Applied Soft Computing 130 (2022) 109650

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Compression and reinforce variation with convolutional neural networks for hyperspectral image classification

Dalal AL-Alimi [a,c], Zhihua Cai [a,*], Mohammed A.A. Al-qaness [b,c,*], Abdelghani Dahou [d], Eman Ahmed Alawamy [e], Sakinatu Issaka [f]

[a] School of Computer Science, China University of Geosciences, Wuhan 43 0074, China
[b] College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China
[c] Faculty of Engineering, Sana'a University, Sana'a 12544, Yemen
[d] Department of Mathematics and Computer Science, Faculty of Science and Technology, University of Ahmed DRAIA, Adrar 01000, Algeria
[e] School of Mathematics and Statistics, Central South University, Changsha 410083, China
[f] School of Environmental Studies, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 5 December 2021
Received in revised form 9 September 2022
Accepted 20 September 2022
Available online 24 September 2022

Keywords:
Feature selection
Hyperspectral image
Classification
Feature fusion
Hybrid model
Dimensionality reduction

ABSTRACT

In Hyperspectral images (HSI), dimensionality reduction methods (DRM) play a critical role in reducing the input data dimension and complexity. As much as the deep learning methods (DLM) have presented very aggressive achievements, preprocessing methods and DRM are very important to enhance the learning of DLMs. This study introduces a novel DRM called Compression and Reinforced Variation (CRV), which is used to reduce the input data dimension. The CRV minimizes the gap between the big and small related data in each class and omits the noise and redundant data. It selects the most informative features and normalizes them to enhance data distribution before inserting them into the learning model. The learning model of this study is multi-hybrid deep learning (MHDL) model to improve the extraction of multi-class HSI and spectral–spatial features. MHDL is a novel classification model that includes hybrid layers of conventional neural networks and batch normalization to avoid overfitting, normalizing the training, and extracting the spectral–spatial features for HSI. The proposed CRV provided highly efficient methods for reducing the HSI dimension and improving the classification accuracy of the MHDL model. In contrast to other conventional DRMs, CRV gave the highest accuracy in the shortest time. CRV-MHDL was also compared to seven existing classification models for three distinct datasets, and the findings demonstrated that the CRV-MHDL outperforms all of them by more than 2%. The code of this study is available at this link: https://github.com/DalalAL-Alimi/CRV.

1. Introduction

The classification of Hyperspectral images (HSI) is a necessary process for different earth observation applications [1], such as war areas [2], military [3,4], environmental monitoring [5], agriculture [6], small object detection [7–9], food quality [10], medical [11,12], and others. HSI can extract spectral data from hundreds of surface object continuous spectrum segments. The spatial resolution of HSI data sets has substantially improved due to the rapid development of remote-sensing technology, which vastly improves the ability of HSI data sets to express distinct objects appropriately.

As described in [1], there are several critical challenges with HSI classification tasks. For example, hyperspectral data has hundreds of band values, and the information between the spectral bands is usually redundant, resulting in a large data dimension and a high computing demand. More so, the presence of mixed pixels causes significant interference in the categorization of HSI, as a single pixel frequently correspondings to numerous object categories and is commonly misclassified. Furthermore, manually labeling HSI samples are expensive, resulting in a tiny number of off-the-shelf labeled samples.

In high-dimensional data analysis, visualization, and modeling, dimensionality reduction methods (DRM) are commonly employed as preprocessing. DRM seeks to increase the performance of estimated accuracy, visualization, and comprehension of learned knowledge in general. DRMs can generally be divided

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding authors.
E-mail addresses: dalal@cug.edu.cn (D. AL-Alimi), zhcai@cug.edu.cn (Z. Cai), alqaness@zjnu.edu.cn (M.A.A. Al-qaness), dahou.abdghani@univ-adrar.edu.dz (A. Dahou), emanahmed@csu.edu.cn (E.A. Alawamy), sakina@cug.edu.cn (S. Issaka).

https://doi.org/10.1016/j.asoc.2022.109650
1568-4946/

Abbreviations

HSI	Hyperspectral Image
DRM	Dimensionality Reduction Method
CRV	Compression And Reinforced Variation
MHDL	Multi-Hybrid Deep Learning
FSA	Feature Selection Algorithm
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
ICA	Independent Component Analysis
KPCA	Kernel PCA
mRMR	Minimum-Redundancy Maximum-Relevance
DL	Deep Learning
CNN	Convolution Neural Network
CDSCN	Cascade Dual-Scale Crossover Neural Network
3D	Three Dimensional
2D	Two Dimensional
FSM	Feature Selection Method
BN	Batch Normalization
RAF	ReLu Activation Function
FC	Fully Connected Layer
MP	Max-Pooling Layer
MLP	Multilayer Perceptron
HybridSN	Hybrid Spectral CNN

into feature extraction and feature (band) selection [13–17]. The DRM is one of the most critical HSI processes, aiming to reduce model complexity and overfitting, and these new lower dimensions of features represent the original ones. The feature extraction approach reduces the dimensionality through particular mathematical processes to generate a new subset of features that are a part of the original dataset and retain only the pertinent data that can improve the final goal while discarding the rest. On the other hand, feature selection algorithms (FSA) select a subset of features most relevant to the problem to improve computational efficiency and reduce generated model errors by deleting unrelated features or noise. FSA methods have three types, filter, wrapper, and embedded [14,18,19].

Principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), kernel PCA (KPCA) [5,20,21], and region-aware latent features fusion-based clustering [18] are examples of feature extraction methods. The most common compressing method in HSI is PCA. PCA uses the correlation between features to find data patterns. It aims to identify the highest variance directions in high-dimensional data and project them onto a subspace with the same or fewer dimensions as the original. On the other hand, KPCA is nonlinear unsupervised features extraction, the kernelized version of PCA [5,10,22–24]. LDA is a linear supervised feature extraction method that aims to minimize class variations. The general concept behind LDA is very similar to PCA.

In contrast, PCA attempts to find the orthogonal component axes of maximum variance. LDA aims to find the feature subspace that optimizes class separability. Thus, it increases computational efficiency, reduces overfitting, and highlights the quality of the classification. It has been widely used to classify agricultural and food products and other applications based on hyperspectral data [25–27]. ICA is a linear, supervised feature extraction; considered a further step of PCA and a powerful tool for extracting source signals or valuable information from the original data. Compared to the PCA and LDA, ICA optimizes higher-order statistics such as kurtosis (non-Gaussian), yielding independent components [20].

In FSA, filter methods nominate features according to specific predefined criteria before feeding to a learning model such as minimum-redundancy maximum-relevance (mRMR), trivariate mutual information-clonal selection algorithm, distance-based criteria, consistency-based criteria, and manifold learning-based criteria [28,29]. The wrapper method chooses and evaluates the candidate features through a chosen training model. So, the research algorithm of the best subset of features is basically "wrapped" around the model. This feature selection method is considered costly due to its computational complexity and the long execution time. It is better in classification than in the filter methods, but filter methods are faster, less complex, and better chosen for high dimensional datasets compared to wrapper methods. Recursive feature elimination (SVM-RFE) is one of the wrapper methods. The SVM-REF employs the weight vector as a ranking criterion to select the features that lead to the most considerable margin of class separation. The embedded methods use the advantages of filters and wrappers methods like the least absolute shrinkage and selection operator and the partial least square [14,19,30,31].

Over the past decades, automatic feature representation and extraction using machine learning techniques have gained popularity over handcrafted techniques for HSI classification [32,33].

For instance, invariant attribute profiles [34] and texture profiles [35] were effective techniques for extracting spatial–spectral features from HSI. In addition, methods such as sparse representation, known as subspace-based learning and manifold learning [36,37], have proven their ability to capture the high-dimensional structure of HSI by mapping the high-dimensional original space to low-dimensional subspace. However, the methods mentioned above are limited in data fitting and representation ability [38,39]. In recent years, deep learning models (DL) have superseded the methods mentioned above on many levels, including feature extraction or representation, feature selection, and classification [40].

2. Literature review

Many DL models have been proposed to address the problems of traditional feature representation and HSI classification. The convolution neural network (CNN) and its variants are well-known DL models based on hierarchical feature learning and classification [41]. CNNs are widely used in HSI classification problems. They are commonly composed of a stack of the convolutional layer with different kernel sizes and activation functions to represent and extract features. CNNs can be used to build shallow or deep structures based on the complexity and abstraction of the extracted features. The DL models can detect different levels of abstraction, thus extracting various features from the HSI. Different designed structures and networks can be distinguished based on the DL model architecture. For instance, Zhao et al. [42] used a combination of two CNN networks representing shallow and deep network structures. The shallow network extracts the small object's features (low-abstract features) while the deep network extracts the big object's features.

Moreover, Cao et al. [43] proposed a cascaded network architecture named cascade dual-scale crossover neural network (CDSCN) to elicit the features from the HSI. The CDSCN was trained on a small number of HSI using a CNN with different kernel sizes to extract spatial and spectral features. The developed CDSCN was used to overcome the overfitting problem and to improve the classification accuracy by employing batch normalization and dropout regularization techniques [38]. Cao

et al. [44] performed HSI classification using an active deep learning model with Markov random field to improve the model performance in classification accuracy. The model was trained using a CNN network on a small set of 30 labeled pixels with feature selection to actively select informative pixels only. The proposed HSI classification model showed better accuracy than the baseline and other existing models. Hang et al. [45] integrated the attention mechanism in their HSI classification model. The authors developed a CNN with an attention mechanism where two network branches for spectral–spatial classifications were built using spectral–spatial attention. In addition, the adaptively weighted summation technique combined the results of the two branches. The experiment demonstrated that the developed sub-networks based on the attention mechanism outperformed the existing models. Yu et al. [46] used a fusion mechanism to build an HSI classification model using 2D-3D CNN architecture.

The model employed the 2D CNN sub-network as a feature extractor while the 3D sub-network as a band correlation exploitation network to extract the spectral features using small kernels. The proposed model performed well on the 40-feature extraction task, which improved the overall classification accuracy. Kang et al. [47] combined different levels of features automatically extracted using a dual-path CNN and residual learning mechanism to improve the classification accuracy. Paoletti et al. [48] developed a novel residual pyramid bottleneck to balance the model's computational complexity and accuracy. The proposed model elongated the spectral domain while reducing the extracted spatial shape features. Moreover, the residual connection and attention mechanism were widely used to perform HSI classification and feature extraction, such as in the works conducted by Alipour-Fard [49], Haut et al. [50], Wu et al. [51], and Xu et al. [52].

Even though DRMs were introduced to enhance the computation complexity and the data dimension, enhancing the feature selection methods and data distribution are still needed before feeding the input data into the learning model. Furthermore, many studies were proposed for HSI classification in the literature, using various classification techniques, such as the classical machine learning approaches, support vector machine [9], and decision tree [53]. Advance deep learning approaches have been used recently to enhance the classification process of the HSI, such as recurrent neural networks [54,55] and conventional neural networks (CNN) [56–58]. Although deep learning methods demonstrated impressive performance in classifying HSI, they still face certain limitations, such as dealing with and training different class sizes in multi-classes datasets (HSI) equally and enhancing the feature extraction of spectral–spatial features simultaneously. So, this study sums up the challenges of HSI as follows:

- Each class does not have a fixed size or appearance in HSI. Further, these images were mainly collected from various sources with varying resolutions, which complex the classification process.
- The HSIs are well-known for their enormous size variations, the number of classes and bands, the varied sizes of each class, and the data distribution's intricacy. All of these factors make selecting effective features more difficult.
- Designing a training model that can handle all the HSI challenges and achieve a high classification accuracy for the spectral–spatial features of various classes and sample sizes simultaneously.

This article proposes a novel feature selection method (FSM) that increases the variance and filters the data to reduce the band number in the input dataset and enhance the classification and training performance. This FSM is called Compression and Reinforced Variation (CRV). It minimizes the gap between the big and small related data of the same class and omits the noise and redundant data. Moreover, CRV normalizes the selected features before feeding them into the classification model to reduce its complexity and overfitting. The classification model of this study is the multi-hybrid deep learning (MHDL) model, whose structure enhances the classification of spectral–spatial features of multi-class HSI. It is based on the hybrid layers of 2 and 3D-CNN and batch normalization (BN) to optimize the extraction and void the overfitting. The significant contributions of this study are as follows:

- To introduce a novel feature selection method called Compression and Reinforced Variation (CRV) thus to pick critical features and reduce dimensionality.
- To implement CRV to improve data compression and classification. The CRV can improve the data distribution of all HSI to be more efficient and decreases performance time.
- To develop a multi-hybrid deep learning (MHDL) model to enhance the extraction of multi-class HS datasets and spectral–spatial features.
- To utilize the multi-size kernel of CNN layers and BN layers at the time of feature extraction to improve the extraction of the various classes and prevent overfitting.

3. Proposed methodologies

This section provides complete information on the operation of the compression and reinforced variation (CRV) method and explains this study model's structure, multi-hybrid deep learning (MHDL).

3.1. Compression and reinforced variation method

The hyperspectral dataset is an extensive dataset with many classes and bands. These classes mostly share the same values in all classes, called outliers or noise. In many cases, these values are important to distinguish some classes and support their differences, so they cannot be eliminated. The other problem is that hyperspectral datasets have many repeated bands and redundant data. Therefore, compressing and choosing quality features are essential to enhance the classification and the performance speed. In addition, knowing the input data distribution is necessary to know if it has a normal or skewed distribution before training the data; that helps to choose the best data normalization. Many preprocessing should be applied to the input dataset before training.

The compression and reinforced variation operation (CRV) works firstly by reducing the gap between large and small values to separate and know the critical noise values and taking the mean of each instance in the dataset. Then subtract it from the main dataset (x) to get the marker image. This marker image is used in the dilation process, which reduces the gap between small and large values according to the x—next, omitting unnecessary data to figure out the most critical features and choose them for the final processing. The final process enhances the extracted data distribution to be normal-like by using Gaussian transformation.

The following parameters are provided in Table 1 to facilitate the comprehension of the proposed method:

In this study, the process of compressing and choosing features depends on two images. The first is the main image (x), the input data. The second is the marker image (S), created based on x.

1. The first step is to generate the marker image as follows:

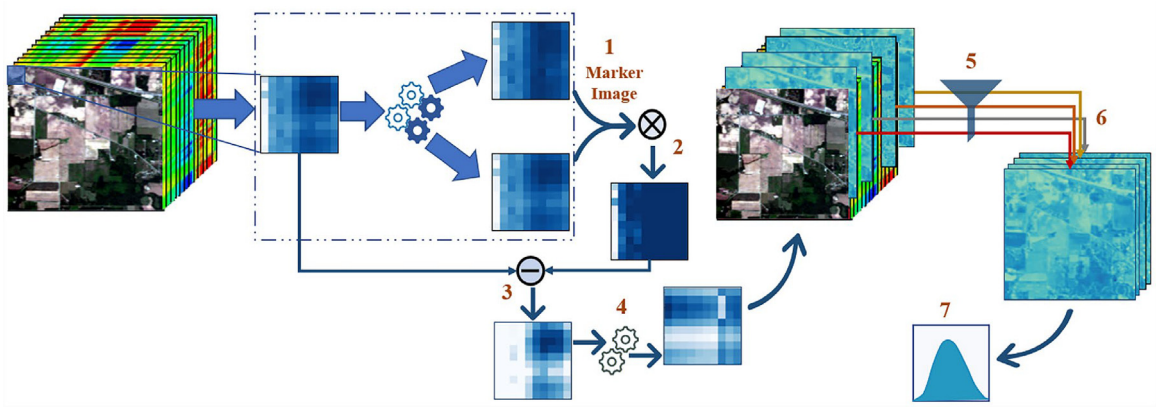


Fig. 1. The compression and reinforced variation (CRV) method.

Table 1
The description of the used parameters in the compression and reinforced variation (CRV) method.

Symbol	Description
x	The main image, which is the input data
c	The number of features for each instance
μ	The mean of each instance of the input dataset (x)
r	The number of instances
S	The marker image, which is created based on x and μ
N	The number of samples in each feature
k	The reduced dimension
$\mathbb{z}$	The final generated subset with the k dimension.
y	The normalization operation for $\mathbb{z}$ and the final output

(a) The mean of each instance should be calculated to create the marker image (S). The mean is the average value of each instance for all the features representing one class. The mean of each instance of the input dataset:

$$\mu_c = 1/c \sum_{c=1}^{c} (x_c) \tag{1}$$

Where c is the number of features for each instance, the output is a one-dimensional matrix that includes the mean value of each row of (x), $x \in R^{n \times d}$.

(b) Then this matrix is repeated to have the exact dimension of the main image (x). If c represents the number of features or bands in x, and r is the number of instances, the dimension of x is ($r \times c$). $q \times g$ are the numbers of features and instances in μ. So, the dimension of μ is changed to equal the dimension of x ($q \times g = r \times c$). So, the multi-dimensional of μ *is* $\mu_{(r,c)}$, as shown in Fig. 2(b).

(c) Next, the marker image (S) is created by subtracting the multi-dimension of μ from the input data x, Fig. 3:

$$S = x_{(r,c)} - \mu_{(r,c)} \tag{2}$$

2. After getting the marker image, the dilation can be calculated. Dilation (D) processes the marker image (S) based on the characteristics of x. It reconstructs the image by removing all the intensity fluctuations except the intensity peak. The high points or peaks of the S specify where processing begins, and the processing continues until the S values stop changing. Hence, D works to reconstruct the image by removing all the intensity fluctuations except the intensity peak. It reduces the gap between small and large values and the outliers, as can be seen in Fig. 4:

$$D = S \oplus x = \bigcup_{b \in x} S_b \tag{3}$$

where S_b is the translation of S by b.

3. The next step is to omit the output of Eq. (3) from the main image (x) to get only the essential and rich features that enhance the variance and classification, Fig. 5:

$$F = x - D \tag{4}$$

where x is the main input data, and D is the dilation; the results of Eq. (3).

4. After that, the mean of Eq. (4) was calculated to know the top richest features that affect the classification:

$$M_N = 1/\mathrm{N} \sum_{N=1}^{N} (F_c) \tag{5}$$

where N is the number of samples in each feature, and F is the output of Eq. (4).

5. Then, Eq. (4) results are sorted by decreasing order according to the rank of the mean of Eq. (5):

$$FD = sort\ (-1 \times F_m) \tag{6}$$

where F_m is the F dataset extracted from Eq. (4) and sorted by M of Eq. (5).

6. In this step, the same column indexes of FD_k from the input data (x_d) are taken to get the main subset ($\mathbb{z}_k$). $\mathbb{z}$ has been placed in an ordered sequence, beginning with the largest, k is the reduced dimension, and B_x and B_f represent the band indexes of the input and FD dataset. Hence the new subset data is part of the main dataset x, where $\mathbb{z}_k \in x_d$, $B_x = B_f$ *and* $k \leq d$; $\{\mathbb{z}^j\}_{j=1}^{k} \subseteq \{x^j\}_{j=1}^{d}$. In such a scenario, $\mathbb{z}_1$ denotes the most critical set of data bands, with subsequent sequences denoting progressively fewer essential features, Fig. 6. These steps extract the more practical features and omit the redundant bands (B), which enhance the classification and variance between classes and reduce the performance time.

7. Finally, this subset ($\mathbb{z}_k$) is fed into normalized data distribution operation to enhance data distribution. The normalization operation is called Gaussian transformation (GT):

$$y = G_y^{-1}\ (F_{\mathbb{z}}\ (\mathbb{z})) \tag{7}$$

where $F_{\mathbb{z}}$ is the Cumulative Distribution Function for the $\mathbb{z}$, and G_y^{-1} is the Gaussian Cumulative Distribution Function (GCDF) in y. Eq. (7) normalizes the data to be normal-like,

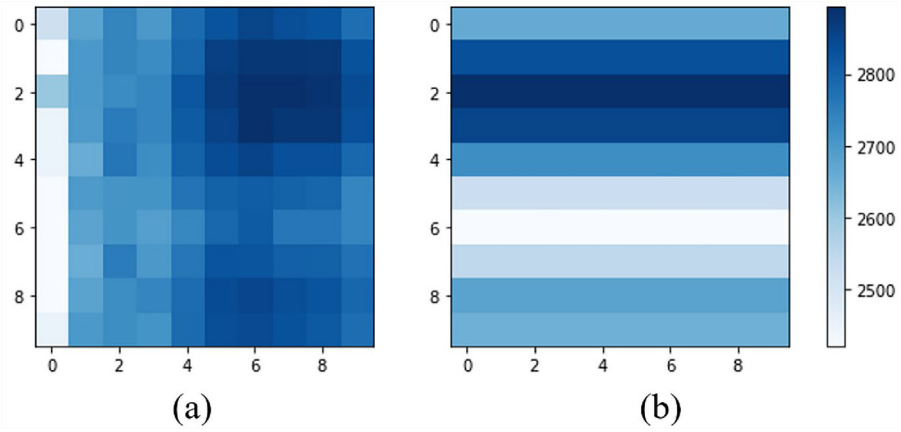


Fig. 2. (a) and (b) are the first 10 × 10 data arrays of the IPs dataset. (a) represents the main input image (x), and (b) represents the repeated mean data of each instance of x, which represents $\mu_{(r,c)}$.

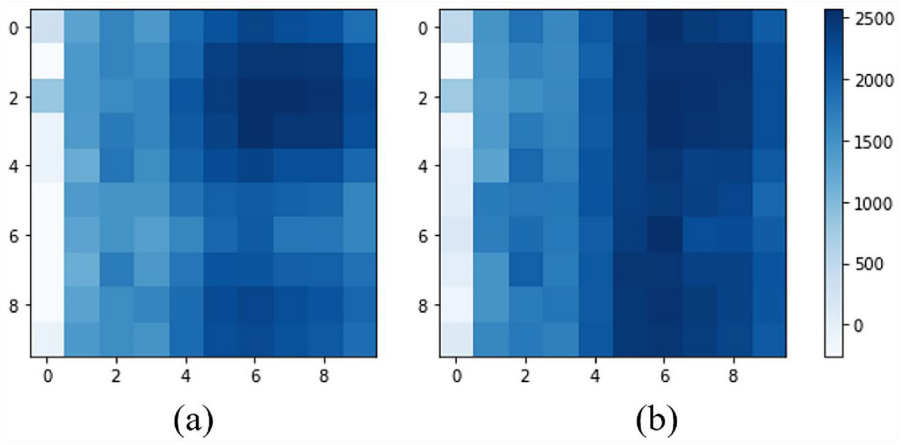


Fig. 3. (a) and (b) are the first 10 × 10 data arrays of the IPs dataset. (a) represents the main input image (x), and (b) represents the marker image (S).

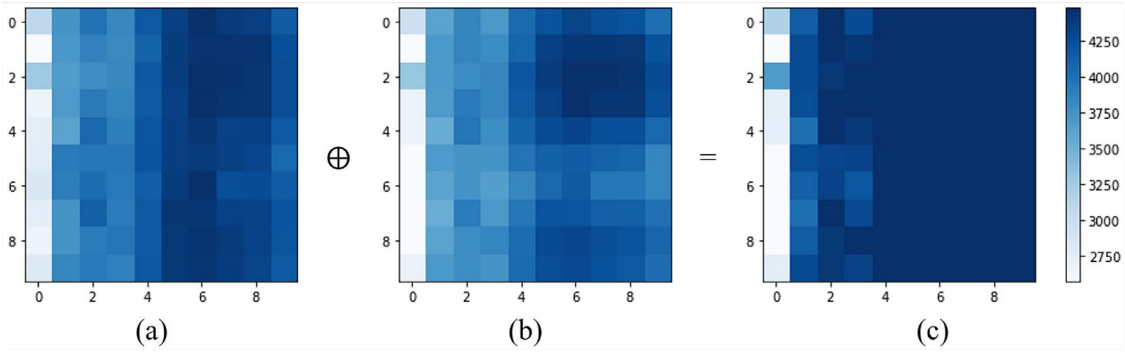


Fig. 4. (a), (b), and (c) are the first 10 × 10 data arrays of the IPs dataset (the first 10 × 10 data). (a) represents the marker image (S). (b) represents the main input image (x). (c) is the dilation operation results (D).

hence reducing the outliers and skewness. Fig. 1 describes and summarizes the whole steps and processes of the CRV.

3.2. The multi-hybrid deep learning model

Due to the nature of HSI, using only spectral information is not enough. When many different classes of the input data have identical spectral signatures, they can be discriminated through their shapes and texture. Furthermore, spatial information helps identify each neighboring pixel belonging to which class [5]. Thus, fusing spectral–spatial information is necessary to improve the classification. Because of the ability of conventional neural networks (CNN) to extract the features and enhance them by their layers, this study used 3D-CNN to extract spectral–spatial features and 2D-CNN to enhance the localization of each pixel.

The multi-hybrid deep learning (MHDL) model structure includes three consecutive layers of 3D-CNN followed by one layer of 2D-CNN, as seen in Fig. 7; both types of CNN layers have multi-scale kernels and ReLu as activation function (RAF). RAF is utilized for a variety of reasons. It has the advantage of introducing non-linearity in CNN. It does not activate all neurons

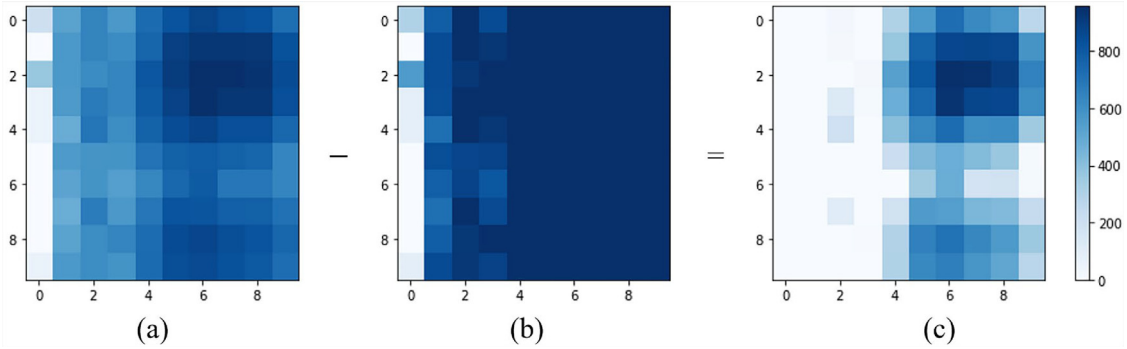


Fig. 5. (a), (b), and (c) arrays from the IPs dataset (the first 10 × 10 data). (a) represents the main input image (x). (b) is the dilation operation results (D). (c) represents the rich and most influential data.

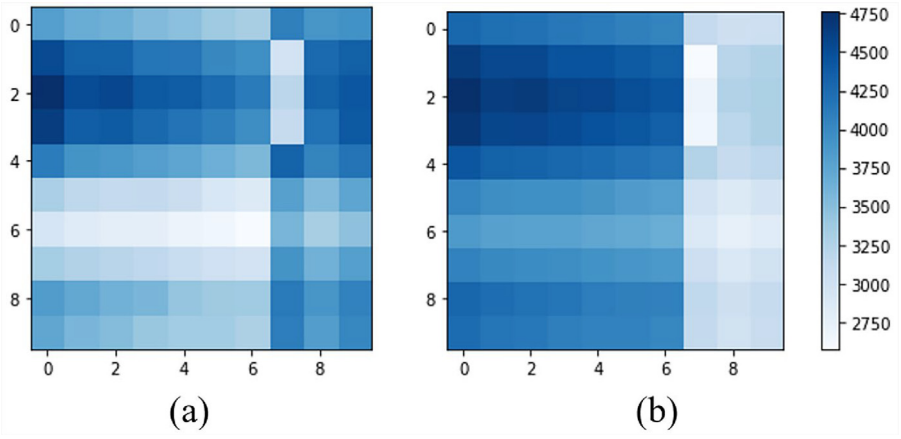


Fig. 6. (a) is the F data from Eq. (4) sorted descending (FD). (b) The new subset data from the main image x, its features indexes match FD's features indexes.

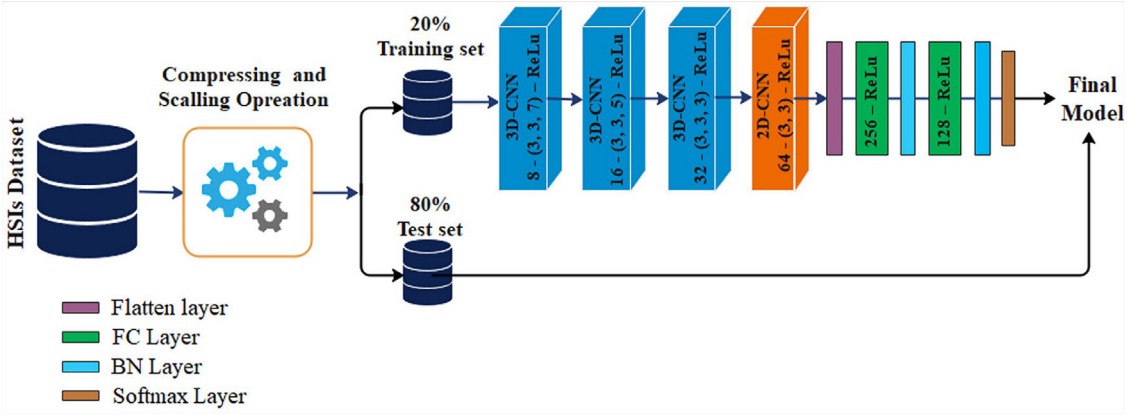


Fig. 7. The main framework of the study.

simultaneously, reducing the network's computational load, so it is faster than other activation functions. RAF works to prevent the problem of vanishing gradient. It is also commonly utilized in deep networks [21,59,60].

3D-CNN layers extract the feature maps of spectral–spatial information, and 2D-CNN enhances the spatial features. Hence to avoid the overfitting caused by deep learning, this study used batch normalization (BN) after each fully connected layer (FC), and also RAF was used in each FC layer [60–62], as indicated in Fig. 7 and Table 2. In addition, because CNN minimizes the size of the generated feature map, zero-padding was employed for the input data to avoid losing the edge data.

4. Experimental

The proposed model in this study and preprocessing operation were used to train three commonly used hyperspectral image datasets: the Indian Pines, Pavia-University, and Kennedy Space Center.

4.1. Datasets

The Indian Pines (IPs) dataset, which contains 16 classes, was collected by the AVIRIS sensor in Northwest India. The spatial resolution is 20 m/p with 145 × 145 spatial dimensions. The IPs

Table 2
The structure of the multi-hybrid deep learning (MHDL) model and its layers.

Layer	Input	Kernel-Size	Activation	Output
3D-CNN	25×25×BNo[a]×1	(3, 3, 7)	ReLu	23×23×24×8
3D-CNN	23×23×24×8	(3, 3, 5)	ReLu	21×21×20×16
3D-CNN	21×21×20×16	(3, 3, 3)	ReLu	19×19×18×32
Reshape	19×19×18×32	–	–	19×19×576
2D-CNN	19×19×576	(3,3)	ReLu	17×17×64
Flatten	17×17×64	–	–	18496
Dense	18496	–	ReLu	256
BN	256	–	–	256
Dense	256	–	ReLu	128
BN	128	–	–	128
Dense	128	–	Softmax	Classes number

[a]BNo means the number of input data bands.

Table 3
The classes of each dataset and their names and samples number.

#	IPs dataset		Pavia-University dataset		KS Center dataset	
	Classes	Samples	Classes	Samples	Classes	Samples
1	Alfalfa	46	Asphalt	6631	Scrub	761
2	Corn-no-till (CN)	1428	Meadows	18649	Willow swamp (WS)	243
3	Corn-mintill (CM)	830	Gravel	2099	Cabbage palm hammock (CPH)	256
4	Corn	237	Trees	3064	Cabbage palm/oak hammock (CPOH)	252
5	Grass-pasture (GP)	483	Painted metal sheets (BMS)	1345	Slash pine (SP)	161
6	Grass-trees (GT)	730	Bare Soil (BS)	5029	Oak/broadleaf hammock (OBH)	229
7	Grass-pasture-mowed (GPM)	28	Bitumen	1330	Hardwood swamp (HS)	105
8	Hay-windrowed (HW)	478	Self-Blocking Bricks (SBB)	3682	Graminoid marsh (GM)	431
9	Oats	20	Shadows	947	Spartina marsh (SM)	520
10	Soybean-no-till (SN)	972			Cattail marsh (CM)	404
11	Soybean-mintill (SM)	2455			Salt marsh (SM)	419
12	Soybean-clean (SC)	593			Mudflats (MF)	503
13	Wheat	205			Water	927
14	Woods	1265				
15	Buildings-Grass-Trees-Drives (BGTD)	386				
16	Stone-Steel-Towers (SST)	93				

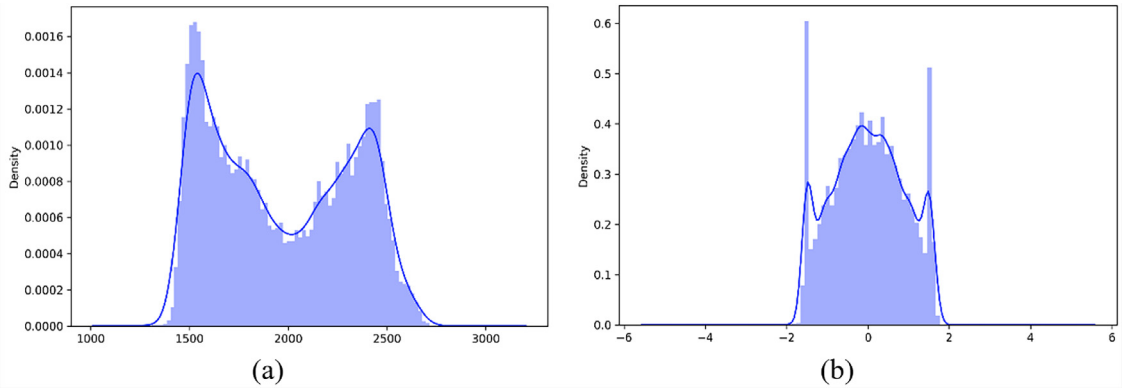


Fig. 8. The data distribution for one band of the IPs dataset, (a) is the data distribution before using Gaussian transformation and (b) is the distribution after using Gaussian transformation.

dataset has 220 spectral bands, then decreased to 200 after eliminating bands encompassing the water absorption zone, spanning from 0.4 μm to 2.5 μm, as illustrated in Table 3.

Pavia-University is the second dataset gathered in Northern Italy using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor and includes nine urban land-cover categories with 610 × 340 spatial dimensions. The spatial resolution is 1.3 m/p. There are 103 spectral bands in the original data set, with wavelengths ranging from 0.43 to 0.86 μm, Table 3.

The Kennedy Space Center (KS Center) data set was obtained over KS Center, Florida, by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It has 13 classes, a resolution of 18 m/p, a size of 512 × 614 pixels, and 176 spectral bands. All the datasets are at this link (https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

4.2. Empirical results and compares

This part explains the output of CRV and four preprocessing methods of DRM and compares them with CRV. It also explains the execution of multi-hybrid deep learning (MHDL) and its comparison with other models. Keras python was used to design the purpose model. The GPU with 26 GB RAM was utilized to run all the experiments.

4.2.1. Compression and reinforced variation results

As reported, the HSIs come with a large number of bands (features) with high spectral resolution and repetition. That brings significant challenges in data transfer, storage, and analysis. Reducing and selecting the crucial features are necessary steps

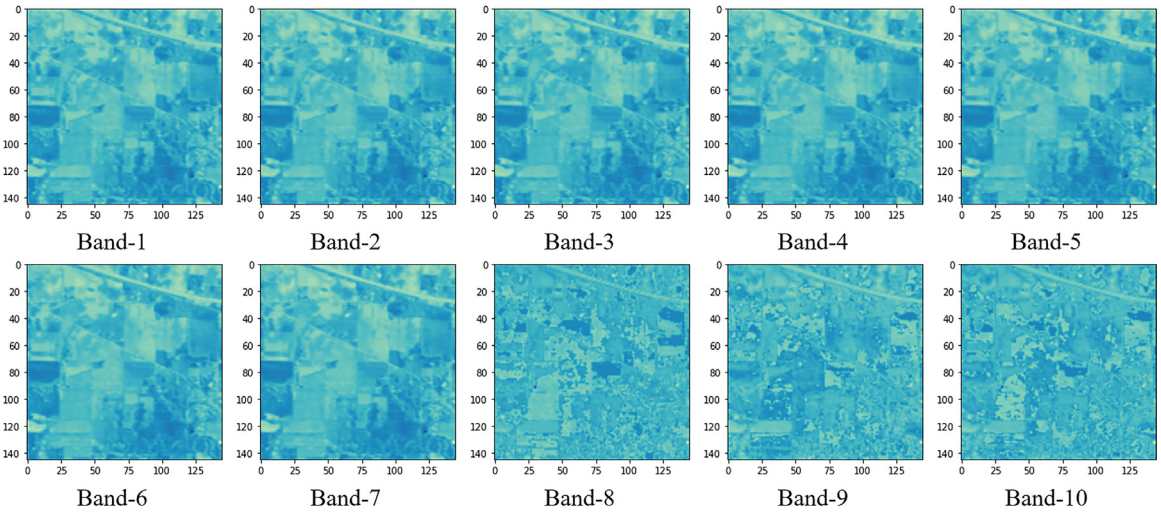


Fig. 9. The first ten bands of compression and reinforced variation (CRV) preprocessing for the IPs dataset.

in the preprocessing stage. Choosing features should facilitate learning and enhance the accuracy as much as possible.

In the multi-class HS dataset, it can be noted that the classes mostly share the same values, especially the smallest values or what are called outliers. These values can be outliers for some classes but are not for others, and they play a big role in enhancing the variance. So, the compression and reinforced variation (CRV) method reinforce the variance by reducing the distance between the big and small values in each class. That helped to detect the main and substantial values from the outliers in each class. After reducing the gap, CRV can choose the informative features after omitting all the outlier and unnecessary values. Fig. 1 illustrates and summarizes all of the CRV's operations and phases.

Fig. 8 shows a big difference between the distribution of the dataset before and after using GT. The output of GT is stabler and more normalized than the data of , Eq. (7). GCDF does not desert any value, but it considers the repeated and correlated data as the frequency data, giving them more area under a standard curve, improving the classification. Fig. 8(a) shows the data distribution for one band of the input data of the IPs dataset. This band has two or more peaks, and each band of the input data has a different distribution with different skewed. After using CRV, the data distribution was enhanced and became more normal, Fig. 8(b). Fig. 9 shows the first ten bands after using CRV operation for the IPs database.

After applying CRV, the number of bands for the IPs dataset was decreased to 30 bands and 15 bands for Pavia-University and KS Center datasets, voiding the redundant features. Then each training set was fed into the MHDL framework for analysis and classification. Tables 4–9 prove that CRV successfully extracted and compressed the critical characteristics and increased classes' variation.

4.2.2. Comparisons results of different RDMs

CRV's effectiveness is evaluated by comparing it to other DRMs. These methods are independent component analysis (ICA) [20,31], factor analysis (FA) [29,63], minimum-redundancy maximum-relevance (mRMR), and principal component analysis (PCA) [20]; all of them are linear unsupervised reduction methods. Independent component analysis (ICA) generates new statistically independent features by reducing the higher-order dependencies in a given dataset [64]. ICA searches for the features that are non-Gaussian and statistically independent. Factor Analysis (FA) is a technique for uncovering correlations between latent and manifest variables, similar to PCA. It describes the variance in multi-class datasets, which helps explain data by reducing the number of bands and enhancing the variance by dropping the factors of the lowest variance [64]. The minimum-redundancy maximum-relevance (mRMR) technique is one of the selection feature algorithms, which increases individual feature reliance while minimizing redundancy between any two features. The redundancy measure in mRMR is not reliant on the classification problem [14,29,30]. PCA determines the highest variance of the data to produce the new features, called principal components (PCs). The orthogonal axes are the directions of the largest data variance in PCA, which projects the highly dimensional dataset to a new subspace. The first PC has the highest variance during the transformation, and the subsequent PCs have decreasing variances [30,31].

In this part, all the above methods were used as preprocessing and then fed their output to the proposed model, multi-hybrid deep learning (MHDL). The output of mRMR was standardized by StandardScaler of the scikit-learn. The whitening operation is applied in PCA and ICA. The number of components (features) for all these preprocessing methods is 30 for the IPs dataset and 15 for the others. Thus, Tables 4–6 represent the results of MHDL models for ICA, FA, mRMR, PCA, and CRV preprocessing of the three datasets.

FA and CRV work to choose the most informative features and normalize them; this help to get very high accuracy. mRMR has a very strong way of categorizing the data. It prefers to choose a subset of attributes with the highest correlation within a class and the lowest correlation with other classes. PCA obtains the features mean to get their variations. All of these make FA, mRMR, PCA, and CRV obtain the highest accuracy, and they were almost the same because they enhanced the variance between the classes, as seen in Tables 4 and 5.

The KS Center dataset is the most complicated, and because ICA did not encourage the distribution of the data and develop independent components, it could not achieve any accuracy with the IPs dataset. In the KS_Center dataset, FA and CRV with MHDL got the best results, as shown in Table 6. With the IPs dataset, CRV acquired the lowest testing time, then PCA. While, with the KS _Center dataset, the CRV method gave the lowest testing time after the mRMR.

The performance time of the learning model is not the only one that can be observed and calculated; there is an exhaustive time in the preprocessing stage that should be calculated. Fig. 10 shows the execution time of the preprocessing of all methods for

Table 4
The comparison of the five dimensionality reduction methods for the first dataset (IPs).

#	Classes	Training samples	Tasting samples	ICA	FA	mRMR	PCA	CRV
1	Alfalfa	9	37	0.973	1	1	1	1
2	CN	285	1143	0.972	0.981	0.970	0.973	0.979
3	CM	166	664	0.976	1	0.991	1	1
4	Corn	47	190	1	1	0.958	1	1
5	GP	97	386	1	0.997	0.997	0.990	1
6	GT	146	584	0.995	0.998	0.997	0.998	0.998
7	GPM	6	22	1	1	0.818	1	1
8	HW	96	382	1	1	0.995	1	1
9	Oats	4	16	1	0.938	0.875	1	0.75
10	SN	194	778	1	0.990	0.982	0.996	0.999
11	SM	491	1964	0.998	0.998	0.996	0.998	0.997
12	SC	118	475	0.975	0.979	0.96	0.989	0.968
13	Wheat	41	164	1	1	1	1	1
14	Woods	253	1012	0.999	1	0.999	1	1
15	BGTD	77	309	1	1	0.974	1	0.997
16	SST	19	74	0.932	0.973	0.865	0.973	0.878
	Kappa accuracy (%)			98.971	99.332	98.358	99.319	99.165
	Overall accuracy (%)			99.098	99.415	98.561	99.402	99.268
	Average accuracy (%)			98.874	99.088	96.109	99.487	97.294
	Training time (S)			130.140	142.919	129.882	142.923	143.139
	Testing time (S)			2.347	3.327	3.362	2.311	2.223

Table 5
The compare results of the five dimensionality reduction methods for the second dataset (Pavia-university).

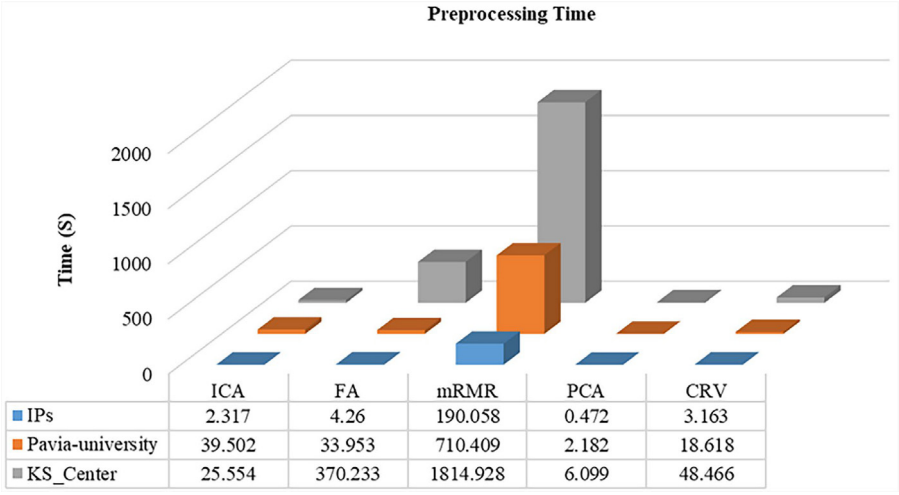
#	Classes	Training samples	Tasting samples	ICA	FA	mRMR	PCA	CRV
1	Asphalt	1326	5305	1	0.999	0.999	1	1
2	Meadows	3730	14919	1	1	1	1	1
3	Gravel	420	1679	0.996	1	1	1	0.998
4	Trees	613	2451	0.994	0.999	1	0.997	1
5	BMS	269	1076	1	1	1	1	1
6	BS	1006	4023	1	1	1	1	1
7	Bitumen	266	1064	1	1	1	1	1
8	SBB	736	2946	0.998	1	0.999	0.999	0.997
9	Shadows	189	758	0.993	0.992	1	0.996	1
	Kappa accuracy (%)			99.868	99.945	99.969	99.950	99.954
	Overall accuracy (%)			99.901	99.962	99.977	99.962	99.965
	Average accuracy (%)			99.786	99.888	99.977	99.913	99.946
	Training time (S)			263.102	263.37	264.303	264.739	203.782
	Testing time (S)			4.798	4.742	3.936	5.248	4.001

Table 6
The compare results of the five dimensionality reduction methods for the KS_Center dataset.

#	Classes	Training samples	Tasting samples	ICA	FA	mRMR	PCA	CRV
1	Scrub	152	609	0	1	1	0.998	1
2	WS	49	194	0	1	0.948	0.923	0.974
3	CPH	51	205	0	1	1	0.956	0.941
4	CPOH	50	202	0	0.995	0.881	0.703	0.955
5	SP	32	129	0	0.977	0.876	0.891	1
6	OBH	46	183	0	1	1	0.973	0.995
7	HS	21	84	0	1	1	0.893	1
8	GM	86	345	1	1	1	0.977	1
9	SM	104	416	0	1	1	1	1
10	CM	81	323	0	1	0.910	0.926	1
11	SM	84	335	0	1	1	1	1
12	MF	101	402	0	1	0.990	0.930	1
13	Water	185	742	0	1	1	1	1
	Kappa accuracy (%)			0.0	99.893	97.782	95.376	99.279
	Overall accuracy (%)			8.275	99.904	98.009	95.850	99.352
	Average accuracy (%)			7.692	99.783	96.968	93.615	98.967
	Training time (S)			83.351	80.030	69.913	85.660	33.075
	Testing time (S)			0.6	0.597	0.560	0.629	0.593

the three datasets. Evidently, the PCA had the quickest time, but the CRV provided the best time performance with good accuracy across all datasets. Fig. 10 further shows that, in contrast to the others, the CRV delivered an exceptional smooth training operation.

As known, preprocessing and deep learning classification models play a considerable role in enhancing spectral–spatial feature extraction accuracy of HSI. Most studies used PCA to reduce dimensionality, nonetheless, it does not classify patterns by itself. As a result, it merely chooses features with the most extensive



	ICA	FA	mRMR	PCA	CRV
IPs	2.317	4.26	190.058	0.472	3.163
Pavia-university	39.502	33.953	710.409	2.182	18.618
KS_Center	25.554	370.233	1814.928	6.099	48.466

Fig. 10. The time of the five preprocessing and reducing dimensionality methods.

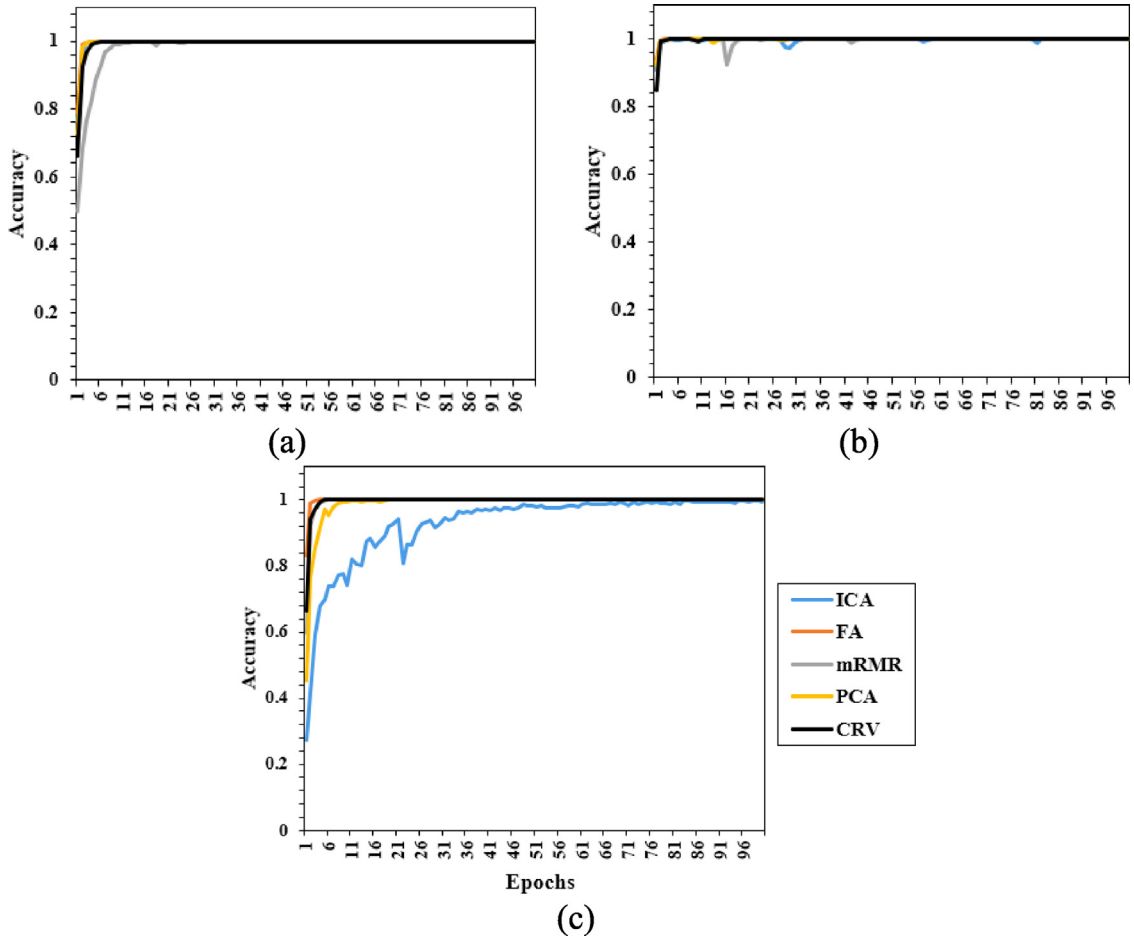


Fig. 11. The accuracy values of each epoch during the training time. (a) is for the first dataset, (b) represents the second dataset, and (c) represents the accuracy of the third dataset.

variability that does not guarantee to perform well in classifications. Furthermore, the FSAs are expensive due to computational complexity and the time of execution, especially with a complex data distribution dataset. Also, the differences in the number of samples of classes in each dataset are computational complexity to simultaneously enhance the detection of the small and large number of samples. In general, notwithstanding the difficulties it encounters, the CRV enhanced the data distribution, choosing the most effective features and performance time simultaneously.

The compression and reinforced variation (CRV) method with multi-hybrid deep learning (MHDL) model were tested with different ranges of split for dataset to see their ability. In the first test, the output of CRV for the three datasets was split into 10% as a training set and 90% as a testing set before feeding them into the MHDL classification model. The second experiment was run

Table 7
Test the compression and reinforced variation (CRV) method with the multi-hybrid deep learning (MHDL) model on different split rates of the three used HSI.

/Dataset	IPs			Pavia-university			KS_Center		
Training set rate	10%	20%	30%	10%	20%	30%	10%	20%	30%
Kappa accuracy (%)	95.33	99.17	99.76	99.31	99.95	99.96	96.74	99.28	99.48
Overall accuracy (%)	95.91	99.27	99.79	99.08	99.97	99.97	97.08	99.35	99.53
Average accuracy (%)	91.29	97.29	98.83	99.31	99.95	99.97	95.22	98.97	99.23
Training time (S)	83.32	143.14	203.47	100.24	203.78	323.37	22.40	33.08	45.13
Testing time (S)	2.70	2.22	2.12	4.54	4	3.81	0.68	0.59	0.58

with 20% as a training set and 80% as testing set for all datasets. The final experiment was to test the output of CRV+ MHDL by splitting 30% of the data as a training set and 70% as a testing set. As seen in Table 7, the accuracy smoothly increased by increasing the size of the training set.

4.2.3. MHDL results and comparisons

The output of CRV was fed into three sequence layers of 3D-CNN to extract the special–spectral feature maps. Then into one layer of 2D-CNN to enhance the localization of the extraction features. Because the sample numbers and dimension values of the classes in each dataset are not the same, different kernel sizes were used in each layer of the MHDL model. These differences in kernel size help to cover and extract more features. Furthermore, using BN helps to normalize, stable the training, and void overfitting during the training operation [65], Fig. 11.

This study trained and tested the MHDL model with CRV in three different datasets. These three datasets are characterized by having different numbers of classes and bands and data distribution, so this study tested the ability of the MHDL framework with CRV on different challenges of HSI. From the results of the three datasets in Tables 8–10, it can be noted that the MHDL model successfully classified the input data of CRV operation for all datasets. Therefore, this study successfully compressed, extracted, and classified the different HSI.

Also, to quantitatively evaluate the proposed MHDL framework with CRV, this study was compared with seven existing methods: multilayer perceptron (MLP), CNN1d, CNN2d, CNN3d, VGG-16 [66], HybridCNN [60], HybridSN [67].

MLP is one layer of FC with RAF, and CNN1d is one 1D-CNN layer with RAF, 24 KS, two FC layers, Max-Pooling layer (MP) (2, 2), and BN. MLP and CNN1d are used to train spectral information [68]. CNN2d includes two layers of 2D-CNN, (5,5) KS, one MP (2, 2), and one FC with 100 units. CNN2d model focuses on spatial information. CNN3D is two layers of 3D-CNN ((5, 5, 24, 32), (5, 5, 7, 64)) and one layer of FC (300 units) with BN and RAF to extract and train spectral–spatial information. In MLP, CNN1d, CNN2d, and CNN3d, each dataset was minimized to 30 bands by PCA and split into 20% as training and 80% as testing sets [23].

The VGG-16 model has 14 2D-CNN layers with the same kernel size (KS) (3×3), two FC layers, both use RAF, and four layers of MP layer (2, 2); it is a deep model. The input dataset is minimized to 30 bands for the IPs dataset and 15 bands for Pavia-University and KS_Center datasets. The rates of training and testing set are 20% and 80%. The HybridCNN framework trains three parallel networks of 3D-CNN with different window sizes (15, 13, 9). The input data is diminished to 15 bands for all datasets and divided into 20%, 10%, and 70% for training, validation, and testing sets. HybridSN model is a multi-size kernel for hybrid layers of 3D-CNN and 2D-CNN with two dropout layers to avoid overfitting. In HybridSN, the number of bands is 30 for the IPs and 15 for the others. In all these compared models, the input data was decreased by PCA.

As can be seen from Tables 8–10, MLP, CNN1d, and CNN2d got less accuracy because they do not train the spectral–spatial data. In the Pavia-University dataset, the CNN3D model is only 0.04% higher than the MHDL. In Table 10, although the VGG-16 model has many 2D-CNN layers, its accuracy is low since the data distribution in the KS_Center dataset is the most complex, and the spectral information of many classes is similar. Because the HybridCNN and HybridSN deal with spectral–spatial data, they were better than the previous models in all datasets evaluated. On the other hand, this study framework received the highest accuracy because it was the best to make the distribution normal-like and enhanced the localization of spectral–spatial data feature extraction. To sum up, the proposed CRV-MHDL model obtained the best results for all data sets.

In addition, using CRV to reduce the input data and enhance the variation, compared to the other models' results, the MHDL model obtained high accuracy for the small sample numbers in all used datasets, as shown in Tables 8–10. The CRV operation enhanced classes' variation and chose the crucial bands. The window size is 25 for all datasets to consider more local similarity. The big size of the window may lead to producing some noise and may cause less classification capability. The CRV worked previously to enhance the variance and to reduce the noise, so that, the window size does not affect the accuracy, as shown in the final results of CRV-MHDL in Tables 8–10 and Figs. 12–14.

Fig. 15 shows the situation of training accuracy for used models. As can be seen, the most smooth and stable model in the training time is MHDL.

5. Conclusion

The HSI dataset has a large number of classes and bands; therefore, these classes typically share the same values across all classes. This study proposed a novel feature selection method called compression and reinforced variation (CRV) to reduce the dimension of HSI. Furthermore, the structure learning model of this study, multi-hybrid deep learning (MHDL), enhanced the extraction of spectral–spatial features by using hybrid layers of CNN and kernel size, and it provided more stable results than others. MHDL with CRV was the best to classify and extract the best accuracy for the small samples. The experiments of the CRV method with MHDL show the following: (1) The MHDL model with CRV provided the most steady performance and the best classification for the small sample number of classes. (2) The CRV+MHDL succeeded in extracting multi-class HSI datasets. (3) The CRV improved all HSI data distributions to have a normal distribution. It also reduced the performance time and the data dimensionality. (4) The CRV successfully compressed the input data and chose only the crucial features that enhance the variation between the different classes. (5) The CRV introduced the best running time with high accuracy compared to the other preprocessing methods. (6) The MHDL successfully extracted the spectral–spatial features for HSI and produced the optimum classification results for the three datasets. (7) The CRV operation increased the accuracy of MHDL to be higher than other models that used PCA to represent and compress data. The future work will focus on reducing the execution time of feature selection operations.

Table 8
The comparison of the seven models and compression and reinforced variation (CRV) + multi-hybrid deep learning model (MHDL) for the first dataset.

#	Classes	Training samples	Tasting samples	MLP	CNN1d	CNN2d	CNN3d	VGG-16	Hybrid-CNN	Hybrid SN	CRV + MHDL
1	Alfalfa	9	37	0.462	0.108	0.026	0.757	0	0.875	0.973	1
2	CN	285	1143	0.626	0.636	0.411	0.904	0.764	0.983	0.979	0.979
3	CM	166	664	0.177	0.291	0.359	0.928	0.937	1	0.985	1
4	Corn	47	190	0.179	0.126	0.313	0.916	0	0.952	0.995	1
5	GP	97	386	0.64	0.790	0.192	0.956	0.896	0.970	0.995	1
6	GT	146	584	0.820	0.937	0.931	0.973	0.954	0.984	0.997	0.998
7	GPM	6	22	0	0.739	0.292	0.913	0	0.85	1	1
8	HW	96	382	0.845	0.984	0.953	1	0.950	1	1	1
9	Oats	4	16	0.294	0	0.118	0.938	0	1	1	0.75
10	SN	194	778	0.308	0.569	0.173	0.969	0.221	0.975	0.991	0.999
11	SM	491	1964	0.620	0.851	0.774	0.980	0.951	0.997	0.990	0.997
12	SC	118	475	0.232	0.420	0.109	0.968	0.909	0.935	0.967	0.968
13	Wheat	41	164	0.730	0.866	0.477	0.994	0.994	1	1	1
14	Woods	253	1012	0.916	0.956	0.940	0.998	0.826	0.999	0.999	1
15	BGTD	77	309	0.220	0.453	0.527	0.867	0.142	1	0.994	0.997
16	SST	19	74	0.823	0.797	0	0.973	0	1	0.973	0.878
	Kappa accuracy (%)			50.681	66.244	50.003	95.223	72.908	98.411	98.777	99.165
	Overall accuracy (%)			57.089	70.899	56.813	95.817	76.524	98.606	98.927	99.268
	Average accuracy (%)			49.316	59.518	41.219	93.956	53.403	97.003	98.973	97.294

Table 9
The comparison of the seven models and compression and reinforced variation (CRV) + multi-hybrid deep learning model (MHDL) for the second dataset.

#	Classes	Training samples	Tasting samples	MLP	CNN1d	CNN2d	CNN3d	VGG-16	Hybrid-CNN	Hybrid SN	CRV + MHDL
1	Asphalt	1326	5305	0.906	0.954	0.988	1	1	1	1	1
2	Meadows	3730	14919	0.960	0.981	0.997	1	1	1	1	1
3	Gravel	420	1679	0.575	0.804	0.932	1	1	0.999	1	0.998
4	Trees	613	2451	0.900	0.936	0.992	1	0.986	0.999	0.991	1
5	BMS	269	1076	0.977	0.989	0.999	1	0.999	1	0.998	1
6	BS	1006	4023	0.819	0.909	0.996	1	1	1	1	1
7	Bitumen	266	1064	0.843	0.914	0.943	1	1	0.992	1	1
8	SBB	736	2946	0.883	0.891	0.950	1	0.997	0.983	0.996	0.997
9	Shadows	189	758	0.993	0.999	0.983	1	0.956	1	0.996	1
	Kappa accuracy (%)			87.083	92.983	98.142	99.996	99.679	99.752	99.853	99.954
	Overall accuracy (%)			90.286	94.719	98.597	99.997	99.757	99.81	99.889	99.965
	Average accuracy (%)			87.286	93.078	97.547	99.995	99.312	99.697	99.788	99.946

Table 10
The comparison of the seven models and compression and reinforced variation (CRV)-multi-hybrid deep learning model (MHDL) for the KS_Center dataset.

#	Classes	Training samples	Tasting samples	MLP	CNN1d	CNN2d	CNN3d	VGG-16	Hybrid-CNN	Hybrid SN	CRV + MHDL
1	Scrub	152	609	0.836	0.966	0.836	0	0	0.985	0.992	1
2	WS	49	194	0.251	0.108	0.256	0.572	0.562	0.9	0.933	0.974
3	CPH	51	205	0.312	0	0.028	0.098	0	0.888	0.941	0.941
4	CPOH	50	202	0	0	0.103	0.248	0.658	0.642	0.916	0.955
5	SP	32	129	0.504	0	0.555	0.512	0.977	0.805	0.946	1
6	OBH	46	183	0.256	0	0.108	0.191	0	0.525	0.973	0.995
7	HS	21	84	0	0	0.067	0	0	0.946	0.869	1
8	GM	86	345	0.380	0.528	0.385	0.974	0	0.858	0.962	1
9	SM	104	416	0.543	0.788	0.224	0.538	0	0.538	0.998	1
10	CM	81	323	0.163	0.130	0.318	0.963	1	0.912	0.966	1
11	SM	84	335	0.854	0.809	0.607	0.994	0.976	1	1	1
12	MF	101	402	0.532	0	0.642	0.811	0.980	0.966	0.970	1
13	Water	185	742	0.985	0.996	0.768	1	1	1	1	1
	Kappa accuracy (%)			51.637	44.721	42.270	56.889	43.861	85.984	97.141	99.279
	Overall accuracy (%)			56.943	52.063	48.973	61.252	51.667	87.445	97.433	99.352
	Average accuracy (%)			43.198	33.267	37.658	53.079	47.333	84.348	95.889	98.967

CRediT authorship contribution statement

Dalal AL-Alimi: Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Zhihua Cai:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **Mohammed A.A. Al-qaness:** Conceptualization, Formal analysis, Project administration, Validation, Writing – review & editing. **Abdelghani Dahou:** Formal analysis, Validation, Writing – original draft. **Eman Ahmed Alawamy:** Formal analysis, Validation, Writing – review & editing. **Sakinatu Issaka:** Validation, Writing – review & editing.

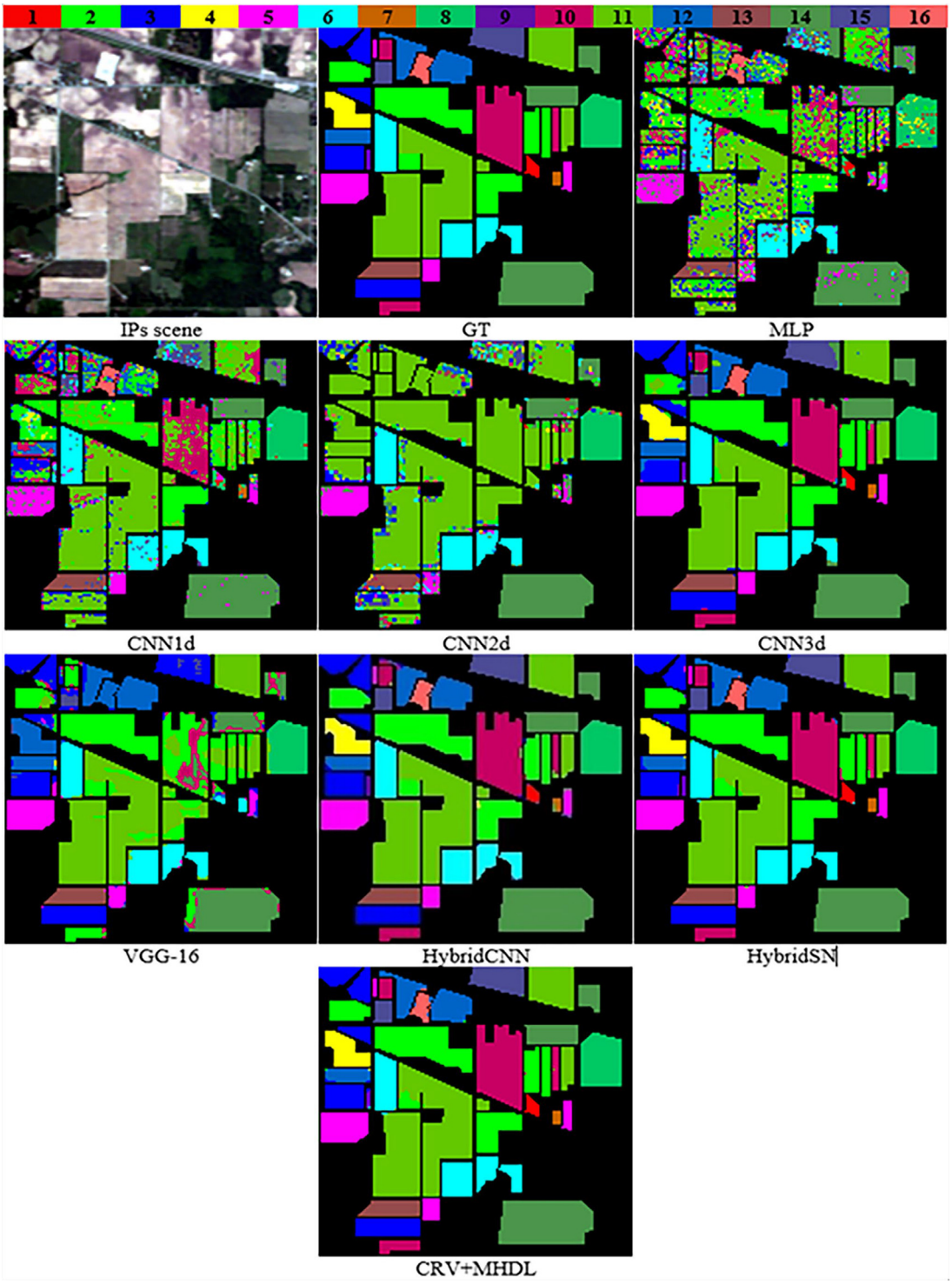


Fig. 12. The output of the different models and compression and reinforced variation (CRV) + multi-hybrid deep learning model (MHDL) for the IPs dataset.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62150410434).

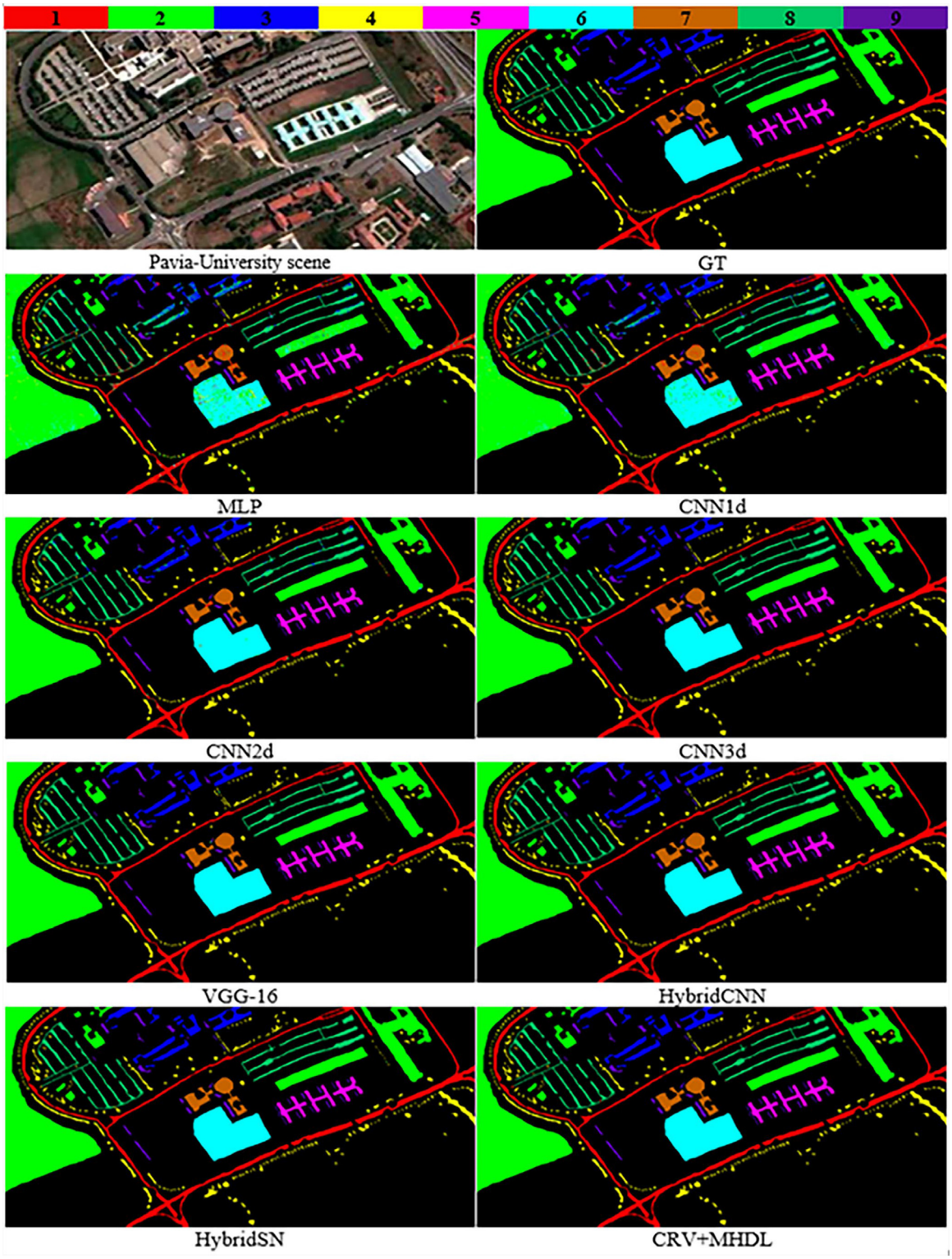


Fig. 13. The output of the different models and compression and reinforced variation (CRV) -multi-hybrid deep learning model (MHDL) for the Pavia-university dataset.

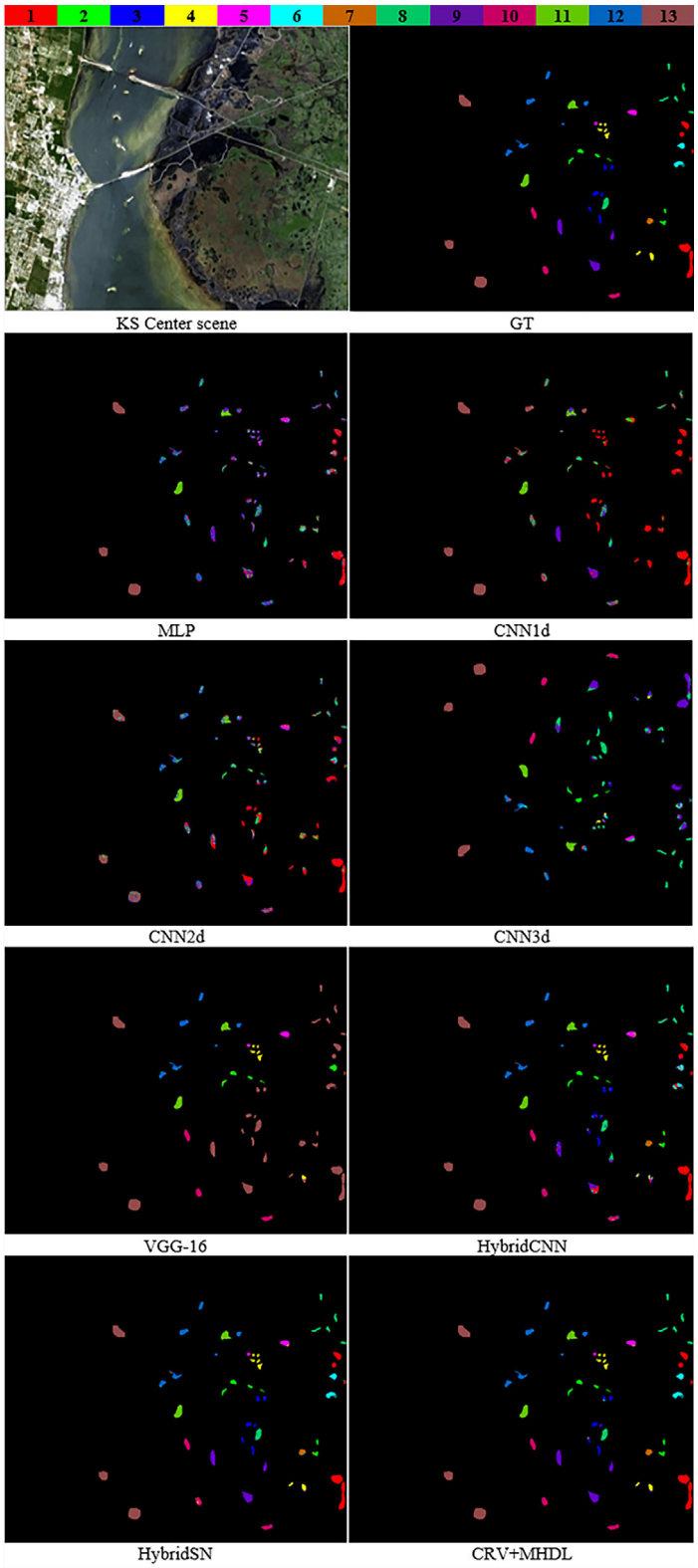


Fig. 14. The output of the different models and compression and reinforced variation (CRV)-multi-hybrid deep learning model (MHDL) for the third dataset.

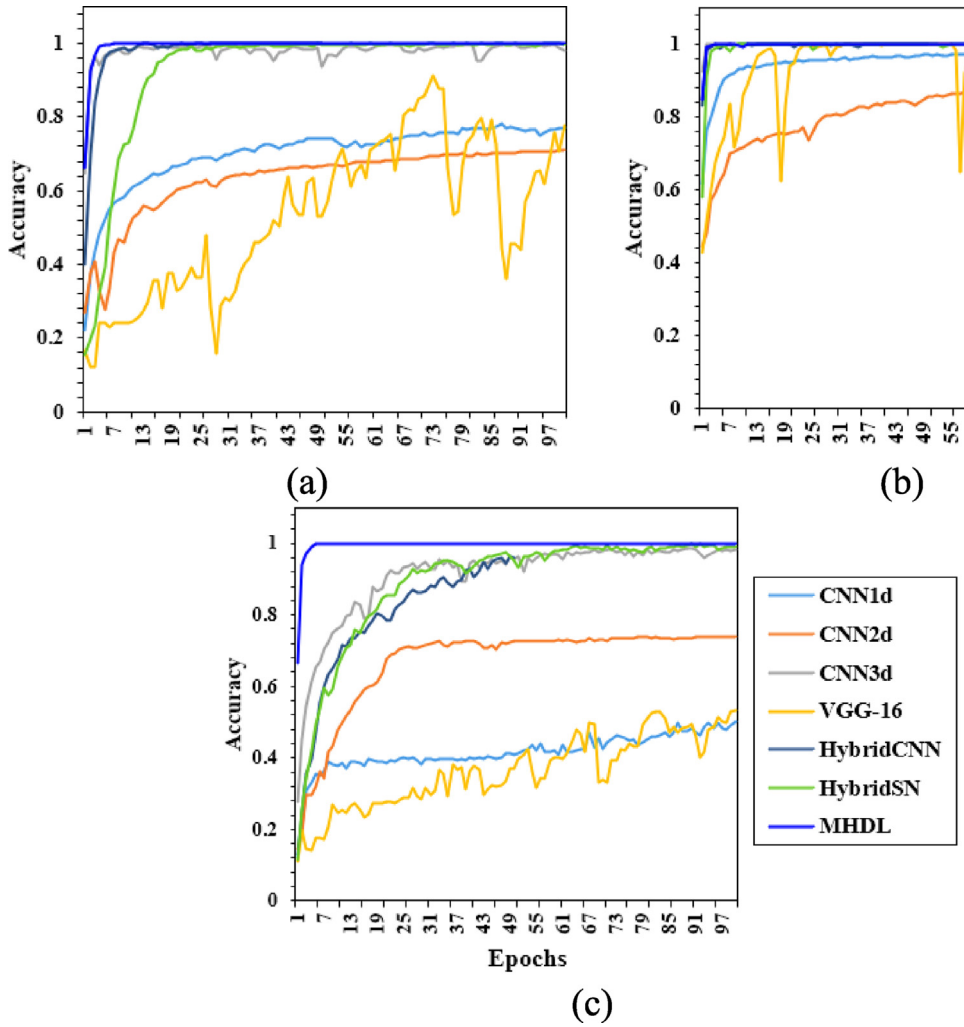


Fig. 15. The accuracy values of each epoch during the training time for the eight models. (a) represents the accuracy of the first dataset (IPs), (b) is the accuracy values of the second dataset, and (c) represents the training accuracy of the last dataset.

References

[1] J. Zheng, Y. Feng, C. Bai, J. Zhang, Hyperspectral image classification using mixed convolutions and covariance pooling, IEEE Trans. Geosci. Remote Sens. 59 (2021) 522–534, http://dx.doi.org/10.1109/TGRS.2020.2995575.

[2] C. Camino, V. González-Dugo, P. Hernández, J.C. Sillero, P.J. Zarco-Tejada, Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture, Int. J. Appl. Earth Obs. Geoinf. 70 (2018) 105–117, http://dx.doi.org/10.1016/j.jag.2018.04.013.

[3] R. Nigam, B.K. Bhattacharya, R. Kot, C. Chattopadhyay, Wheat blast detection and assessment combining ground-based hyperspectral and satellite based multispectral data, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII-3/W6 (2019) 473–475, http://dx.doi.org/10.5194/isprs-archives-XLII-3-W6-473-2019.

[4] M. Shimoni, R. Haelterman, C. Perneel, Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques, IEEE Geosci. Remote Sens. Mag. 7 (2019) 101–117, http://dx.doi.org/10.1109/MGRS.2019.2902525.

[5] M. Imani, H. Ghassemian, An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges, Inf. Fusion. 59 (2020) 59–83, http://dx.doi.org/10.1016/j.inffus.2020.01.007.

[6] R.J. Murphy, B. Whelan, A. Chlingaryan, S. Sukkarieh, Quantifying leaf-scale variations in water absorption in lettuce from hyperspectral imagery: a laboratory study with implications for measuring leaf water content in the context of precision agriculture, Precis. Agric. 20 (2019) 767–787, http://dx.doi.org/10.1007/s11119-018-9610-5.

[7] L. Yan, N. Noro, Y. Takara, F. Ando, M. Yamaguchi, Using hyperspectral image enhancement method for small size object detection on the sea surface, in: L. Bruzzone (Ed.), Proc.SPIE, 2015, 96430H, http://dx.doi.org/10.1117/12.2194606.

[8] C. Lin, S.-Y. Chen, C.-C. Chen, C.-H. Tai, Detecting newly grown tree leaves from unmanned-aerial-vehicle images using hyperspectral target detection techniques, ISPRS J. Photogramm. Remote Sens. 142 (2018) 174–189, http://dx.doi.org/10.1016/j.isprsjprs.2018.05.022.

[9] C. Jiao, C. Chen, R.G. McGarvey, S. Bohlman, L. Jiao, A. Zare, Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection, ISPRS J. Photogramm. Remote Sens. 146 (2018) 235–250, http://dx.doi.org/10.1016/j.isprsjprs.2018.08.012.

[10] N. Hussain, D.-W. Sun, H. Pu, Classical and emerging non-destructive technologies for safety and quality evaluation of cereals: A review of recent applications, Trends Food Sci. Technol. 91 (2019) 598–608, http://dx.doi.org/10.1016/j.tifs.2019.07.018.

[11] B. Fei, Hyperspectral imaging in medical applications, in: J.M. Amigo (Ed.), Data Handling in Science and Technology, Elsevier, 2020, pp. 523–565, http://dx.doi.org/10.1016/B978-0-444-63977-6.00021-3.

[12] M.A. Calin, S.V. Parasca, D. Savastru, D. Manea, Hyperspectral imaging in the medical field: Present and future, Appl. Spectrosc. Rev. 49 (2014) 435–447, http://dx.doi.org/10.1080/05704928.2013.838678.

[13] C. Tang, X. Liu, X. Zhu, J. Xiong, M. Li, J. Xia, X. Wang, L. Wang, Feature selective projection with low-rank embedding and dual Laplacian regularization, IEEE Trans. Knowl. Data Eng. 32 (2019) 1, http://dx.doi.org/10.1109/TKDE.2019.2911946.

[14] J. Feng, J. Chen, Q. Sun, R. Shang, X. Cao, X. Zhang, L. Jiao, Convolutional neural network based on bandwise-independent convolution and hard thresholding for hyperspectral band selection, IEEE Trans. Cybern. 51 (2021) 4414–4428, http://dx.doi.org/10.1109/TCYB.2020.3000725.

[15] C. Tang, M. Bian, X. Liu, M. Li, H. Zhou, P. Wang, H. Yin, Unsupervised feature selection via latent representation learning and manifold regularization, Neural Netw. 117 (2019) 163–178, http://dx.doi.org/10.1016/j.neunet.2019.04.015.

[16] B. Xu, X. Li, W. Hou, Y. Wang, Y. Wei, A similarity-based ranking method for hyperspectral band selection, IEEE Trans. Geosci. Remote Sens. 59 (2021) 9585–9599, http://dx.doi.org/10.1109/TGRS.2020.3048138.

[17] J. Feng, D. Li, J. Gu, X. Cao, R. Shang, X. Zhang, L. Jiao, Deep reinforcement learning for semisupervised hyperspectral band selection, IEEE Trans. Geosci. Remote Sens. (2021) 1–19, http://dx.doi.org/10.1109/TGRS.2021.3049372.

[18] J. Wang, C. Tang, Z. Li, X. Liu, W. Zhang, E. Zhu, L. Wang, Hyperspectral band selection via region-aware latent features fusion based clustering, Inf. Fusion 79 (2022) 162–173, http://dx.doi.org/10.1016/j.inffus.2021.09.019.

[19] B. Ghojogh, M.N. Samad, S.A. Mashhadi, T. Kapoor, W. Ali, F. Karray, M. Crowley, Feature selection and feature extraction in pattern analysis: A literature review, 2019, ArXiv E-Prints. arXiv:1905.02845. https://ui.adsabs.harvard.edu/abs/2019arXiv190502845G.

[20] F. Anowar, S. Sadaoui, B. Selim, Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE), Comput. Sci. Rev. 40 (2021) 100378, http://dx.doi.org/10.1016/j.cosrev.2021.100378.

[21] D. Saha, A. Manickavasagan, Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review, Curr. Res. Food Sci. 4 (2021) 28–44, http://dx.doi.org/10.1016/j.crfs.2021.01.002.

[22] J. Dong, W. Guo, F. Zhao, D. Liu, Discrimination of hayward kiwifruits treated with forchlorfenuron at different concentrations using hyperspectral imaging technology, Food Anal. Methods. 10 (2017) 477–486, http://dx.doi.org/10.1007/s12161-016-0603-6.

[23] M.E. Paoletti, J.M. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: A review, ISPRS J. Photogramm. Remote Sens. 158 (2019) 279–317, http://dx.doi.org/10.1016/j.isprsjprs.2019.09.006.

[24] D. AL-Alimi, M.A.A. Al-qaness, Z. Cai, A. Dahou, Y. Shao, S. Issaka, Meta-learner hybrid models to classify hyperspectral images, Remote Sens. 14 (2022) 1038, http://dx.doi.org/10.3390/rs14041038.

[25] S.R. Delwiche, I.T. Rodriguez, S.R. Rausch, R.A. Graybosch, Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging, J. Cereal Sci. 87 (2019) 18–24, http://dx.doi.org/10.1016/j.jcs.2019.02.008.

[26] C. Xia, S. Yang, M. Huang, Q. Zhu, Y. Guo, J. Qin, Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis, Infrared Phys. Technol. 103 (2019) 103077, http://dx.doi.org/10.1016/j.infrared.2019.103077.

[27] T.V. Bandos, L. Bruzzone, G. Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, IEEE Trans. Geosci. Remote Sens. 47 (2009) 862–873, http://dx.doi.org/10.1109/TGRS.2008.2005729.

[28] L. Zhang, Q. Zhang, B. Du, X. Huang, Y.Y. Tang, D. Tao, Simultaneous spectral-spatial feature selection and extraction for hyperspectral images, IEEE Trans. Cybern. 48 (2018) 16–28, http://dx.doi.org/10.1109/TCYB.2016.2605044.

[29] Hanchuan. Peng, Fuhui. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238, http://dx.doi.org/10.1109/TPAMI.2005.159.

[30] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Sci. Inf. Conf., IEEE, 2014, pp. 372–378, http://dx.doi.org/10.1109/SAI.2014.6918213.

[31] S. Wang, M.E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J.M. Gorriz, I. Tyukin, Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects, Inf. Fusion 76 (2021) 376–421, http://dx.doi.org/10.1016/j.inffus.2021.07.001.

[32] J. Peng, W. Sun, Q. Du, Self-paced joint sparse representation for the classification of hyperspectral images, IEEE Trans. Geosci. Remote Sens. 57 (2019) 1183–1194, http://dx.doi.org/10.1109/TGRS.2018.2865102.

[33] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, X.X. Zhu, X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data, ISPRS J. Photogramm. Remote Sens. 167 (2020) 12–23, http://dx.doi.org/10.1016/j.isprsjprs.2020.06.014.

[34] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, X.X. Zhu, Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (2020) 3791–3808, http://dx.doi.org/10.1109/TGRS.2019.2957251.

[35] Y. Liang, L. Wei, Q. Lu, Multiple feature fusion for fine classification of crops in UAV hyperspectral imagery, in: 2021 IEEE Int. Geosci. Remote Sens. Symp, IGARSS, IEEE, 2021, pp. 5059–5062, http://dx.doi.org/10.1109/IGARSS47720.2021.9553490.

[36] D. Hong, N. Yokoya, X.X. Zhu, Learning a robust local manifold representation for hyperspectral dimensionality reduction, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (2017) 2960–2975, http://dx.doi.org/10.1109/JSTARS.2017.2682189.

[37] D. Hong, N. Yokoya, J. Chanussot, J. Xu, X.X. Zhu, Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction, ISPRS J. Photogramm. Remote Sens. 158 (2019) 35–49, http://dx.doi.org/10.1016/j.isprsjprs.2019.09.008.

[38] Y. Guo, H. Cao, J. Bai, Y. Bai, High efficient deep feature extraction and classification of spectral-spatial hyperspectral image using cross domain convolutional neural networks, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (2019) 1–12, http://dx.doi.org/10.1109/JSTARS.2018.2888808.

[39] M. Ahmad, S. Shabbir, S.K. Roy, D. Hong, X. Wu, J. Yao, A.M. Khan, M. Mazzara, S. Distefano, J. Chanussot, Hyperspectral image classification—Traditional to deep models: A survey for future prospects, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 968–999, http://dx.doi.org/10.1109/JSTARS.2021.3133021.

[40] J. Nalepa, M. Antoniak, M. Myller, P. Ribalta Lorenzo, M. Marcinkiewicz, Towards resource-frugal deep convolutional neural networks for hyperspectral image segmentation, Microprocess. Microsyst. 73 (2020) 102994, http://dx.doi.org/10.1016/j.micpro.2020.102994.

[41] J. Zhang, L. Dai, F. Cheng, Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network, J. Food Meas. Charact. 15 (2021) 484–494, http://dx.doi.org/10.1007/s11694-020-00646-3.

[42] G. Zhao, G. Liu, L. Fang, B. Tu, P. Ghamisi, Multiple convolutional layers fusion framework for hyperspectral image classification, Neurocomputing 339 (2019) 149–160, http://dx.doi.org/10.1016/j.neucom.2019.02.019.

[43] F. Cao, W. Guo, Cascaded dual-scale crossover network for hyperspectral image classification, Knowledge-Based Syst. 189 (2020) 105122, http://dx.doi.org/10.1016/j.knosys.2019.105122.

[44] X. Cao, J. Yao, Z. Xu, D. Meng, Hyperspectral image classification with convolutional neural network and active learning, IEEE Trans. Geosci. Remote Sens. 58 (2020) 4604–4616, http://dx.doi.org/10.1109/TGRS.2020.2964627.

[45] R. Hang, Z. Li, Q. Liu, P. Ghamisi, S.S. Bhattacharyya, Hyperspectral image classification with attention-aided CNNs, IEEE Trans. Geosci. Remote Sens. 59 (2021) 2281–2293, http://dx.doi.org/10.1109/TGRS.2020.3007921.

[46] C. Yu, R. Han, M. Song, C. Liu, C.-I. Chang, A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 2485–2501, http://dx.doi.org/10.1109/JSTARS.2020.2983224.

[47] X. Kang, B. Zhuo, P. Duan, Dual-path network-based hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 16 (2019) 447–451, http://dx.doi.org/10.1109/LGRS.2018.2873476.

[48] M.E. Paoletti, J.M. Haut, R. Fernandez-Beltran, J. Plaza, A.J. Plaza, F. Pla, Deep pyramidal residual networks for spectral–spatial hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2019) 740–754, http://dx.doi.org/10.1109/TGRS.2018.2860125.

[49] T. Alipour-Fard, M.E. Paoletti, J.M. Haut, H. Arefi, J. Plaza, A. Plaza, Multi-branch selective kernel networks for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 18 (2021) 1089–1093, http://dx.doi.org/10.1109/LGRS.2020.2990971.

[50] J.M. Haut, M.E. Paoletti, J. Plaza, A. Plaza, J. Li, Visual attention-driven hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2019) 8065–8080, http://dx.doi.org/10.1109/TGRS.2019.2918080.

[51] P. Wu, Z. Cui, Z. Gan, F. Liu, Residual group channel and space attention network for hyperspectral image classification, Remote Sens. 12 (2020) 2035, http://dx.doi.org/10.3390/rs12122035.

[52] Q. Xu, Y. Xiao, D. Wang, B. Luo, CSA-MSO3DCNN: Multiscale octave 3D CNN with channel and spatial attention for hyperspectral image classification, Remote Sens. 12 (2020) 188, http://dx.doi.org/10.3390/rs12010188.

[53] G. Ren, Y. Wang, J. Ning, Z. Zhang, Using near-infrared hyperspectral imaging with multiple decision tree methods to delineate black tea quality, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 237 (2020) 118407, http://dx.doi.org/10.1016/j.saa.2020.118407.

[54] S. Mei, X. Li, X. Liu, H. Cai, Q. Du, Hyperspectral image classification using attention-based bidirectional long short-term memory network, IEEE Trans. Geosci. Remote Sens. (2021) 1–12, http://dx.doi.org/10.1109/TGRS.2021.3102034.

[55] L. Mou, X. Lu, X. Li, X.X. Zhu, Nonlocal graph convolutional networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (2020) 8246–8257, http://dx.doi.org/10.1109/TGRS.2020.2973363.

[56] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, A. Tariq, Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 7570–7588, http://dx.doi.org/10.1109/JSTARS.2021.3099118.

[57] M.E. Paoletti, J.M. Haut, S.K. Roy, E.M.T. Hendrix, Rotation equivariant convolutional neural networks for hyperspectral image classification, IEEE Access 8 (2020) 179575–179591, http://dx.doi.org/10.1109/ACCESS.2020.3027776.

[58] J. Wang, X. Song, L. Sun, W. Huang, J. Wang, A novel cubic convolutional neural network for hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 4133–4148, http://dx.doi.org/10.1109/JSTARS.2020.3008949.

[59] Z. Ge, G. Cao, X. Li, P. Fu, Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 5776–5788, http://dx.doi.org/10.1109/JSTARS.2020.3024841.

[60] A. Mohan, M. Venkatesan, HybridCNN based hyperspectral image classification using multiscale spatiospectral features, Infrared Phys. Technol. 108 (2020) 103326, http://dx.doi.org/10.1016/j.infrared.2020.103326.

[61] D. AL-Alimi, Y. Shao, R. Feng, M.A.A. Al-Qaness, M.A. Elaziz, S. Kim, Multi-scale geospatial object detection based on shallow-deep feature extraction, Remote Sens. 11 (2019) 2525, http://dx.doi.org/10.3390/rs11212525.

[62] C. Garbin, X. Zhu, O. Marques, Dropout. vs, Dropout vs. batch normalization: an empirical study of their impact to deep learning, Multimedia Tools Appl. 79 (2020) 12777–12815, http://dx.doi.org/10.1007/s11042-019-08453-9.

[63] J.D. Tarnas, J.F. Mustard, X. Wu, E. Das, K.M. Cannon, C.B. Hundal, A.C. Pascuzzo, J.R. Kellner, M. Parente, Successes and challenges of factor analysis/target transformation application to visible-to-near-infrared hyperspectral data, Icarus 365 (2021) 114402, http://dx.doi.org/10.1016/j.icarus.2021.114402.

[64] C.O.S. Sorzano, J. Vargas, A.P. Montano, A survey of dimensionality reduction techniques, 2014, ArXiv E-Prints. arXiv:1403.2877. https://ui.adsabs.harvard.edu/abs/2014arXiv1403.2877S.

[65] J. Bjorck, C. Gomes, B. Selman, K.Q. Weinberger, Understanding batch normalization, 2018, ArXiv E-Prints. arXiv:1806.02375. https://ui.adsabs.harvard.edu/abs/2018arXiv180602375B.

[66] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, X. Cao, Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 55 (2017) 5585–5599, http://dx.doi.org/10.1109/TGRS.2017.2710079.

[67] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 17 (2020) 277–281, http://dx.doi.org/10.1109/LGRS.2019.2918719.

[68] J. Fang, X. Cao, Multidimensional relation learning for hyperspectral image classification, Neurocomputing 410 (2020) 211–219, http://dx.doi.org/10.1016/j.neucom.2020.05.034.