposed a combined stacked auto encoder and LR method to extract deep spectral features. Deep belief network-based classification is one of the other DL methods used for HSIC. This method is proposed to join the spatial features of HSI with the spectral features. As a result of the applications using PU and IP datasets and 10% training sample resulted in 98.64% and 89.35% overall accuracy results, respectively. Chen et al. (2015) proposed a method in which hierarchical learning-based feature extraction, PCA, and LR were used together for deep belief network-based HSIC. IP and PU were used to analyze the classification accuracy of the proposed method. The overall accuracy results obtained 95.95% for IP and 99.05% for PU. Yang et al. (2018), investigated deep learning techniques to solve the HSI classification problem. Four DL models such as 2D-CNN, 3D-CNN, R-2D-CNN and R-3D-CNN were designed and developed and traditional ML methods were compared. The proposed methods have been observed to give more than 30% performance. Mou et al. (2017) proposed an RNN model for HSIC, inspired by their observation that hyperspectral pixels can be considered as ordered data. PRE tanh has been proposed as a newly designed activation function for hyperspectral data processing in RNN. They are the first to consider a hyperspectral pixel's intrinsic ordered data structure, which represents a new methodology for better understanding, modeling and processing hyperspectral data. Zhang et. al. (2018) proposed a new diverse region-based CNN model for HSIC. The proposed DR-CNN model has extracted spectral-spatial features through a well-designed network with a "multi-scale aggregation" module. Research shows that the proposed DR-CNN can provide statistically higher accuracy than the latest techniques. Paoletti et al. (2018) proposed a new 3D CNN architecture that uses both spatial and spectral information to classify HSIs. Studies have shown that the proposed network works correctly and efficiently, reduces

computation time compared to other traditional CNN techniques and increases the accuracy in classifying HSIs. Li et al. (2017) proposed a new pixel-pair method to demonstrate the advantage of CNN. Experimental results based on HSI datasets showed that the method can provide better classification performance than the traditional DL based method. Zhu et al. (2018), designed a CNN in Generative Contention Network (GAN) to distinguish between input and output. The proposed system is implemented on three widely used HSI datasets: SA, IP and Kennedy Space Center (KSC). As a result of the studies, it has been revealed that the proposed models give competitive results compared to the most advanced methods. Cheng et al. (2018) proposed a new method of extracting hierarchical deep spatial features for HSIC by exploring the power of ready-to-use CNN models without additional retraining on the target dataset. The study aimed to learn stronger spectral-spatial features by designing a new objective function that inserts a metric learning term into the SVM to obtain better classification accuracy. For these purposes, a method based on metric learning is proposed. The proposed method was evaluated on three commonly used HSI benchmarks and the most advanced results were obtained. In Okwuashi and Ndehedehe (2020), the Deep Support Vector Machine (DSVM) with four different kernel functions was used for HSIC. The DSVM algorithm is compared with well-known SVM, Deep Neural Network (DNN), Gaussian Mixture Model (GMM), KNN and K Means (KM) classifiers. The results of the study suggest that DSVM outperforms other algorithms. Zhou et al. (2019) proposed an HSIC method based on Long Short Term Memory (LSTM) network for both spectral feature extraction and spatial feature extraction. Experiments were conducted on three different HSIs and the proposed method was compared with the most advanced methods, including CNN. As a result of the experimental results, it has been observed that the simultaneous use of spectral and spatial information improves the classification performance. Song et al. (2018) proposed a new DL-based method for HSIC with Deep Feature Fusion Network (DFFN). The experiment results, conducted with three real HSI datasets, showed that this method gave better results than several known deep learning methods in terms of both visual qualities in the classification map and quantitative metrics. Hong et al. (2022) reused HSIC with transformers from an sequential perspective and proposed a new method called SpectralFormer. Beyond bandwise representations in classic transformers, SpectralFormer is capable of learning spectrally local sequence information from neighboring bands of HSIs, yielding groupwise spectral embeddings. Ding et al. (2022) proposed a new multcale fusion network for HSI classification by integrating multiscale graph convolutional network and multiscale CNN, which

can extract multiscale superpixel-based graphics features and local pixel features. Gao et al. (2021) proposed a sandwich CNN based on spectral feature enhancement (SFE-SCNN) for HSIC. The SFE-SCNN, presents the SFE operation, which makes the data reflect more discriminative spectral feature details to suppress the interference of mixed pixels.

The current study proposes a method consisting of a combination of 3D CNN, SE network and 2D depthwise separable convolution layers for HSIC. One of the most critical problems in HSIC is considering both spectral and spatial features. When the latest literature is examined, it is seen that CNN-based methods are used quite frequently for HSIC. In studies using 2D CNN, only spatial features are extracted. Since distinctive features are not extracted from the spectral dimension, important spectral features are lost. Unlike 2D CNN, since spatial-spectral features are extracted simultaneously in 3D CNN using 3D convolutions, it started to be used in HSIC. However, the use of 3D CNN causes an increase in computational complexity. To overcome the problems that arise when 2D CNN and 3D CNN are used separately, the focus is on the hybrid CNN method, which is used together. The hybrid CNN method is aimed to make full use of both spatial and spectral feature information. This increases the classification accuracy. In addition, combining 2D CNN and 3D CNN methods with different kernel sizes allows multiple features to be used together. Moreover, a multipath network structure is used in the proposed method. The use of a multipath network structure provides further enrichment of the features extracted from the HSIs. In our multipath hybrid method, 2D depthwise separable convolution layers are used instead of 2D convolution layers to decrease the number of trainable parameters and increase classification accuracy in 2D CNN. Finally, the SE network is combined with Hybrid CNN to improve its feature extraction and HSIC performance. The SE network is designed to develop the representation quality of a CNN. The main contributions of this study are as follows:

1. More feature extraction from HSIs is provided by using a multipath network structure.
2. By using 2D DSC instead of 2D CNN, the number of trainable parameters is reduced and classification performance is increased.
3. Hybrid CNN method was used to reduce the number of trainable parameters and increase classification performance.
4. Squeeze-and-excitation blocks are integrated into the proposed method as they increase the classification performance while minimally increasing the total number of parameters.

5. WHLK, WHHC and WHHH datasets were used in the applications to evaluate the classification performance of the proposed method. Using a 5% training sample with WHLK, WHHC and WHHH, the overall accuracy of 99.86%, 97.51%, 97.64%.

The rest of this study is organized as follows; 3D/2D CNN, in-depth separable convolution and SE block, which form the theoretical background of the proposed method, are explained in Chapter 2. Detailed information about the datasets used in the applications is explained in Chapter 3, and the proposed hybrid method is explained in Chapter 4. The experimental process and results are given in Chapter 5, and a general summary of the study is given in Chapter 6.

Materials and Method

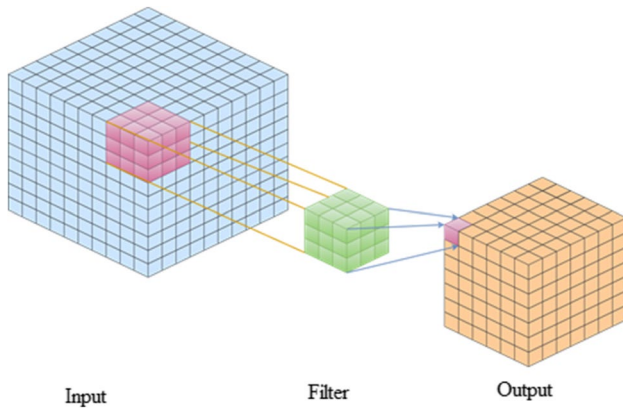
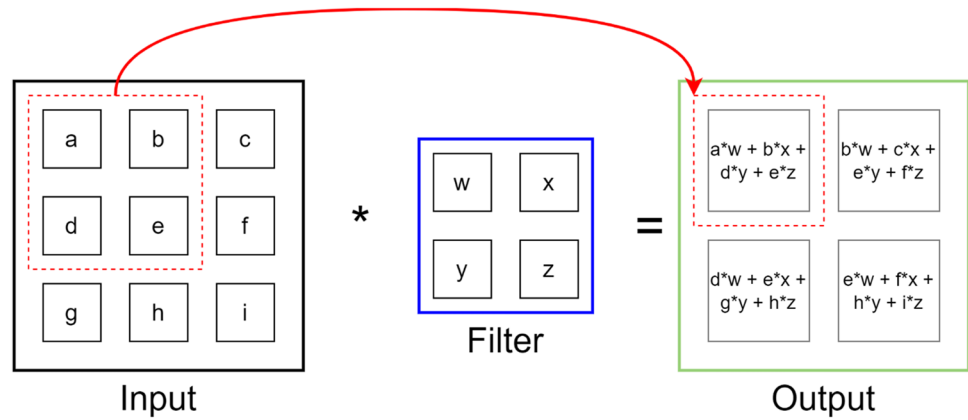
3D/2D CNN

CNN, which is successfully used in analysing images, is also used in solving classification problems of HSI (Karadag et al. 2021). Traditional CNN models use 2D CNN for hyperspectral images, but feature maps cannot be obtained with 2D CNN (Firat et al. 2022a). 3D CNNs are used to solve this problem. Since 3D CNNs have more parameters than 2D, they cause costly computation. For this reason, 3D CNN is used in 3-dimensional feature extraction or when it is desired to establish a relationship between 3 dimensions. 2D CNN usually consists of convolution, pooling, and fully connected layers. 3D CNN model consists of convolution, smoothing and fully connected layers (Firat and Hanbay 2022). In 2D CNN, a new 2×2 output is obtained after the 3×3 input images are multiplied with a 2×2 filter in the convolution operation (Fig. 1). In 3D CNN, 3D cube is used to log the network (Figs. 2 and 3).

Depthwise Separable Convolution

Depthwise separable convolutions (DSC) are divided into two categories called depthwise convolution and pointwise convolution (Lu et al. 2022). In depthwise convolution is a separate convolution operation on each channel of the input image. The convolution operation is used to extract spatial features in all dimensions. Pointwise convolution is a 1×1 standard convolution operation on the output feature map (Dang et al. 2020).

The depthwise convolution kernel receives data from only one channel at any given time and accumulates the outputs from all channels in a matrix of size $c \times n \times n$; The size of the kernel convolution network is $n \times n$, and the number of channels is c . In particular, the feature map is taken as input and applied to the expansion and depth of the 2D filter (for $n \times n$

Fig. 1 2D convolution sample**Fig. 2** 3D Convolution

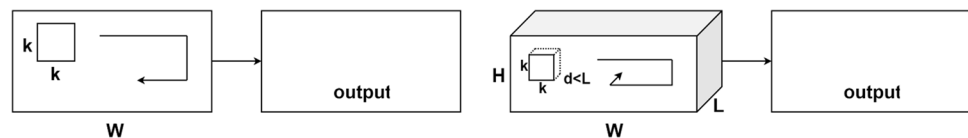
kernel). Depthwise convolution kernel is applied iteratively to all channels (Lu et al. 2022).

The output of the depthwise convolution kernel is given to the point convolution kernel, which iteratively applies the 1×1 filter at all points. The depth of this kernel is equal to the number of input channels. Compared to standard depthwise convolution, In-depth decomposable convolution reduces the number of input transformations required (Lu et al. 2022).

The mathematical formulas are as follows:

$$\text{Conv}(W, y)_{(i,j)} = \sum_{k,l,m}^{K,L,M} W_{(k,l,m)} \cdot y_{(i+k,j+l,m)} \quad (1)$$

$$\text{PointwiseConv}(W, y)_{(i,j)} = \sum_m^M W_m \cdot y_{(i,j,m)} \quad (2)$$

Fig. 3 Comparison of 2D CNN and 3D CNN (Lee et al. 2018)

$$\text{DepthwiseConv}(W, y)_{(i,j)} = \sum_{k,l}^{K,L} W_{k,l} \odot y_{(i+k,j+l)} \quad (3)$$

Squeeze-Excitation (SE) Block

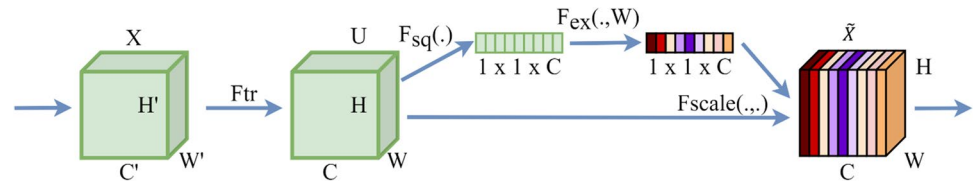
The SE block improves network performance by simulating interdependencies between channels and adaptively recalibrating the characteristic response (Lin et al. 2019). Because of the performance advantage of SE block, it can decrease model parameters and increase detection speed. A learning method with SE block can characterize the spatial- spectral features quite well. The application of the SE block is indicated in the Eqs. (4) and (5).

$$U = F_{tr}(X), \quad X \in R^{H' \times W' \times C'}, \quad U \in R^{H \times W \times C} \quad (4)$$

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * X^s, \quad (5)$$

This equations also shows the X input example. F_{tr} is defined as the convolution operator. $V = [v_1, v_2, \dots, v_c]$ represents the set of learned filter kernels and v_c , specifies the parameters of the c 'th filter. $U = [u_1, u_2, \dots, u_c]$ represents the output. The s 'th input data is X^s . U represents the properties acquired after convolution. Generally, an SE block consists of two essential steps, squeeze and excitation (Rajendran et al. 2022). The structure of the SE block is shown in Fig. 4.

Fig. 4 SE Block



Squeeze Process

The squeeze operation is performed through overall mean pooling, which obtains dependencies between channels (Han et al. 2020). By converting the two-dimensional feature channel to a real number, feature squeeze is performed across the spatial dimension. This real number has the global receiver area, and the output size matches the number of input-capable channels. As stated in Eq. 6, the general average pool, which is the simplest addition technique, is preferred. Equation 6 converts the input of $H \times W \times C$ to the output data of $1 \times 1 \times C$ and is shown in Fig. 4 as the operation F_{sq} . The result is equivalent to showing the numerical distribution of the layer's C feature maps.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

Excitation Process

The excitation process creates weights for each feature channel with parameters. The parameters are used to model the correlation between feature channels. A second process is applied to use the information gathered in the squeeze process, which aims to capture the dependencies on a per-channel basis (Gong et al. 2021). To achieve this goal, the function must meet two criteria: first, it should be flexible, i.e. it should be able to learn a non-linear interaction between channels, and secondly, it should learn a non-exclusive relationship as it is desirable to ensure that more than one channel is allowed to be highlighted. A simple gate mechanism with sigmoid activation is used to meet these criteria.

$$z_c = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (7)$$

Equation 7 demonstrates σ sigmoid and δ ReLU functions as ($W_1 \in R^{(C/r) \times C}$ ve $W_2 \in R^{C \times (C/r)}$). The final outcome of the block is acquired by rescaling the U transform output with activations:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (8)$$

In Eq. 8, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_c]$ ve $F_{scale}(u_c, s_c)$, refers to the ductal collision between scaler s_c and feature map $u_c \in R^{H \times W}$.

Proposed Method

This study's HSIC method is based on a feature extraction method by combining a SE block with a multipath Hybrid CNN. Hybrid CNN consists of multipath 3D CNN and 2D DSC layers (Fig. 5). DL is challenging to implement to HSI because the data structure of HSI is complex. In general, the neural network has strong representation capability and a larger volume of training samples. The first aim of this study is to provide simultaneous extraction of spatial-spectral features for HSIC. With 2D CNN, only spatial features are extracted. However, since HSIs are 3D, spectral features are significant. The 3D CNN method was used to extract spectral features. SE Network is joined with CNNs in this study to improve its performance in feature extraction and classification of HSI. The SE network is designed to develop the representation quality of a CNN. DSC layers are used to decrease the computational cost and the number of trainable parameters. Finally, using a multipath method, richer feature information from HSI is obtained by using a combination of different kernel sizes and extracted features in the PM. It is a hybrid method because when only 2D CNN is used, spectral feature information is lost. Computational complexity increases when only 3D CNN is used. To overcome above mentioned challenges, a hybrid method has been proposed.

The HSI is specified as a 3D cube of size $M \times N \times D$. M and N refer to the spatial width and height of the image. D indicates the number of spectral bands. The ground truth of the input image Y is converted using one-hot encoding and displayed as $Y = (y_1, y_2, \dots, y_C)$. Here C denotes the number of classes present in the input image. HSI pixels exhibit high inter-class similarity, high intra-class variability, overlapping and nested regions, which require extensive effort for any classification model. To overcome these problems, it is necessary to remove spectral redundancy. Conventional PCA across the spectral bands is first applied to the original HSI to remove spectral redundancy. PCA reduces the number of spectral bands from D to B while maintaining the same spatial dimensions (i.e. width M and height N). Only their spectral bands were decreased, protecting the spatial information crucial to recognizing any object. Thus, it was decreased to the desired number of bands. The hyperspectral data cube was divided into small overlapping 3D spatial segments from which real labels based on the central

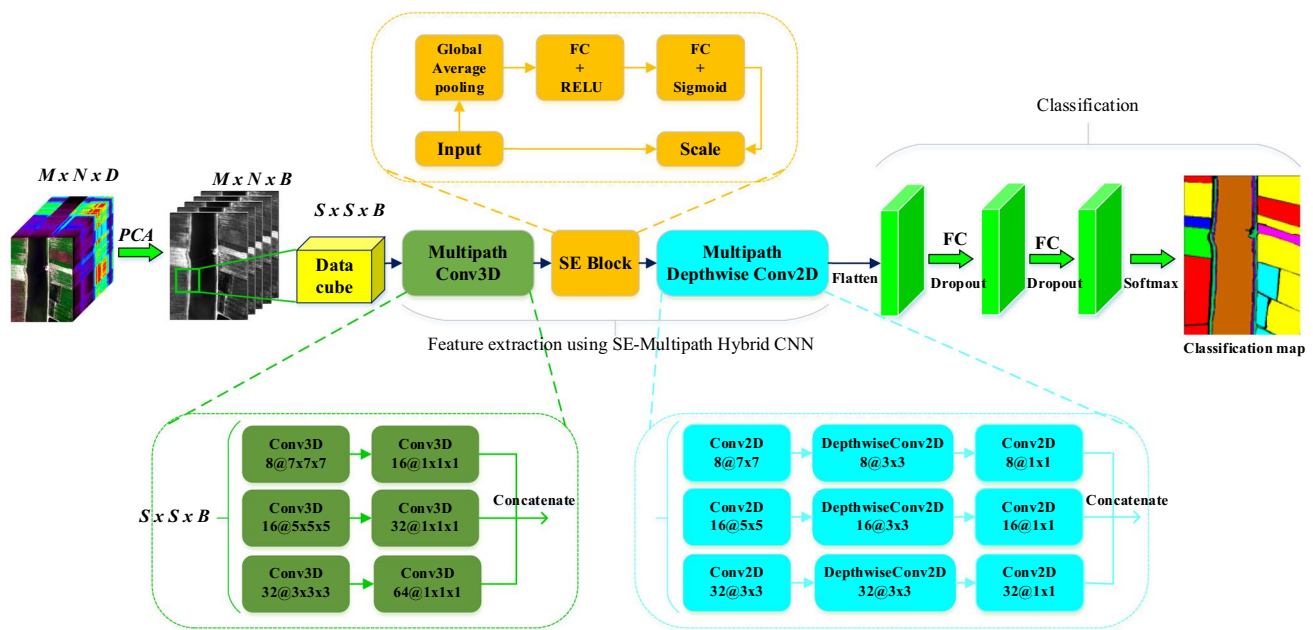


Fig. 5 Proposed method

pixel were generated to use image classification techniques. The 3D neighbor fragments $S \times S \times B$ were created from the hyperspectral data cube centred at the spatial location (a,b) and covering the $S \times S$ window or the spatial dimension and all B spectral bands. The total number of 3D parts (n) produced from the hyperspectral data cube is found by (n), $(M - S + 1) \times (N - S + 1)$.

The obtained 3D parts are first given to the input of the multipath 3D CNN block. This block consists of 3 different paths to extract more spatial spectral features. Each path consists of 3D convolution operations. The first 3D convolution layers in all three paths have different filter and kernel sizes. The dimensions of the 3D convolution layers are as follows: In the first path, the core size of the 3D convolution is $7 \times 7 \times 7$, and the number of filters is 8. In the second path, the core size of the 3D convolution is $5 \times 5 \times 5$, and the number of filters is 16. In the third path, the core size of the 3D convolution is $3 \times 3 \times 3$, and the number of filters is 32. The kernel dimensions of all the second 3D convolution layers in all three paths are $1 \times 1 \times 1$. However, the number of filters in each path is different. The numbers of filters on the paths are 16, 32, and 64, respectively. Thus, a different neural network method was designed for each path. Using more than one path, different features are extracted to obtain a more distinctive feature representation. These extracted features were combined afterwards. Feature cubes, removed from all three paths when combined, have the same width and height but different depths. After merging, the reshape process is done in 2D image size. The obtained 2D image

is passed through the SE block and given to the input of the block where multipath separable convolution layers would be applied. In this block, 2D convolution and 2D DSC operations are used to learn more spatial properties. The first 2D convolution layers in this block in all three paths have different filter and kernel sizes. The dimensions of the 2D convolution layers are as follows: In the first path, the kernel size of the 2D convolution is 7×7 , and the number of filters is 8. In the second path, the core size of the 2D convolution is 5×5 , and the number of filters is 16. In the third path, the kernel size of the 2D convolution is 3×3 , and the number of filters is 32.

Then, in the first path, depthwise convolution with 8 filters and 3×3 kernel size and pointwise convolution with 8 filters and 1×1 kernel size was applied. Similarly, in the second path, depthwise convolution with 16 filters and 3×3 kernel size, and pointwise convolution with 16 filters and 1×1 kernel size were applied. Finally, depthwise convolution with 32 filters and 3×3 kernel size and pointwise convolution with 1×1 kernel size with 32 filters were applied in the third path. After the most recent processing, the features extracted are flattened and given as input to the fully connected (FC) layers for HSIC. The PM uses two FC layers with 128 and 256 neurons. A dropout layer with a dropout rate of 0.4% is implemented after each FC layer to prevent overfitting. The output of the FC layer is given to the simple softmax classifier to generate the classification result. Table 1 includes detailed information about the PM. The number of trainable parameters in the PM for WHHH is 3.453.650.

Table 1 WHHH summary of dataset for the proposed method

Layer Name	Layer details	Parameters	Output	Connected to
InputLayer	-	0	$7 \times 7 \times 20 \times 1$	-
Conv3D	filters = 8, kernel_size = $7 \times 7 \times 7$, padding = 'same'	2752	$7 \times 7 \times 20 \times 8$	InputLayer
Conv3D_1	filters = 16, kernel_size = $5 \times 5 \times 5$, padding = 'same'	2016	$7 \times 7 \times 20 \times 16$	InputLayer
Conv3D_2	filters = 32, kernel_size = $3 \times 3 \times 3$, padding = 'same'	896	$7 \times 7 \times 20 \times 32$	InputLayer
Conv3D_3	filters = 16, kernel_size = $1 \times 1 \times 1$, padding = 'same'	144	$7 \times 7 \times 20 \times 16$	Conv3D
Conv3D_4	filters = 32, kernel_size = $1 \times 1 \times 1$, padding = 'same'	544	$7 \times 7 \times 20 \times 32$	Conv3D_1
Conv3D_5	filters = 64, kernel_size = $1 \times 1 \times 1$, padding = 'same'	2112	$7 \times 7 \times 20 \times 64$	Conv3D_2
Concatenate			$7 \times 7 \times 20 \times 112$	Conv3D_3, Conv3D_4, Conv3D_5
Reshape	-	0	$7 \times 7 \times 2240$	Concatenate
GlobalAveragePooling2D	-	0	2240	Reshape
Dense (FC)		313,740	140	GlobalAveragePooling2D
Dense_1 (FC)		315,840	2240	Dense (FC)
tf.reshape (TFOpLambda)	-	0	$1 \times 1 \times 2240$	Dense_1 (FC)
multiply (Multiply)	-	0	$7 \times 7 \times 2240$	Reshape, tf.reshape
Conv2D	filters = 8, kernel_size = 7×7 , padding = 'same'	878,088	$7 \times 7 \times 8$	multiply
Conv2D_2	filters = 16, kernel_size = 5×5 , padding = 'same'	896,016	$7 \times 7 \times 16$	multiply
Conv2D_4	filters = 32, kernel_size = 3×3 , padding = 'same'	645,152	$7 \times 7 \times 32$	multiply
Depthwise_conv2D	kernel_size = 3×3 , depth_multiplier = 1	80	$5 \times 5 \times 8$	Conv2D
Depthwise_conv2D_1	kernel_size = 3×3 , kernel_size = 1×1	160	$5 \times 5 \times 16$	Conv2D_2
Depthwise_conv2D_2	kernel_size = 3×3 , depth_multiplier = 1	320	$5 \times 5 \times 32$	Conv2D_4
Conv2D_1	filters = 8, kernel_size = 1×1 , padding = 'same'	72	$5 \times 5 \times 8$	Depthwise_conv2D
Conv2D_3	filters = 16, kernel_size = 1×1 , padding = 'same'	272	$5 \times 5 \times 16$	Depthwise_conv2D_1
Conv2D_5	filters = 32, kernel_size = 1×1 , padding = 'same'	1056	$5 \times 5 \times 32$	Depthwise_conv2D_2
Concatenate_1	-	0	$5 \times 5 \times 56$	Conv2D_1, Conv2D_3, Conv2D_5
Flatten	-	0	1400	Concatenate_1
Dense_2(FC)	units = 256	358,656	256	Flatten
Dropout	dropout-ratio 0.4	0	256	Dense_2(FC)
Dense_3(FC)	units = 128	32,896	128	Dropout
Dropout_1	dropout-ratio 0.4	0	128	Dense_3(FC)
output_layer	Output_units = 22	2838	22	Dropout_1
Total number of trainable parameters		3,453,650		

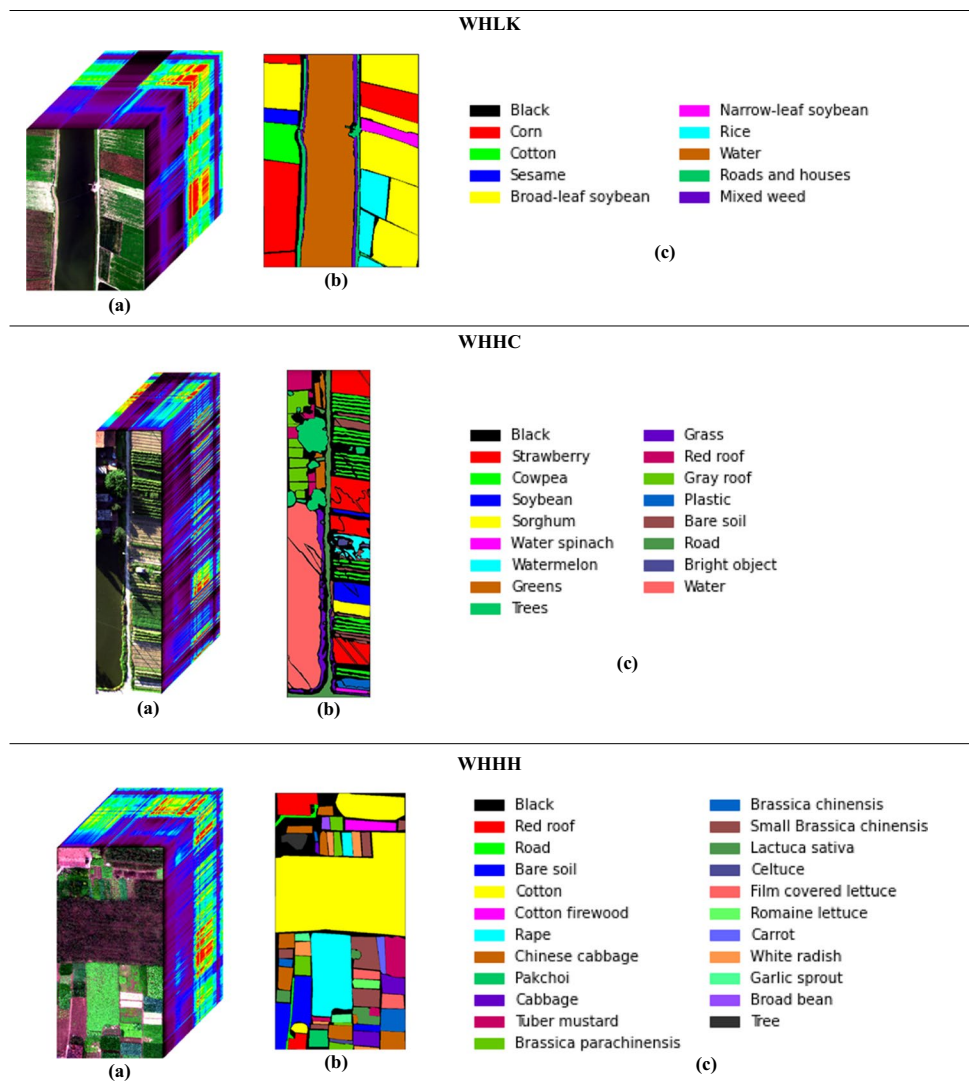
Experimental results & analysis

Dataset

In this study, images from WHLK, WHHC and WHHH datasets in WHU-Hi dataset (Zhong et al. 2020) were used, as shown in Fig. 6. The number of classes belonging to the 3 datasets used and the number of images belonging to each class is given in Table 2. All datasets were obtained 2 in agricultural fields with various crop types in Hubei province, China, via a Headwall Nano-Hyperspec sensor mounted on

a UAV platform. UAV hyperspectral data preprocessing includes geometric correction and radiometric calibration performed in HyperSpec software ensured by the instrument manufacturer. For radiometric calibration, raw digital number values were converted to radiance values with the laboratory calibration parameters of the sensor. The WHLK dataset was created in Longkou Town, Hubei province, China. The dimension of each image in the dataset is 550×400 pixels and includes 270 bands between 400 and 1000 nm, and the spatial resolution of the HSIs is about 0.463 m. The WHHC dataset was created in Hanchuan, Hubei province, China.

Fig. 6 (a) False-color images, (b) ground truth maps and (c) color map for WHLK, WHHC and WHHH datasets



The dimension of each image in the dataset is 1217×303 pixels, includes 274 bands between 400 and 1000 nm, and the spatial resolution of the HSIs is about 0.109 m. The WHHH dataset was created in Honghu City, Hubei province, China. The dimension of each image in the dataset is 940×475 pixels, includes 270 bands between 400 and 1000 nm, and the spatial resolution of the HSIs is about 0.043 m.

Applications created by using three datasets are written in the python programming language. Google Colab is used in applications. In addition, Tensor Processing Units were used as hardware accelerators. Tensor Processing Units provide 107.77 GB of storage and 35 GB of RAM to run python codes in Colab. The training-test ratio for all datasets is 5–95%. Also, the learning rate was 0.001 and Adam was used as the optimizer. The training process was carried out in 256 batch sizes and 100 epochs. The width and height of the 3D data cube obtained in all datasets

were taken as 7×7 ($S = 7$). After PCA is applied to the hyperspectral image (HSI) of size $M \times N \times D$, the number of spectral bands decreases, and the newly created image becomes $M \times N \times B$. Within the scope of the implementations, $B = 20$ was taken for WHLK, WHHC and WHHH datasets. The 3D part dimensions given as input image data to all deep learning-based methods used for comparison are $7 \times 7 \times 20$ ($S \times S \times B$) for all three datasets. Average accuracy (AA), Overall accuracy (OA) and Kappa (K) coefficient evaluation criteria were used to analyze the classification accuracy of the PM on three datasets. AA is calculated based on class classification accuracy. OA is determined by the ratio of the number of correctly classified samples to the number of all test samples. K is a statistical metric that can evaluate the consistency between the classification map obtained as a result of the estimation and the ground truth map. The PM is compared with 8 different deep learning-based methods studied in

Table 2 Detailed information about datasets

No	WHHH		WHHC		WHHH	
	Classes	Samples	Classes	Samples	Classes	Samples
1	Red roof	14,041	Strawberry	44,735	Corn	34,511
2	Road	3512	Cowpea	22,753	Cotton	8374
3	Bare soil	21,821	Soybean	10,287	Sesame	3031
4	Cotton	163,285	Sorghum	5353	Broad-leaf soybean	63,212
5	Cotton firewood	6218	Water spinach	1200	Narrow-leaf soybean	4151
6	Rape	44,557	Watermelon	4533	Rice	11,854
7	Chinese cabbage	24,103	Greens	5903	Water	67,056
8	Pakchoi	4054	Trees	17,978	Roads and houses	7124
9	Cabbage	10,819	Grass	9469	Mixed weed	5229
10	Tuber mustard	12,394	Red roof	10,516		
11	Brassica parachinensis	11,015	Gray roof	16,911		
12	Brassica chinensis	8954	Plastic	3679		
13	Small Brassica chinensis	22,507	Bare soil	9116		
14	Lactuca sativa	7356	Road	18,560		
15	Celtuce	1002	Bright object	1136		
16	Film covered lettuce	7262	Water	75,401		
17	Romaine lettuce	3010				
18	Carrot	3217				
19	White radish	8712				
20	Garlic sprout	3486				
21	Broad beans	1328				
22	Tree	4040				
Total Number		386,693		257,530		204,542

Table 3 Classification accuracies for the WHLK (%)

No	Train/Test	2D CNN	3D CNN	FC3D CNN	HybridSN	S3EResBoF	A2S2KRes	FuSENet	DLEM	PM
1	1725/32786	99.95	99.89	99.90	99.99	99.97	99.31	99.83	99.99	99.96
2	419/7955	98.89	99.20	99.26	99.87	99.01	99.30	99.42	99.89	99.90
3	152/2879	97.48	99.83	99.76	98.98	100	99.13	99.82	99.42	100
4	3161/60051	99.67	99.71	99.62	99.46	99.88	99.81	99.31	99.85	99.94
5	207/3944	97.07	98.46	97.76	97.59	99.46	99.84	99.54	98.34	98.86
6	593/11261	99.57	99.86	99.92	99.88	99.99	99.63	99.97	99.87	100
7	3353/63703	99.98	99.97	99.98	99.96	99.88	99.98	99.99	99.98	99.97
8	356/6768	95.45	98.16	98.41	98.16	92.85	98.26	95.79	98.36	99.02
9	261/4968	94.75	94.08	97.06	98.13	99.55	97.64	98.91	97.08	98.05
OA	10,227	99.42	99.59	99.64	99.63	99.59	99.63	99.54	99.76	99.86
K	/	99.24	99.46	99.64	99.52	99.46	99.51	99.40	99.68	99.81
AA	194,315	98.09	98.79	99.07	99.11	98.96	99.21	99.18	99.20	99.52
Training Time (min.)		1.63	2.7	4.1	6.33	10.54	29.38	13.8	19	2.66
Testing Time (sec.)		10.3	19	26	29	36	47.7	37	20.1	15

recent years, such as 3D CNN (Ben Hamida et al. 2018), 2D CNN (Makantasis et al. 2015), FuSENet (Roy et al. 2020b), A2S2KRes (Roy et al. 2020c), S3EResBoF (Roy et al. 2020a), FC3DCNN (Ahmad et al. 2020), HybridSN (Roy et al. 2019), and DLEM (Iyer et al. 2021).

The OA, AA, K values and class-based classification accuracies of all methods using WHLK are shown in Table 3. The estimated classification maps and ground truth map for all methods are given in Fig. 7. When Table 3 is examined, it is observed that the highest classification

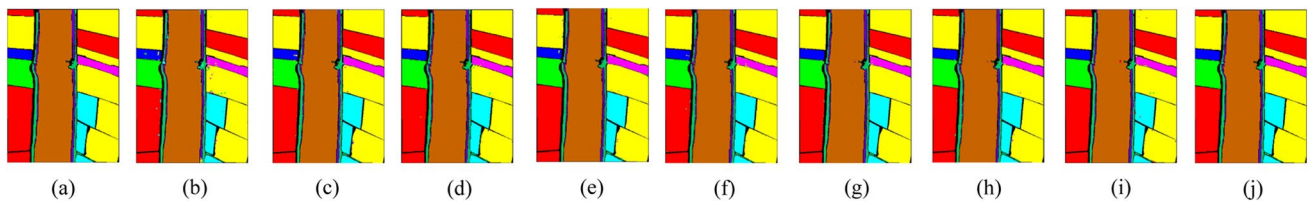


Fig. 7 Classification maps for WHLK. (a) Ground truth, (b)–(j) 2D CNN, 3D CNN, FC3DCNN, HybridSN, S3EResBoF, A2S2KRes, FuSENet, DLEM, and PM

performance were acquired in the PM with 99.86% OA, 99.81% K and 99.52% AA as a result of the applications performed with 5% training sample. When the OA, K and AA classification criteria were examined, the classification accuracies closest to the PM were acquired in DLEM with 99.76%, 99.68%, 99.20%, and A2S2KRes with 99.63%, 99.51% and 99.21%, respectively. Classification accuracies (OA, K, and AA) obtained by other methods are as follows: 99.54%, 99.40%, 99.18% with FuSENet, 99.63%, 99.52%, 99.11% with HybridSN, 99.64%, 99.64%, 99.07% with FC3DCNN, S3EResBoF with 99.59%, 99.46%, 98.96%, with 3D CNN 99.59%, 99.46%, 98.79%, with 2D CNN 99.42%, 99.24%, 98.09%. The lowest AA result was obtained with 2D CNN. With the PM, classification results of 99.90%, 100%, 99.94%, 100% and 99.02% were obtained in 2, 3, 4, 6 and 8 classes, respectively, according to their classification accuracies on a class basis. The PM, among these five classes, gives the best results. The best classification accuracy was found in the DLEM method, with 99.99% for class 1. The best classification accuracy from the other classes was found in A2S2KRes at 99.84% for class 5, FuSENet at 99.99% for class 7, and S3EResBoF at 99.55% for class 9. Except for classes 2, 3, 4, 6 and 8, the PM achieved a classification accuracy close to the methods that gave the best classification accuracies in all the other classes. The PM in the WHLK dataset was compared with eight methods in terms of training–testing times, and the results are given in Table 3. According to Table 3, the training–testing processes of the PM were carried out in a shorter time than the other seven methods except 2D CNN. Considering all the information in Table 3, it is concluded that the PM achieves better performance in a shorter time.

The OA, AA, K values and class-based classification accuracies of all methods using WHHC are shown in Table 4. The classification maps and ground truth map in the estimation result of all methods are given in Fig. 8. When Table 4 is examined, it is observed that the highest classification performance were acquired in the PM with 97.51% OA, 97.09% K and 94.10% AA as a result of the applications performed with 5% training sample. When the OA, K and AA classification criteria were examined, the classification accuracies closest to the PM were acquired

from A2S2KRes with 96.87%, 96.34%, 92.97% and DLEM with 96.47%, 95.87% and 92.47%, respectively. Classification accuracies (OA, K and AA) obtained by other methods are as follows: 94.27%, 93.28%, 92.06% with S3EResBoF, 94.05%, 93.05%, 89.24% with FuSENet, 95.09%, 94.25%, 89.38%, 3D with FC3DCNN 94.28%, 93.33%, 87.82% with CNN, 94.41%, 93.45%, 86.94% with HybridSN, 93.40%, 92.28%, 85.30% with 2D CNN. Considering all methods, it is observed that the lowest classification accuracies are acquired with 2D CNN. When the classification accuracies on a class basis are examined, with the PM, 99.05%, 99.68%, 83.95%, 94.51%, 96.23%, 95.83%, 99.31% and 89.87% classification in grades 1, 4, 6, 7, 8, 9, 10, and 12, respectively. Results have been obtained. The best classification accuracy was found in the FuSENet method with 99.09%, 98.74% and 97.43% for classes 3, 14 and 15, respectively. The best classification accuracy was found in FC3DCNN method with 98.85% for class 11, S3EResBoF method with 99.95% for class 16, A2S2KRes method with 97.44% and 86.51% for class 2 and 13, and DLEM method with 90.55% for class 5. In addition, the training–testing times for all methods are given in Table 4. According to Table 4, the PM has shorter training–test times than all other advanced methods except 2D CNN. Considering all the information in Table 4, it is concluded that the PM achieves better performance in a shorter time.

The OA, AA, K values and class-based classification accuracies of all methods using WHHH are given in Table 5. The classification maps and ground truth map of the estimation results of all methods are shown in Fig. 9. According to Table 5, it is observed that the highest classification accuracies were acquired in the PM with 97.64% OA, 97.02% K and 93.89% AA as a result of the applications using 5% training sample. The PM gives better results than A2S2KRes, which is one of the methods used for comparison, with 0.54% OA, 0.68% K and 0.43% AA. Similarly, 2.13%, 2.71% and 3.1% compared to FuSENet, 1.33%, 1.69% and 3.73% compared to DLEM 4.1%, 5.12% and 8.13% compared to S3EResBoF, 5.49% compared to HybridSN 6.97, 11.43%, compared to FC3DCNN 2.26%, 1.64%, 5.81%, 3.52% compared to 3D CNN, 4.46%, 6.29%, 3.83%, 4.85%, 9.19% compared to 2D CNN. it appears to do. When all methods are examined,

Table 4 Classification accuracies for the WHHC (%)

No	Train/Test	2D CNN	3D CNN	FC3D CNN	HybridSN	S3EResBoF	A2S2KRes	FuSENet	DLEM	PM
1	2237/42498	96.61	96.80	98.46	97.19	91.61	95.26	97.49	98.41	99.05
2	1137/21616	93.65	93.14	95.73	94.38	95.07	97.44	97.16	96.69	97.27
3	514/9773	86.85	91.03	89.49	89.96	92.12	94.71	99.09	94.38	97.58
4	268/5085	98.21	97.82	98.46	98.28	98.14	97.67	99.61	98.77	99.68
5	60/1140	69.59	67.18	74.74	61.43	96.94	89.86	87.30	90.55	89.03
6	227/4306	68.41	67.36	69.82	70.00	79.46	78.51	57.36	75.96	83.95
7	295/5608	88.02	90.24	87.53	87.63	93.61	91.80	93.58	93.24	94.51
8	899/17079	87.54	90.41	94.72	91.68	94.11	92.96	92.87	95.30	96.23
9	473/8996	89.13	89.01	90.58	92.75	92.31	92.99	91.03	94.28	95.83
10	526/9990	99.01	97.80	98.13	98.99	98.33	98.84	96.78	98.67	99.31
11	845/16066	96.74	96.96	98.85	98.25	92.37	96.42	94.90	97.51	98.60
12	184/3495	66.74	87.31	80.44	69.36	84.98	84.95	57.56	82.18	89.87
13	456/8660	77.10	75.30	81.08	75.67	76.54	86.51	67.25	84.27	84.14
14	928/17632	90.05	91.82	87.45	91.65	91.31	98.42	98.74	94.68	98.03
15	57/1079	57.89	73.23	84.94	72.41	96.10	91.42	97.43	84.93	82.85
16	3770/71631	99.33	99.73	99.71	99.38	99.95	99.78	99.65	99.76	99.81
OA	12,876	93.40	94.28	95.09	94.41	94.27	96.87	94.05	96.47	97.51
K	/	92.28	93.30	94.25	93.45	93.28	96.34	93.05	95.87	97.09
AA	244,654	85.30	87.82	89.38	86.94	92.06	92.97	89.24	92.47	94.10
Training Time (min.)		1.7	4.40	4.5	4.71	11	26.05	12.17	29	3.38
Testing Time (sec.)		19	23	23.3	43	37	117	37	60	22

it is observed that the classification accuracies closest to the PM are acquired with A2S2KRes. According to class-based classification performance in the PM, the highest classification accuracies were obtained in the 6th, 12th, 13th and 22nd classes with 98.99%, 92.30%, 95.31% and 96.74%, respectively. The best classification accuracies for Classes 1, 15, 16, 19, and 20 were obtained in FuSENet with

99.34%, 99.71%, 99.13%, 98.53%, and 98.01%. The best classification accuracies were in S3EResBoF with 95.38%, 98.24%, 100%, 99.20% for classes 2, 7, 9, 14%, DLEM with 98.08% and 99.89% for classes 3 and 4, FC3DCNN with 97.66% for class 18 obtained. Finally, the best classification accuracies were obtained with A2S2KRes with 91.65%, 87.18%, 95.95%, 93.14%, 97.96% for classes 5, 8, 10, 11,

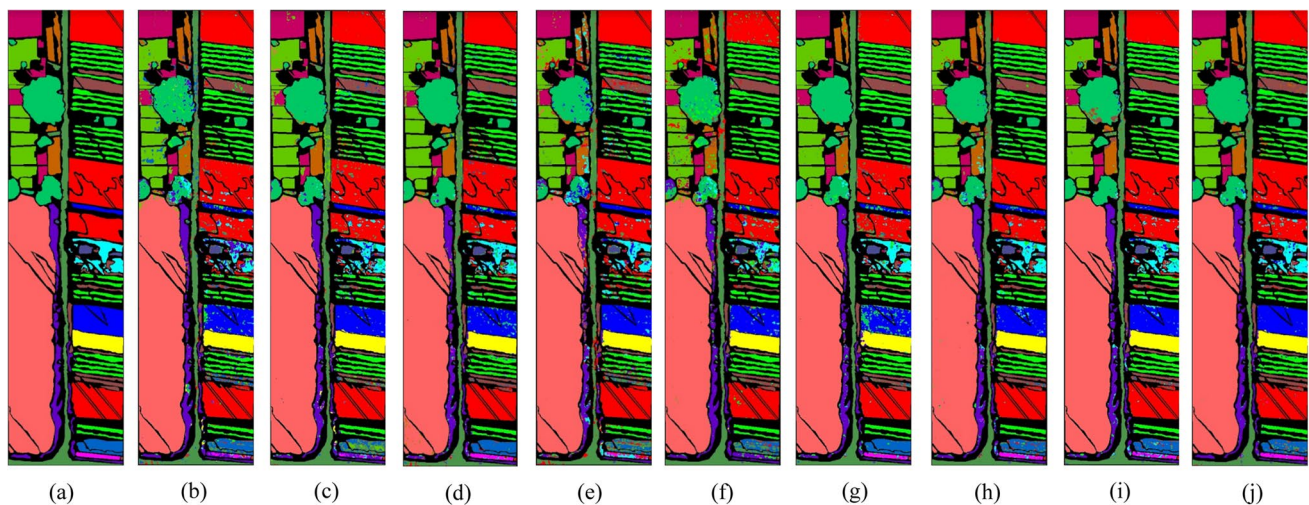


Fig. 8 Classification maps for WHHC. (a) Ground truth, (b)–(j) 2D CNN, 3D CNN, FC3DCNN, HybridSN, S3EResBoF, A2S2KRes, FuSENet, DLEM, and PM

Table 5 Classification accuracies for the WHHH (%)

No	Train/Test	2D CNN	3D CNN	FC3D CNN	HybridSN	S3EResBoF	A2S2KRes	FuSENet	DLEM	PM
1	702/13339	98.41	98.05	98.05	97.60	93.77	99.28	99.34	97.94	99.10
2	176/3336	83.97	89.87	70.32	86.32	95.38	94.70	84.14	94.83	94.81
3	1091/20730	95.75	92.26	97.77	95.38	95.17	97.56	96.03	98.08	97.78
4	8164/155121	99.30	99.47	99.12	99.34	99.59	99.40	98.15	99.89	99.70
5	311/5907	78.36	71.53	81.22	57.72	52.48	91.65	88.43	78.25	89.64
6	2228/42329	97.13	97.31	97.89	96.98	98.52	98.95	97.77	98.26	98.99
7	1205/22898	90.16	90.70	95.65	88.30	98.24	93.99	94.47	95.92	96.68
8	203/3851	60.73	63.50	76.19	47.30	86.02	87.18	83.29	68.97	83.93
9	541/10278	97.69	97.90	99.32	98.11	100	99.29	99.38	99.01	98.65
10	620/11774	88.19	82.52	88.96	71.34	90.06	95.95	94.63	89.58	95.45
11	551/10464	84.29	82.31	85.93	75.27	82.45	93.14	89.65	90.90	91.38
12	448/8506	69.24	74.66	81.22	77.96	73.33	91.97	74.18	89.64	92.30
13	1125/21382	83.26	87.44	88.58	78.65	82.91	91.39	90.46	88.97	95.31
14	368/6988	92.35	91.74	95.35	90.22	99.20	96.73	99.05	93.29	96.82
15	50/952	71.60	80.14	85.18	91.36	50.10	92.21	99.71	93.31	91.70
16	363/6899	94.46	95.00	97.61	93.03	98.73	98.22	99.13	96.14	98.90
17	150/2860	87.57	89.83	77.12	87.16	76.87	97.96	92.86	96.74	93.00
18	161/3056	92.57	95.39	97.66	89.14	91.77	90.08	83.03	97.56	96.24
19	435/8277	90.85	92.59	93.35	88.26	89.79	94.69	98.53	93.99	96.34
20	174/3312	82.64	90.09	91.36	75.24	91.32	94.71	98.01	93.37	90.88
21	66/1262	34.24	76.94	44.72	43.71	50.26	64.35	55.78	39.13	71.24
22	202/3838	90.69	87.88	95.25	85.68	90.81	92.79	81.42	89.72	96.74
OA	19,334	93.81	94.12	95.38	92.15	93.54	97.10	95.51	96.31	97.64
K	/	92.17	92.56	95.38	90.05	91.90	96.34	94.31	95.33	97.02
AA	367,359	84.70	87.60	88.08	82.46	85.76	93.46	90.79	90.16	93.89
Training Time (min.)		2.64	19.78	13.48	11.12	24.07	43.9	27.75	47	5.7
Testing Time (sec.)		65	334	269	195	126.5	341	254	277	177

17 and 76.94% for class 21 with 3D CNN. According to Table 5, the PM has shorter training–testing times than all other advanced methods except 2D CNN. Considering all the information in Table 5, it was concluded that the PM achieved better performance in a shorter time.

In addition, the overall accuracy (OA) value obtained with different principal components for the three datasets is given in Fig. 10. When Fig. 10 is examined, it is seen that the best classification results are obtained when 20 principal components are taken. Therefore, within the scope of the study, after applying PCA, 20 principal components were taken and applications were carried out. For the proposed method, the effect of different 3D data cubes on the classification results in three different datasets was examined and shown in Fig. 11. When Fig. 11 is examined, it is seen that the best classification accuracy is obtained in 7×7 . Therefore, in our study, 3D cubes of 7×7 size were used in all datasets.

Detailed analyzes of the components of the proposed method were performed on three different data sets and the

classification results are given in Table 6. The proposed method is a powerful network architecture consisting of multiple components. As shown in Table 6, the proposed method (Model_PM) showed a robust structure with Multipath 3D CNN, SENet, Multipath 2D DSC components and produced the highest classification results. By using these components, OA values of 99.86%, 97.51%, and 97.64% were obtained in WHLK, WHHC and WHHH datasets, respectively. When Table 6 is examined, it is seen that the lowest classification performance is obtained in Model_6, where 3D CNN and 2D CNN are not used together. In Model_4, where the classification results closest to the proposed method were obtained, only 2D CNN was not used. Considering the methods in which 3D CNN is not used, it is seen that the classification results are low. The following conclusion can be drawn from here. Since both spectral and spatial features are extracted with 3D CNN, the use of 3D CNN increases the classification performance. Also, when Model_2 is examined, it is seen that hybrid 3D/2D CNN and SENet are used together. The classification

Fig. 9 Classification maps for WHHH. (a) Ground truth, (b)–(j) 2D CNN, 3D CNN, FC3DCNN, HybridSN, S3EResBoF, A2S2KRes, FuSE-Net, DLEM, and PM

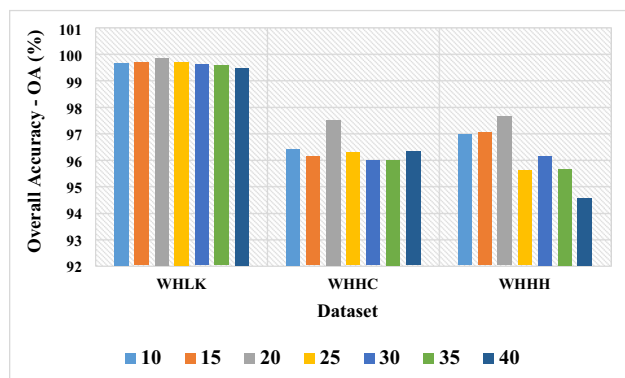
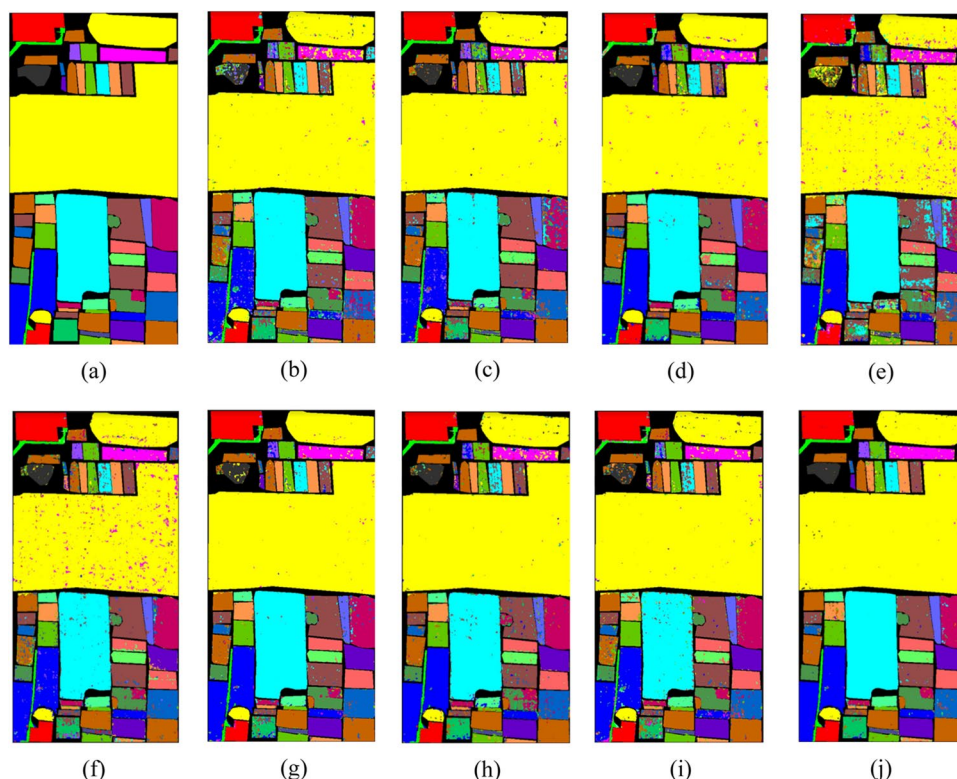


Fig. 10 OA values of different principal components obtained after PCA

results obtained are 99.75%, 96.36% and 95.07%, respectively. In the Hybrid CNN model, in Model_4 where 2D DSC is used instead of 2D CNN, 99.74%, 96.42% and 95.99% results were obtained, respectively. Finally, using the WHLK dataset, the parameter numbers and training times of the different components are given in Table 7. When Table 7 is examined, it is seen that the highest number of parameters is obtained in the use of 3D CNN. The minimum number of parameters was obtained with 2D CNN from which only spatial features were extracted. When Table 6 and Table 7 are examined together, although the use of 3D CNN increases the number of parameters, it increases

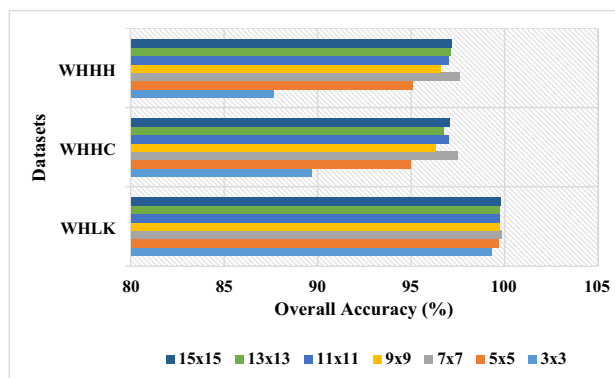


Fig. 11 Effect of different 3D data cube dimensions on classification accuracy on three datasets

the classification performance. It can be seen from Table 7 that the methods in which 3D CNN and 2D CNN are used together reduce the number of parameters. When 2D DSC is used instead of 2D CNN, the number of trainable parameters can be further reduced.

Conclusions

This study aims to develop a new deep feature extraction method for HSI classification. For this purpose, a method consisting of a combination of a multipath

Table 6 Analysis of the components of proposed method

<i>Models</i>	<i>Components</i>				<i>Overall Accuracy—OA (%)</i>		
	<i>3D CNN</i>	<i>SENet</i>	<i>2D CNN</i>	<i>2D DSC</i>	<i>WHLK</i>	<i>WHHC</i>	<i>WHHH</i>
<i>Model_1</i>	✓	X	✓	✓	%99.70	%95.85	%93.59
<i>Model_2</i>	✓	✓	✓	X	%99.75	%96.36	%95.07
<i>Model_3</i>	✓	X	✓	X	%99.72	%95.82	%94.41
<i>Model_4</i>	✓	✓	X	✓	%99.74	%96.42	%95.99
<i>Model_5</i>	X	✓	✓	✓	%99.68	%96.43	%92.79
<i>Model_6</i>	X	✓	X	✓	%99.60	%95.11	%91.53
<i>Model_PM</i>	✓	✓	✓	✓	%99.86	%97.51	%97.64

Table 7 Number of parameters and training times of different components using the WHLK dataset

<i>Component</i>	<i>Parameters</i>	<i>Training Time (min.)</i>
<i>2D CNN+SENet</i>	553.424	1.2
<i>3D CNN+SENet</i>	7.161.560	12.55
<i>Hybrid 3D/2D CNN+SENet</i>	3.463.565	4.36
<i>PM (Hybrid 3D CNN/2D DSC+SENet)</i>	3.451.973	2.66

Hybrid CNN and a SE network is proposed. Hybrid CNN consists of multipath 3D CNN and multipath 2D DSC layers. In the PM, richer feature information is obtained from HSIs using a multipath structure and extracted features with different kernel sizes. 3D CNN was used to extract both spectral and spatial features simultaneously. Using only 3D CNN increases computational complexity. With the hybrid methods obtained by using 2D CNN and 3D CNN, computational complexity is avoided and more spatial features are extracted simultaneously. However, within the scope of this study, 2D convolution layers in 2D CNN have been replaced with 2D DSC layers. It is aimed to increase the classification accuracies by reducing the number of trainable parameters with the use of deeply separable convolution layers. Moreover, SE network is combined with the Hybrid CNN to improve its performance in feature extraction and HSI classification in this study. The SE network is designed to develop the representation quality of a CNN. WHLK, WHHC and WHHH datasets were used in the applications to analyze the PM. Using a 5% training sample with WHLK, WHHC and WHHH datasets, OA values of 99.86%, 97.51% and 97.64% were obtained, respectively. In addition, the PM was compared to the latest technology methods in the literature. The PM outperforms the compared methods for all evaluation metrics across all dataset classes. The classification results show that the PM can be used in real-world applications.

Author's contribution Author 1: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing—review & editing.

Data availability Not applicable.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ahmad M, Khan A, Khan AM et al (2019) Spatial prior fuzziness pool-based interactive classification of hyperspectral images. *Remote Sens* 11:1–19. <https://doi.org/10.3390/rs11091136>
- Ahmad M, Khan AM, Mazzara M, et al (2020) A fast and compact 3-D CNN for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 1–5. <https://doi.org/10.1109/LGRS.2020.3043710>
- Alcolea A, Paoletti ME, Haut JM et al (2020) Inference in supervised spectral classifiers for on-board hyperspectral imaging: An overview. *Remote Sens* 12:1–29. <https://doi.org/10.3390/rs12030534>
- Ben Hamida A, Benoit A, Lambert P, Ben Amar C (2018) 3-D deep learning approach for remote sensing image classification. *IEEE Trans Geosci Remote Sens* 56:4420–4434. <https://doi.org/10.1109/TGRS.2018.2818945>
- Blanzieri E, Melgani F (2008) Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Trans Geosci Remote Sens* 46:1804–1811. <https://doi.org/10.1109/TGRS.2008.916090>
- Chen Y, Zhao X, Jia X (2015) Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:2381–2392. <https://doi.org/10.1109/JSTARS.2015.2388577>
- Cheng G, Li Z, Han J et al (2018) Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification. *IEEE Trans Geosci Remote Sens* 56:6712–6722. <https://doi.org/10.1109/TGRS.2018.2841823>
- Dang L, Pang P, Lee J (2020) Depth-wise separable convolution neural network with residual connection for hyperspectral image classification. *Remote Sens* 12:1–20. <https://doi.org/10.3390/rs12203408>
- Data H, Chen Y, Lin Z et al (2015) Deep Learning-Based Classification of Hyperspectral Data. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7:1–14

- Ding Y, Zhang Z, Zhao X et al (2022) Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification. *Neurocomputing* 501:246–257. <https://doi.org/10.1016/j.neucom.2022.06.031>
- Firat H, Asker ME, Bayindir Mİ, Hanbay D (2022a) 3D residual spatial–spectral convolution network for hyperspectral remote sensing image classification. *Neural Comput Appl* 8:.. <https://doi.org/10.1007/s00521-022-07933-8>
- Firat H, Asker ME, Hanbay D (2022b) Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN. *Remote Sens Appl Soc Environ* 100694. <https://doi.org/10.1016/j.rsase.2022.100694>
- Firat H, Hanbay D (2021) Classification of hyperspectral images using 3D CNN based ResNet50. *SIU 2021 - 29th IEEE Conf Signal Process Commun Appl Proc* 6–9. <https://doi.org/10.1109/SIU53274.2021.9477899>
- Firat H, Hanbay D (2022) 3 Boyutlu Evrişimsel Sinir Ağı Kullanılarak Hipspektral Görüntülerin Sınıflandırılması Classification of Hyperspectral Images Using 3D Convolutional Neural Network. *Türk Doğa Ve Fen Derg* 11:19–28
- Firat H, Hanbay D (2023) Comparison of 3D CNN based deep learning architectures using hyperspectral images. *J Fac Eng Archit Gazi Univ* 38:521–534. <https://doi.org/10.17341/gazimmfd.977688>
- Firat H, Asker ME, Ilyas M, Hanbay D (2022a) Spatial-spectral classification of hyperspectral remote sensing images using 3D CNN based LeNet-5 architecture. *Infrared Phys Technol* 127:.. <https://doi.org/10.1016/j.infrared.2022.104470>
- Firat H, Emin M, Mehmet A, et al (2022b) Hybrid 3D / 2D Complete Inception Module and Convolutional Neural Network for Hyperspectral Remote Sensing Image Classification. *Neural Process Lett* 1–44. <https://doi.org/10.1007/s11063-022-10929-z>
- Firat H, Hanbay D (2022) 4CF-Net: New 3D convolutional neural network for spectral spatial classification of hyperspectral remote sensing images. *J Fac Eng Archit Gazi Univ* 37:439–453. <https://doi.org/10.17341/gazimmfd.901291>
- Gao H, Chen Z, Li C (2021) Sandwich convolutional neural network for hyperspectral image classification using spectral feature enhancement. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:3006–3015. <https://doi.org/10.1109/JSTARS.2021.3062872>
- Gong H, Li Q, Li C, et al (2021) Multiscale information fusion for hyperspectral image classification based on hybrid 2D-3D CNN. *Remote Sens* 13:.. <https://doi.org/10.3390/rs13122268>
- Han Y, Wei C, Zhou R, et al (2020) Combining 3D-CNN and squeeze-and-excitation networks for remote sensing sea ice image classification. *Math Probl Eng* 2020:.. <https://doi.org/10.1155/2020/8065396>
- Hanbay K (2020) Hyperspectral image classification using convolutional neural network and two-dimensional complex Gabor transform. *J Fac Eng Archit Gazi Univ* 35:443–456. <https://doi.org/10.17341/gazimmfd.479086>
- Hong D, Han Z, Yao J, et al (2022) SpectralFormer: rethinking hyperspectral image classification with transformers. *IEEE Trans Geosci Remote Sens* 60:.. <https://doi.org/10.1109/TGRS.2021.3130716>
- Hörig B, Kühn F, Oschütz F, Lehmann F (2001) HyMap hyperspectral remote sensing to detect hydrocarbons. *Int J Remote Sens* 22:1413–1422. <https://doi.org/10.1080/01431160120909>
- Huang J, He H, Lv R et al (2022) Non-destructive detection and classification of textile fibres based on hyperspectral imaging and 1D-CNN. *Anal Chim Acta* 1224:340238. <https://doi.org/10.1016/j.aca.2022.340238>
- Iyer P, A S, Lal S (2021) Deep learning ensemble method for classification of satellite hyperspectral images. *Remote Sens Appl Soc Environ* 23:100580. <https://doi.org/10.1016/j.rsase.2021.100580>
- Jia J, Wang Y, Chen J et al (2020) Status and application of advanced airborne hyperspectral imaging technology: A review. *Infrared Phys Technol* 104:103115. <https://doi.org/10.1016/j.infrared.2019.103115>
- Kang X, Duan P, Li S (2020) Hyperspectral image visualization with edge-preserving filtering and principal component analysis. *Inf Fusion* 57:130–143. <https://doi.org/10.1016/j.inffus.2019.12.003>
- Karadağ B, Ari A, Karadağ M (2021) Derin Öğrenme modellerinin sinirsel stil aktarımı performanslarının karşılaştırılması. *J Politeknik* 0900:1611–1622. <https://doi.org/10.2339/politeknik.885838>
- Lanthier Y, Bannari A, Haboudane D et al (2008) Hyperspectral data segmentation and classification in precision agriculture: A multi-scale analysis. *Int Geosci Remote Sens Symp* 2:585–588. <https://doi.org/10.1109/IGARSS.2008.4779060>
- Lee J, Kim Y, Jeong M, et al (2018) 3D convolutional neural networks for soccer object motion recognition. *Int Conf Adv Commun Technol ICACT* 2018-Febru:354–358. <https://doi.org/10.23919/ICACT.2018.8323754>
- Li S, Zhang K, Hao Q et al (2018) Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters. *IEEE Geosci Remote Sens Lett* 15:1605–1609. <https://doi.org/10.1109/LGRS.2018.2853705>
- Li W, Wu G, Zhang F, Du Q (2017) Hyperspectral image classification using deep pixel-pair features. *IEEE Trans Geosci Remote Sens* 55:844–853. <https://doi.org/10.1109/TGRS.2016.2616355>
- Lin Z, Ji K, Leng X, Kuang G (2019) Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geosci Remote Sens Lett* 16:751–755. <https://doi.org/10.1109/LGRS.2018.2882551>
- Liu X, Yu J, Kurihara T et al (2022) Hyperspectral imaging for green pepper segmentation using a complex-valued neural network. *Optik (stuttg)* 265:169527. <https://doi.org/10.1016/j.ijleo.2022.169527>
- Lu G, Zhang W, Wang Z (2022) Optimizing depthwise separable convolution operations on GPUs. *IEEE Trans Parallel Distrib Syst* 33:70–87. <https://doi.org/10.1109/TPDS.2021.3084813>
- Makantasis K, Karantzas K, Doulamis A, Doulamis N (2015) Deep supervised learning for hyperspectral data classification through convolutional neural networks. *Int Geosci Remote Sens Symp* 2015-Novem:4959–4962. <https://doi.org/10.1109/IGARSS.2015.7326945>
- Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens* 42:1778–1790. <https://doi.org/10.1109/TGRS.2004.831865>
- Mohan A, Venkatesan M (2020) HybridCNN based hyperspectral image classification using multiscale spatio-spectral features. *Infrared Phys Technol* 108:.. <https://doi.org/10.1016/j.infrared.2020.103326>
- Mou L, Ghamisi P, Zhu XX (2017) Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55:3639–3655. <https://doi.org/10.1109/TGRS.2016.2636241>
- Mughees A, Tao L (2017) Efficient deep auto-encoder learning for the classification of hyperspectral images. *Proc - 2016 Int Conf Virtual Real Vis ICVRV* 2016 44–51. <https://doi.org/10.1109/ICVRV.2016.16>
- Okwuashi O, Ndehedehe CE (2020) Deep support vector machine for hyperspectral image classification. *Pattern Recognit* 103:107298. <https://doi.org/10.1016/j.patcog.2020.107298>
- Paoletti ME, Haut JM, Plaza J, Plaza A (2018) A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J Photogramm Remote Sens* 145:120–147. <https://doi.org/10.1016/j.isprsjprs.2017.11.021>
- Rajendran T, Valsalan P, Amutharaj J, et al (2022) Hyperspectral image classification model using squeeze and excitation network with deep learning. *Comput Intell Neurosci* 2022:.. <https://doi.org/10.1155/2022/9430779>

- Ratlé F, Camps-Valls G, Weston J (2010) Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 48:2271–2282. <https://doi.org/10.1109/TGRS.2009.2037898>
- Roy SK, Chatterjee S, Bhattacharyya S et al (2020a) Lightweight spectral-spatial squeeze-and- excitation residual bag-of-features learning for hyperspectral classification. *IEEE Trans Geosci Remote Sens* 58:5277–5290. <https://doi.org/10.1109/TGRS.2019.2961681>
- Roy SK, Dubey SR, Chatterjee S, Chaudhuri BB (2020b) FuSENet: Fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. *IET Image Process* 14:1653–1661. <https://doi.org/10.1049/iet-ipr.2019.1462>
- Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) HybridSN: exploring 3D-2D CNN feature hierarchy for hyperspectral image classification. *arXiv* 17:277–281
- Roy SK, Manna S, Song T, Bruzzone L (2020c) Attention-based adaptive spectral-spatial kernel resnet for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 59:7831–7843. <https://doi.org/10.1109/TGRS.2020.3043267>
- Song W, Li S, Fang L (2018) Hyperspectral image classification with deep feature fusion network. *IEEE Trans Geosci Remote Sens* 99:3173–3184. <https://doi.org/10.1109/IGARSS.2019.8898520>
- Üzen H, Turkoglu M, Aslan M, Hanbay D (2022) Depth-wise squeeze and excitation block-based efficient-unet model for surface defect detection. *Vis Comput*. <https://doi.org/10.1007/s00371-022-02442-0>
- Uzen H, Turkoglu M, Hanbay D (2021) Texture defect classification with multiple pooling and filter ensemble based on deep neural network. *Expert Syst Appl* 175:114838. <https://doi.org/10.1016/j.eswa.2021.114838>
- Wang Y, Yu W, Fang Z (2020) Multiple Kernel-based SVM classification of hyperspectral images by combining spectral, spatial, and semantic information. *Remote Sens* 12:. <https://doi.org/10.3390/RS12010120>
- Xu H, Zhang H, He W, Zhang L (2019) Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification. *Neurocomputing* 360:138–150. <https://doi.org/10.1016/j.neucom.2019.06.023>
- Yang X, Ye Y, Li X et al (2018) Hyperspectral image classification with deep learning models. *IEEE Trans Geosci Remote Sens* 56:5408–5423. <https://doi.org/10.1109/TGRS.2018.2815613>
- Zhang M, Li W, Du Q (2018) Diverse region-based CNN for hyperspectral image classification. *IEEE Trans Image Process* 27:2623–2634. <https://doi.org/10.1109/TIP.2018.2809606>
- Zhong Y, Hu X, Luo C et al (2020) WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens Environ* 250:112012. <https://doi.org/10.1016/j.rse.2020.112012>
- Zhou F, Hang R, Liu Q, Yuan X (2019) Hyperspectral image classification using spectral-spatial LSTMs. *Neurocomputing* 328:39–47. <https://doi.org/10.1016/j.neucom.2018.02.105>
- Zhu L, Chen Y, Ghamisi P, Benediktsson JA (2018) Generative adversarial networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 56:5046–5063. <https://doi.org/10.1109/TGRS.2018.2805286>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.