A shallow network for hyperspectral image classification using an autoencoder with convolutional neural network

Heena Patel[1] · Kishor P. Upla[1]

Abstract

This paper addresses an approach for the classification of hyperspectral imagery (HSI). In remote sensing, the HSI sensor acquires hundreds of images with narrow and continuous spectral width in visible and near-infrared regions of the electromagnetic (EM) spectrum. Such nature of data acquisition is very useful in the classification and/or the identification of different objects present in the HSI data. However, the low-spatial resolution and large volume of HS images make it more challenging. In the proposed approach, we use an autoencoder with convolutional neural network (AECNN) for the classification of HS images. Preprocessing with autoencoder enhances the features in the HS images which helps to obtain optimized weights in the initial layers of the CNN model. Hence, shallow CNN architecture can be utilized to extract features from the pre-processed HSI data which are used further for the classification of the same. The potential of the proposed approach has been verified by conducting many experiments on various datasets. The classification results obtained using the proposed method are compared with many state-of-the-art deep learning based methods including the winner of the geoscience and remote sensing society (GRSS) Image Fusion Contest-2018 on HSI classification held at IEEE International Geoscience and Remote Sensing Symposium (IGARSS)-2018 and it shows superiority over those methods.

1 Introduction

Hyperspectral imagery (HSI) is a collection of images acquired in the spectral range of visible and near-infrared regions of the electromagnetic (EM) spectrum. It provides hundreds of spectral channels over the acquired geographical area [5]. Due to this, HS

✉ Kishor P. Upla
  kishorupla@gmail.com

  Heena Patel
  hpatel1323@gmail.com

1 Sardar Vallabhbhai National Institute of Technology, Surat 395007, India

images are very useful in the classification and/or identification of the objects present in the captured scene. The HSI classification has been widely used in a variety of applications. However, HSI classification is a challenging task mainly due to its nature of mixed pixels and large amount of data volume [46]. This demands more efficient and robust techniques for the classification of the same in order to extract features from the HSI data.

Recently, the use of deep learning models in the different tasks of computer vision and image processing has led to an active area of research due to its ability to obtain higher accuracy than that of traditional methods. In specific, a lot of attention has been attracted by a convolutional neural network (CNN), due to its capability of extracting features automatically from various kind of images. Also, it can be applied as an end-to-end framework for multi-task learning which has been utilised in many applications [7, 12, 32, 51]. In addition to that, CNN can be widely used in object localization [40], detection [42], recognition [24] as well as classification problem [22]. In comparison to the traditional image classification methods, CNN can extract features efficiently alongside with accurate classification map using the same network. In the literature, many methods have been proposed for the classification of HSI data using different CNN models which aim to extract efficient high-level deep features from HSI data [44].

In this paper, an autoencoder with CNN (AECNN) is utilized to improve the classification accuracy of the HS data. Literature shows that the use of autoencoder is effective to enhance the features of data with reduced noise by adopting the concept of sparsity [25]. In the image classification problem, it is useful as it represents the image features effectively at abstract level with reduced dimensionality [9, 33]. The enhanced feature map obtained through the autoencoder is fed to the CNN architecture in the proposed method which extracts the efficient features at each subsequent layer. The final classification is performed through the softmax classifier. The proposed framework includes a shallow architecture that does not require any additional post-processing module for classification. Moreover, the CNN architecture consists of only two convolution layers and three fully connected dense layers. Thus, the proposed method has a shallow architecture. The key contributions of the proposed method are as follows:

– The idea of using an autoencoder makes the proposed method more efficient as compared to the other existing methods.
– The shallow CNN architecture is proposed which is efficient to extract the different features from HS image. Additionally, it is more robust against the different kinds of HSI data.
– The use of autoencoder with shallow CNN architecture reduces the overall complexity of the proposed method when it is compared to the existing deep learning based methods.

We published this idea in the Workshop on Computer Vision Application (WCVA) held at Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2018) [41]. However, the following additional works have been appended in this manuscript:

– The recent literature on hyperspectral image classification using CNN has been reviewed and the same is appended in this manuscript.
– The detailed description of the proposed method is elaborated with different block schematic diagram.

– In addition, the motivation behind each component utilized in the proposed method is narrated with the proposed methodology.
– The experimental results section has been expanded with new experiments on additional datasets, and detailed description of the parameter settings used in the proposed method has also been added to this manuscript.
– To show the comparison between the HSI classification results obtained using the proposed method with the recent state-of-the-art methods, the new results obtained using the recently proposed CNN-based methods are also presented along with the traditional methods.

The rest of this paper is organized as follows. The detailed review of the existing deep learning based methods is presented in Sect. 2. In Sect. 3 we describe the overall framework of the proposed method. The experimental results and discussions are presented in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related works

In the last few decades, many researchers have been attempted to address this problem. In the early era, spectral-domain classifiers such as multimodal logistic regression (MLR) [28], random forest (RF) [16] and support vector machines (SVMs) [15, 39] were utilised for HSI classification. Recently, the other methods include sparsity [10], Markov random fields (MRFs) [27] and morphological profiles (MPs) [4] which show the promising improvement over the traditional methods since they use both spatial and spectral details for HSI classification.

The use of CNN in the classification of HSI data is pioneered by Chen et al. [8]. They utilize many convolutional layers in their architecture which extract invariant and non-linear features from the hyperspectral image. Then, in [49], authors have implemented a deep CNN model with a dimension reduction algorithm for spatial-spectral feature extraction from the hyperspectral image. Moreover, authors in [36, 48] have proposed the HSI classification by constructing the CNN network with hierarchical feature extraction. Further, Aptoula et al. [2] learn several attribute profiles as an input to the CNN framework and stacked up on raw hyperspectral data. They classify the hyperspectral data efficiently by capturing its spectral and geometric properties.

The aforementioned CNN models have focused on automatic spatial and spectral feature extraction. Also, Zhao et al. [50] presented a different strategy that composite the extracted features by deep learning at multiple spatial scales which improves the performance of HSI classification to some extent. Moreover, the selection of partial view strategy is utilized to combine the various views and applied as an input to the specific CNN architecture for land usage classification [34]. Deep CNN is also presented to classify HSI in spectral-domain in [21, 47]. Authors in [30] presented the classification using pixel-pair features by combining the centre pixel and each of the surrounding pixels. Recently, Haut et al. [18] proposed a new model that exploits both spatial-contextual and spectral information in HSI by utilizing Bayesian CNN (BCNN). Moreover, authors in [45] address a data classification of multi-source data. They use two tunnel CNN (2T-CNN) framework (i.e. extraction of spatial and spectral features separately by a single module) and fusion is performed on the multi-source data. Further, Fang et al. [14] represent a two-branch network with semi-supervised learning (TBSSL); these training modules are designed and integrated for final classification.

Additionally, Duan et al. [13] represent the local and global edge-preserving operations along with MRF prior for smooth HSI classification. A spectral-spatial attention network (SSAN) is introduced by authors in [37], in which discriminative spectral-spatial features are captured from attention areas and the attention network helps to suppress the effects of interfering pixels. Furthermore, the Invariant Attribute Profiles (IAP) method is utilized to extract spatial and frequency domain features and joint spatial-frequency features are used in the application of HSI classification [20]. To reveal the complex high-order relationships of the HSI, authors in [43] propose the enhanced hybrid-graph learning method which extracts discriminative features with a dimensionality reduction model. Additionally, Hao et al. [17] represent the HSI classification method based on multi-label augmentation in which they extract features using recursive filtering and perform segmentation on super-pixel by removing mislabeled samples. Authors in [6] propose an active learning concept with deep learning approach which selects the most informative samples for labelling the image and hence it reduces the labelling cost.

An extensive amount of HSI classification works using deep learning have been reported; nevertheless, there is a lack of investigations to classify various similar kinds of features from HS image in the above work. Therefore, it is necessary to explore the CNN framework which can extract spectral and spatial features simultaneously from different bands of HSI data and also improve the accuracy of HSI classification.

3 Proposed architecture pipeline

In Fig. 1, we display the framework of the proposed method for the classification of HS image. It consists of a shallow architecture, that includes autoencoder and CNN to improve the performance of the network. The detailed description about the important of the autoencoder module along with various components utilized in the proposed method is described below:

– In the literature, it is experimentally proved that the use of autoencoder is effective in enhancing the non-linear features of high-dimensional data [9]. It is an unsupervised network that acts as a regularizer and hence it affects the starting point of supervised training in the CNN model. Thus, pre-trained networks consistently provide better generalization. It is also observed that pre-trained networks learn qualitatively different functions compared to without using pre-trained networks. Basically, unsupervised training using autoencoder favours the hidden units that compute features of the input $[x]$ (input to CNN network) that correspond to major factors of variation in the true input $[P(x)]$ i.e., probability of occurring of input $[x]$. Assuming that some of these near features that are useful at predicting variations in $[y]$ (actual output of CNN network), unsupervised pre-training sets up the parameters nearer to a solution of low predictive generalization error.

– Additionally, it is also observed that with a small training set, usually, researchers are not giving importance to minimizing the training error, because of overfitting issue. The training error is not a better way to distinguish between the generalized performance of the two models. In that case, unsupervised pre-training helps to find apparent local minima that has better generalization error.

– It is also analyzed in the proposed method that almost 100% reconstruction is obtained at the output of autoencoder and the features are enhanced better than that of the
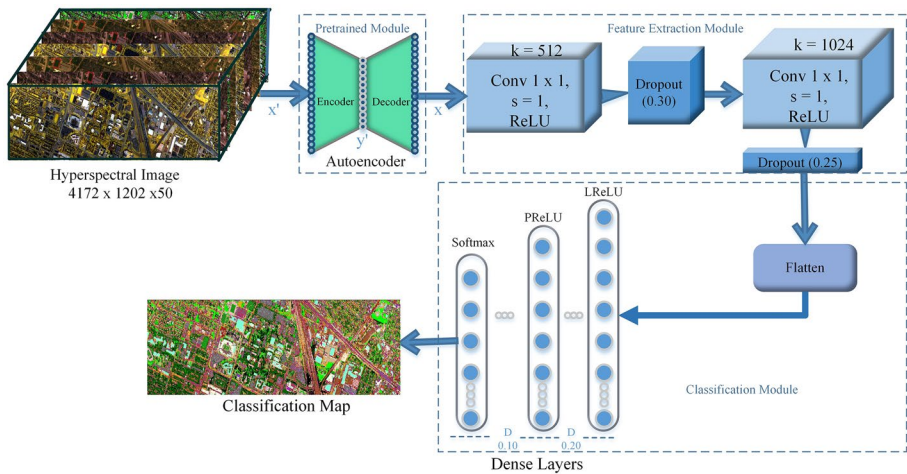
Fig. 1 The architecture framework of the proposed method-AECNN. Here, k, s and D indicate the number of channels, size of stride and dropout in the network, respectively

original hyperspectral image. Thus, the use of autoencoder also acts as a kind of network pre-conditioner, that keeps the parameter values in the appropriate range for further supervised training of CNN framework. Therefore, accurate feature enhancement is performed by pre-processing the hyperspectral image with the autoencoder which also initializes the CNN model to a point in parameter space that somehow renders the optimization process more effective, in the sense of achieving a lower minimum of the empirical cost function. The experimental justification of using autoencoder with CNN is also verified and it is included in Subsection 4.2.

– Dropout simulates sparse activation from a particular layer. Interestingly, this allows the network to learn sparse expressions. Therefore, it can be used an alternative to activity regularization to facilitate sparse representation in the deep learning model. Hence, regularization method such as dropout is set between each layer in the network in order to reduce the overfitting.

– Further, the number of feature maps often increases with the depth of the network. This issue can dramatically increase the number of parameters and calculations required when using larger filter. To address this issue, 1×1 convolution layer is utilized in the proposed method that provides per-channel pooling, and is often referred to as feature map pooling. This can be used for dimensionality reduction. It reduces the number of parameters while retaining the salient features of each feature map.

In the proposed method, an autoencoder takes an original data as input ($[x']$) and encode it in hidden layer ($[y']$) by mapping,

$$[y'] = ReLU(W[x'] + b), \tag{1}$$

where, W and b are the weight and bias of the autoencoder, respectively. The dimensionality of the intermediate latent representation is dynamic for all dataset due to the large variation in size of dataset which are 2^{10}, 2^{15} and 2^{25} for Salinas, Pavia and Houston, respectively. Finally, it generates the output ($[x]$) by decoding the data from hidden layer ($[y']$) as,

$[x] = \sigma(W'[y'] + b')$. Here, σ represents the sigmoid of given function. The reconstructed error is measured by using square error function as,

$$Loss(x', x) = ||x' - x||^2. \tag{2}$$

In the autoencoder, since the number of input nodes are same as the number of pixels in an HS image, the pixel's information is propagated through the network by discarding the nodes which are redundant. Finally, it reconstructs the features at the output layer with better enhancement. The output of the autoencoder i.e., $[x]$ is used as an input to the convolution layer of CNN.

The CNN module consists of two convolution layers with the kernel of 1×1 to extract multiple features. The first layer includes 512 channels to extract deep features. In deep learning, a network without an activation function works as a linear regression model which cannot perform the given task in a better way. Hence, it requires to use of a non-linear activation function in order to learn the complicated and complex form of data. Earlier, the sigmoid function was used as an activation function but due to its vanishing gradient problem, rectified linear unit (ReLU) is used nowadays [38]. It is defined mathematically as, $f(x) = max(0, x)$, where x represents the value of that particular node. ReLU describes that function is activated above the zero value; hence its partial derivative is one. Thus, the problem of vanishing gradient does not exist. Moreover, it saturates at zero which is more helpful to use for features at the initial layer in CNN. However, it has a disadvantage during optimization due to zero gradient whenever the unit is inactive. That means the algorithm would never adjust the weights for initially inactive nodes. In the second layer of CNN, the number of channels are increased (i.e. 1024) from that of the first layer to extract high-level features. The classification task demands more robust features to classify each object present in the image. Hence, there are three dense layers used in the proposed method. In the first dense layer, output data from the convolutional layer is processed with 2048 nodes and here leaky ReLu (LReLU) is utilized instead of Relu activation function. The LReLU is mathematically described as,

$$f(x) = \begin{cases} x, & if \quad x > 0; \\ 0.01x, & otherwise. \end{cases} \tag{3}$$

It allows small non-zero gradient when nodes are inactive and sacrifices the sparsity for the gradient during optimization which is more robust [35]. Moreover, parametric ReLU (PReLU) [19] makes the use of coefficient of leakage (a) into a parameter that is learned along with the other neural network parameters. It improves the model fitting with low computational cost and reduces the risk of over-fitting. Hence, second dense layer is implemented with 1024 nodes and data mapping is performed on that layer by utilizing the PReLU which is described as,

$$f(x) = \begin{cases} x, & if \quad x > 0; \\ ax, & otherwise. \end{cases} \tag{4}$$

Finally, softmax function is used at the output layer with the nodes having the total number of classes. The softmax layer is specifically used for classification problem to compute the probability of each class. It can be expressed as,

$$P(\hat{y} = c|x) = \frac{exp\left(x^T W_c\right)}{\sum_{k=1}^{C} exp\left(x^T W_k\right)}. \tag{5}$$

Equation (5) describes the predicted probability of cth class, defined by given input vector x and weight vector W. The loss function in the network measures the performance of the classification between the predicted ($\hat{y}$) and actual (y) classes among the total number of samples (M) present in the image. For more than two classes, categorical cross-entropy is preferred and it can be calculated by separating the loss for each class label and then sums the result of an individual as,

$$L(y, \hat{y}) = -\sum_{c=1}^{M} y_c log(\hat{y}_c). \tag{6}$$

Furthermore, an appropriate optimizer enhance the results with better efficiency. Hence, Adam is utilized to optimize the model for the training of high dimensional hyperspectral image. Adam takes the benefits of adaptive gradient and RMS propagation. It uses the average of the gradient's second moment [23], instead of using adapted learning rate parameter on the average and hence, it calculates an exponential moving average of the gradient and the squared gradient. It is described as,

$$W'_{ij+1} = W_{ij} - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t, \tag{7}$$

where, $\hat{m}_t = \frac{m_t}{1-\alpha_1^t}$ and $\hat{v}_t = \frac{v_t}{1-\alpha_2^t}$. Also, $m_t = \alpha_1 m_{t-1} + (1 - \alpha_1)g_t$ represents the decaying average of past gradient and $v_t = \alpha_2 v_{t-1} + (1 - \alpha_2)g_t^2$ which is decaying average of past squared gradient, α_1 and α_2 are decay rate that are close to 1.

4 Experimental results

In order to evaluate the performance of the proposed method, we have conducted many experiments on different datasets such as Indian Pines [3], Salinas, Pavia [1] and Houston [26]. All experiments have been performed on the computer with the following specifications: Intel Core $i7 - 7700K$ processor with $8M$ Cache and $3.60GHz$, 32 GB RAM with NVIDIA GeForce GTX 1070 8GB GPU. Here, we compare our classification results with two traditional methods such as SVM [15] and ELM [29] as well as recently proposed CNN based methods such as 3D-CNN [47], CNN-PPF [30], BCNN [18] and two tunnel CNN (2T-CNN) [45]. The experimental results obtained for the Houston dataset using the proposed method are also compared with the results obtained using the winner of the geoscience and remote sensing society (GRSS) Image Fusion Contest-2018 held at IEEE International Geoscience and Remote Sensing Symposium (IGARSS)-2018 (i.e., TBSSL [14]). In order to verify the quantitative performance of the proposed method, we use the different parameters such as accuracy and efficiency coefficient called kappa, (κ) [11] along with other quantitative measures such as precision, recall rate and f1-score on the classification results of the HS image.

4.1 Parameter settings and training details

In the experiments, pixel-wise annotation is performed using ERDAS Imagine and annotated data are exported to generate *.mat* file for importing in the algorithm. The original dataset has been divided into training and testing according to Ref. [45]. Experiments on Indian pines, Salinas and Pavia are conducted with batch size and epochs of 32 and 50, respectively. However, due to the large size of the Houston dataset, batch size of 128 has been chosen for fine-tuning. Additionally, Houston data is trained with 500 epochs due to more number of classes with similar kind of features. Adam optimizer is chosen to optimize the training process with $\eta = 0.001$, $\alpha_1 = 0.9$, $\alpha_2 = 0.999$ and $\epsilon = 10^{-8}$ [23]. Here, η is the learning rate, α_1 and α_2 represent exponential decay for first and second-moment estimation, respectively. ϵ is used to prevent division by zero. In the proposed method, an autoencoder and CNN networks are trained separately. First, the autoencoder is trained as a preprocessing module that reduces the dimensionality and enhances the features of the input image to some extent. Then, the generated output of the autoencoder is directly fed to CNN network. In the proposed method, 1×1 convolution is used in order to reduce the number of parameters. It increases the network capacity by adding non-linearity and also reduces the computation cost compared to 3×3 and 5×5 convolution [31]. Regularization method such as dropout is set between each layer in the network in order to reduce the overfitting.

4.2 Performance of the proposed algorithm without using autoencoder

In the proposed method, an autoencoder is used as a pre-processing step before CNN to obtain optimized weights in the initial layer of CNN network which could also reduce the number of parameters. An experiment has been conducted in order to see the classification results obtained using the proposed method without use of an autoencoder. In this experiment, we have observed that the training fails due to overfitting with accuracy of less than 25% on Indian pines, Pavia and Salinas datasets. In the case of Houston dataset, the proposed method without autoencoder attains 16-18(%) accuracy only. In order to increase the accuracy, deep CNN could be implemented with more number of layers and various filters. However, deep CNN makes architecture more complex and increases the training time. Hence, an autoencoder is utilized as a preconditioning step in the proposed method to obtain the optimized weights of CNN architecture. This also enhances the features of HS image and helps to improve the classification accuracy with less training time.

4.3 Evaluation on hyperspectral datasets

The potential of the proposed method has been verified by conducting the experiments on four different HSI datasets: 1. Indian pines, 2. Pavia University, 3. Salinas and 4. Houston. Table 1 describes the annotation of classes with different color codes of various datasets.

4.3.1 Indian pines dataset [3]

The first experiment is conducted on the well-known Indian pines dataset. It is acquired by an Airborne Visible Infrared Imaging Spectrometer (AVIIS) sensor which includes 200 spectral bands over the Indian Pines test site in North-western Indiana. The size of dataset consists of 145×145 pixels with wavelength range of 0.4×10^{-6} to 2.5×10^{-6} meters. This dataset contains 16 classes including natural perennial vegetation,

Table 1 Annotation of classes of Indian pines, Salinas, Pavia and Houston dataset

Color	Indian Pines	Salinas
	Background	Background
	Alfalfa	Brocoli _green_weeds_1
	Corn-notill	Brocoli _green_weeds_2
	Corn-mintill	Fallow
	Corn	Fallow _rough_plow
	Grass-pasture	Fallow _smooth
	Grass-trees	Stubble
	Grass-pasture-mowed	Celery
	Hay-windrowed	Grapes _untrained
	Oats	Soil _vinyard _develop
	Soybean-notill	Corn _senesced_green_weeds
	Soybean-mintill	Lettuce _romaine_4wk
	Soybean-clean	Lettuce _romaine_5wk
	Wheat	Lettuce _romaine_6wk
	Woods	Lettuce _romaine_7wk
	Building-Grass-Trees-Drives	Vinyard _untrained
	Stone-Steel-Towers	Vinyard _vertical _trellis

Color	Houston	Color	Pavia
	Healthy grass		Asphalt
	Stressed grass		Meadows
	Stadium seat		Gravel
	Paved parking lot		Trees
	Unpaved parking lot		Painted metal sheets
	Deciduous tree		Bare soil
	Bare earth		Bitumen
	Train		Self-Blocking
	Evergreen tree		Shadows
	Artificial turf		Background
	Road		
	Highway		
	Railway		
	Car		
	Residential building		
	Non-residential building		
	Major throughfare		
	Crosswalk		
	Sidewalk		
	Unclassified		

agriculture, forest, a rail line, major dual-lane highways, low-density housing, smaller roads and other built structures [3]. The classification maps obtained using the proposed method along with the other exiting methods are displayed in Fig. 2. In Fig. 2a, b, the original dataset and its corresponding ground-truth classification map are displayed. Figure 2c−h display the classification maps obtained using the different existing traditional and state-of-the-art methods. In Fig. 2i, the HSI classification map obtained using the proposed AECNN method is displayed. By looking at this figure, it can be observed that the classification result obtained using the proposed method is more close to the ground truth map than any other methods. Table 3 describes the quantitative comparison of the proposed method with the other existing HSI classification methods. We obtain 99.90 % accuracy and kappa of 0.9988 value using the proposed method. The other performance parameters for the classification map obtained using the proposed method are also evaluated and the same are depicted in Table 2 with tested samples. It is observed from Fig. 2 as well as from Tables 2 and 3 that the qualitative and quantitative performances of the proposed method i.e., AECNN are better when compared to same with the recently proposed existing methods.

4.3.2 Pavia dataset [1]

Pavia University dataset is acquired by Reflective Optics System Imaging Spectrometer (ROSIS) sensor over university of Pavia, Nothern Italy. It has 103 spectral bands ranging from 0.43 to 0.86 μm with 610×340 pixels and the spatial resolution of 1.3 m. This dataset has 9 classes [1]. Figure 3 shows the classification performance obtained using the proposed method along with the other methods on Pavia dataset. The evaluation parameters for each class obtained using the proposed method are displayed in Table 4 and the quantitative comparison of the proposed method along with the other exiting methods is displayed in Table 5. This shows that the proposed method obtains 99.95 % accuracy with a kappa value of 0.9993. Here, it is observed that the proposed method outperforms on all classes except the class of *Bare Soil*.

(a) Original Image   **(b)** Ground truth   **(c)** SVM

(d) ELM   **(e)** 3D-CNN   **(f)** CNN-PPF

(g) 2T-CNN   **(h)** BCNN   **(i)** AECNN (Proposed)
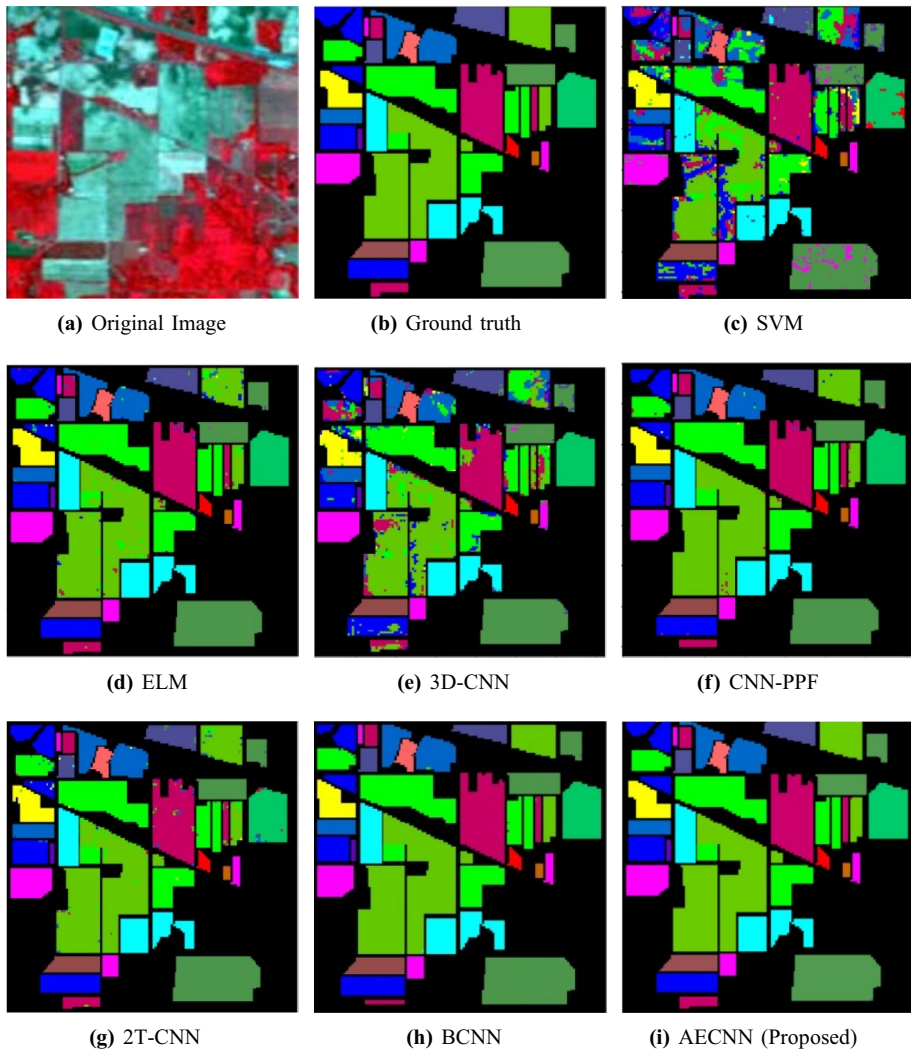
Fig. 2 The classification map obtained using the proposed method on Indian Pines dataset [3] with different existing methods

4.3.3 Salinas dataset [1]

The second experiment is conducted on the Salinas dataset. It is acquired by the AVIIS sensor and it includes 224 spectral bands over Salinas, California. The size of the images in this dataset is 512×217 pixels with a geometric resolution of 3.7 *m*. It contains 16 classes including vegetables, bare soils and vineyard fields [1]. The classification maps obtained using the proposed method along with the other exiting methods are depicted in Fig. 4 along with the original image and its ground truth classification map. Here, it can be observed that all classes except *Grapes_untrained* class are recovered well by using the proposed method. The *Grapes_untrained* class is displayed with dark green color.

Table 2 The different quantitative measures obtained using the proposed method for Indian Pines dataset [3]

Classes	Samples	Precision	Recall	f1-score
Alfalfa	3	1.00	1.00	1.00
Corn-notill	145	1.00	1.00	1.00
Corn-mintill	72	1.00	1.00	1.00
Corn	21	1.00	1.00	1.00
Grass-pasture	40	1.00	1.00	1.00
Grass-trees	67	1.00	1.00	1.00
Grass-pasture-mowed	1	1.00	1.00	1.00
Hay-windrowed	53	1.00	1.00	1.00
Oats	5	1.00	1.00	1.00
Soybean-notill	101	0.99	1.00	1.00
Soybean-mintill	246	1.00	1.00	1.00
Soybean-clean	52	1.00	1.00	1.00
Wheat	27	1.00	1.00	1.00
Woods	138	1.00	0.99	1.00
Building-Grass-Trees-Drives	41	0.98	1.00	0.99
Stone-Steel-Towers	13	1.00	1.00	1.00
Avg./Total	1025	1.00	1.00	1.00

The classification accuracy and kappa coefficient are also evaluated and the same are depicted in Table 7 for the proposed and other existing HSI classification methods. Table 6 describes the other parameters such as precision, recall rate and f1-score obtained using

Table 3 Comparison of the proposed method on Indian Pines dataset [3] with the different existing HSI classification methods. Here, the highest values are mentioned in bold font texts

Class	SVM	ELM	3D-CNN	CNN-PPF	2T-CNN	BCNN	Proposed
Alfalfa	100	100	100	100	100	100	**100**
Corn-notill	73.94	94.40	94.40	98.59	97.75	98.37	**99.31**
Corn-mintill	64.90	95.18	95.18	98.79	96.63	99.51	**100**
Corn	89.83	100	91.52	100	96.61	100	**100**
Grass-pasture	90.08	100	100	100	100	100	**100**
Grass-trees	96.72	100	100	99.54	99.45	100	**100**
Grass-pasture-mowed	71.42	100	100	100	100	100	**100**
Hay-windrowed	94.16	100	100	100	96.66	100	**100**
Oats	80.00	100	100	100	100	100	**100**
Soybean-notill	88.06	94.32	83.12	97.94	95.88	99.58	**100**
Soybean-mintill	48.04	86.96	66.28	97.96	92.50	99.83	**100**
Soybean-clean	58.10	93.27	79.05	98.87	96.62	98.64	**100**
Wheat	100	100	100	100	100	100	**100**
Woods	84.81	99.20	97.78	99.73	99.68	99.68	**100**
Building-Grass-Trees-Drives	85.56	94.80	94.84	99.13	90.72	100	**100**
Stone-Steel-Towers	95.65	100	100	100	100	100	**100**
OA	77.54	95.86	89.23	99.78	97.65	99.86	**100**
AA	73.74	94.48	83.30	98.86	96.37	99.53	**99.90**
κ	0.7062	0.9373	0.8113	0.9870	0.9587	0.9946	**0.9988**

(a) Original Image   **(b)** Ground truth   **(c)** SVM

(d) ELM   **(e)** 3D-CNN   **(f)** CNN-PPF

(g) 2T-CNN   **(h)** BCNN   **(i)** AECNN (Proposed)

Fig. 3 The classification map obtained using the proposed method on Pavia dataset [1] with different existing methods
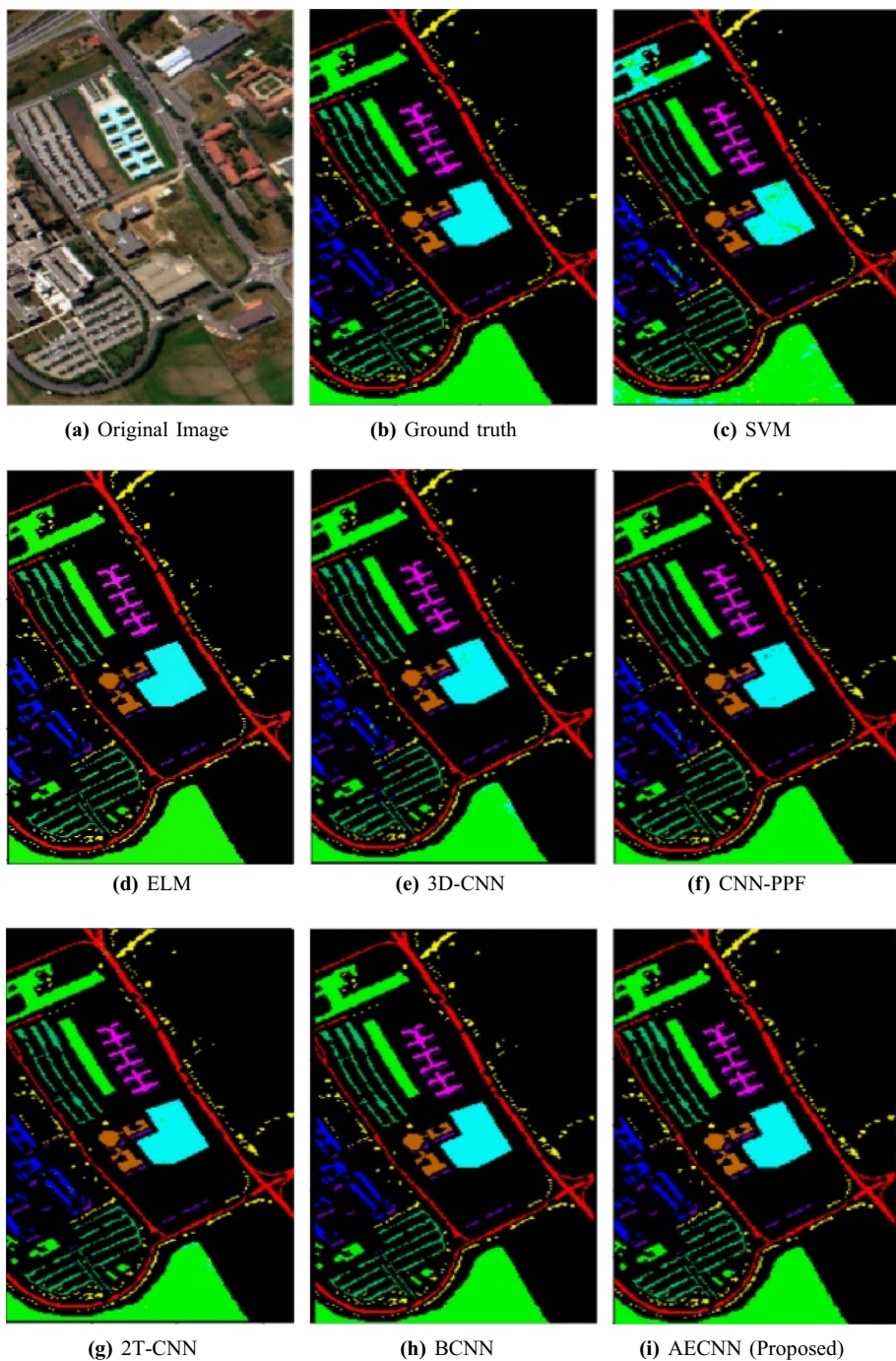
Table 4 The different quantitative measures obtained using the proposed method for Pavia dataset [1]

Classes	Samples	Precision	Recall	f1-score
Asphalt	665	1.00	1.00	1.00
Meadows	1845	1.00	1.00	1.00
Gravel	215	1.00	1.00	1.00
Trees	309	1.00	1.00	1.00
Painted metal sheets	140	1.00	1.00	1.00
Bare Soil	488	1.00	1.00	1.00
Bitumen	146	1.00	1.00	1.00
Self-Blocking Bricks	385	1.00	1.00	1.00
Shadows	85	1.00	1.00	1.00
Avg./Total	4278	1.00	1.00	1.00

the proposed method. In Table 7, it can also observe that the BCNN and 2T-CNN methods obtain 99.17% and 98.58% accuracy, respectively. However, the proposed method attains accuracy of 99.00% which is very close to BCNN method. In this experiment, classification results of the proposed method is close to BCNN method but it is better than the other existing HSI classification methods.

4.3.4 Houston dataset [26]

The last experiment is conducted on a recent dataset called Houston HSI dataset which has the highest number of classes. The HS images in this dataset are acquired by National Center for Airborne Laser Mapping (NCALM) and they are made available through IEEE GRSS Data Fusion Contest 2018. The images in this dataset are of size 4172×1202 with the spectral range of 380-1050 nm and they are acquired at 1-m GSD. It includes 20 classes with a range from natural land cover (e.g., grass, tree and bare earth) to man-made objects (e.g., vehicles, roads and buildings) [26]. The classification results obtained using the proposed method along with the recently proposed 2T-CNN method [45] are depicted in Fig. 5 for this

Table 5 Comparison of the proposed method on Pavia dataset [1] with the different existing HSI classification methods. Here, the highest values are mentioned in bold font texts

Class	SVM	ELM	3D-CNN	CNN-PPF	2T-CNN	BCNN	Proposed
Asphalt	94.27	99.77	98.00	97.83	97.64	99.93	**100**
Meadows	82.68	99.83	98.39	99.82	99.27	100	**100**
Gravel	85.90	96.66	89.33	93.01	90.85	99.61	**100**
Trees	96.08	99.51	99.35	99.34	99.21	99.96	**100**
Painted metal sheets	100	100	99.47	100	100	100	**100**
Bare Soil	91.40	**100**	98.72	98.74	98.74	99.92	99.79
Bitumen	93.97	100	98.79	99.74	100	100	**100**
Self-Blocking Bricks	95.65	98.77	92.94	99.36	95.42	99.67	**99.74**
Shadows	100	100	99.90	100	100	99.94	**100**
OA	91.12	99.75	98.35	99.13	98.54	99.98	**100**
AA	89.02	99.59	97.19	98.98	98.27	99.93	**99.95**
κ	0.8583	0.9944	0.9693	0.9865	0.9771	0.9980	**0.9993**

(a) Original Image   **(b)** Ground truth   **(c)** SVM

(d) ELM   **(e)** 3D-CNN   **(f)** CNN-PPF

(g) 2T-CNN   **(h)** BCNN   **(i)** AECNN (Proposed)
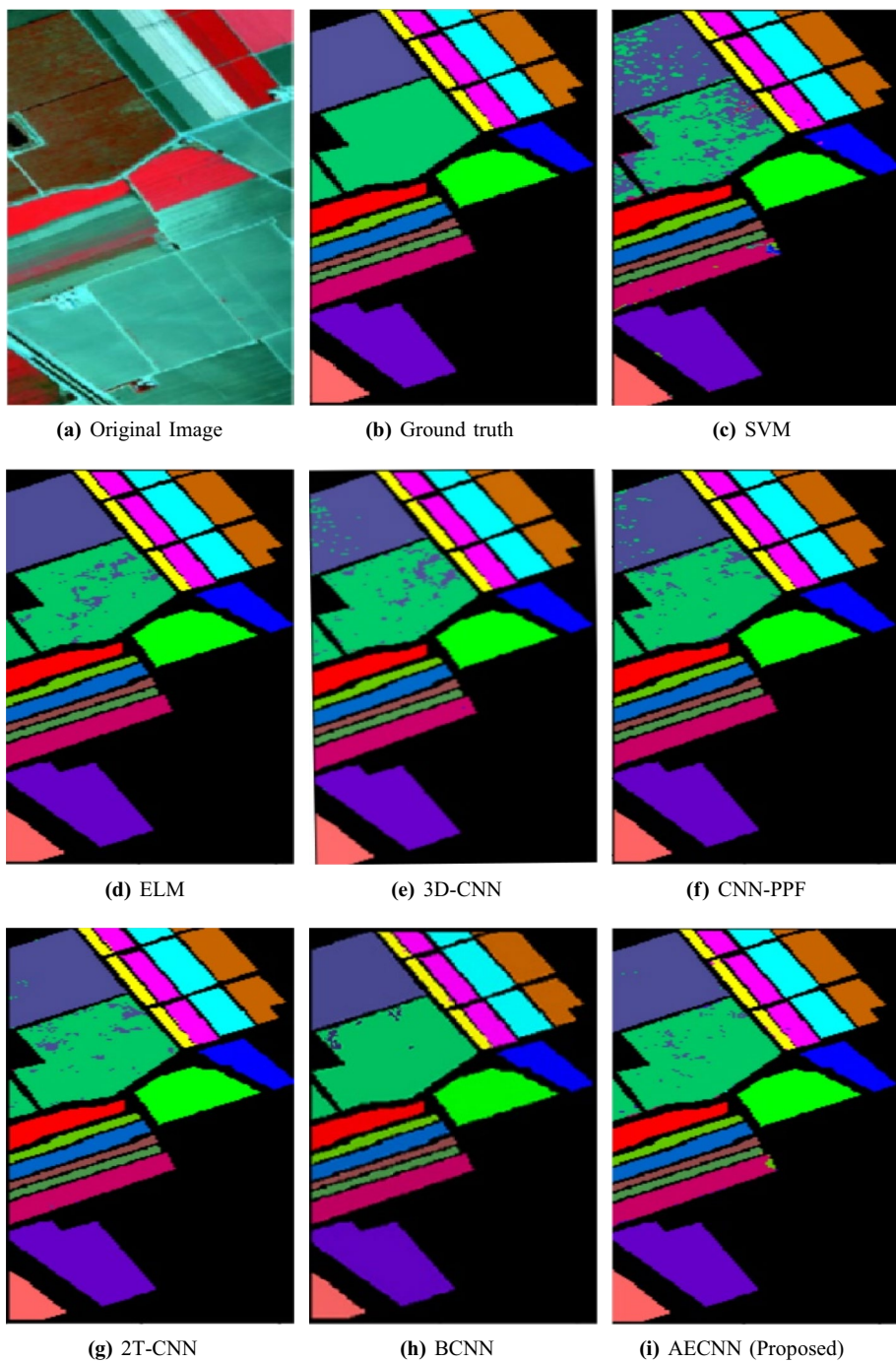
Fig. 4 The classification map obtained using the proposed method on Salinas dataset [1] with different existing methods

Table 6 The different quantitative measures obtained using the proposed method for Salinas dataset

Classes	Samples	Precision	Recall	f1-score
Brocoli_green_weeds_1	409	1.00	1.00	1.00
Brocoli_green_weeds_2	739	1.00	0.99	1.00
Fallow	400	1.00	0.98	0.99
Fallow_rough_plow	271	1.00	1.00	1.00
Fallow_smooth	557	1.00	0.99	1.00
Stubble	755	1.00	1.00	1.00
Celery	703	0.99	1.00	1.00
Grapes_untrained	2219	0.96	0.97	0.96
Soil_vinyard_develop	1263	1.00	1.00	1.00
Corn_senesced_green_weeds	664	0.99	0.98	0.99
Lettuce_romaine_4wk	208	1.00	1.00	1.00
Lettuce_romaine_5wk	378	0.95	1.00	0.97
Lettuce_romaine_6wk	188	1.00	1.00	1.00
Lettuce_romaine_7wk	206	0.99	1.00	0.99
Vinyard_untrained	1489	0.95	0.93	0.94
Vinyard_vertical_trellis	377	1.00	1.00	1.00
Avg./Total	10826	0.98	0.98	0.98

dataset. Here, in order to see the qualitative comparison among the results, some portions of these results are also zoomed and the same are displayed with red and blue color borders along with the HSI results. In the red box, the different details such as evergreen trees, healthy

Table 7 The comparison of the proposed method on Salinas dataset [1] with the different existing HSI classification methods. Here, the highest values are mentioned in bold font texts

Class	SVM	ELM	3D-CNN	CNN-PPF	2T-CNN	BCNN	Proposed
Brocoli_green_weeds_1	100	100	100	100	100	100	**100**
Brocoli_green_weeds_2	100	100	100	100	**100**	99.45	99.05
Fallow	99.39	100	100	100	99.94	**100**	98.00
Fallow_rough_plow	99.71	100	100	100	97.41	99.61	**100**
Fallow_smooth	99.10	99.81	99.85	97.75	98.80	**99.53**	99.46
Stubble	100	100	100	99.83	100	100	**100**
Celery	99.88	100	100	100	99.77	100	**100**
Grapes_untrained	70.33	90.68	89.49	94.11	95.28	**99.68**	99.66
Soil_vinyard_develop	99.61	100	100	100	100	**100**	99.76
Corn_senesced_green_weeds	94.14	100	99.39	99.69	96.95	**99.88**	99.49
Lettuce_romaine_4wk	99.25	100	100	99.68	100	100	**100**
Lettuce_romaine_5wk	100	100	100	100	100	99.97	**100**
Lettuce_romaine_6wk	100	100	100	99.98	100	99.95	**100**
Lettuce_romaine_7wk	99.25	100	100	100	97.75	**100**	99.51
Vinyard_untrained	85.80	**99.72**	95.81	99.35	99.50	99.66	93.14
Vinyard_vertical_trellis	99.77	100	100	100	100	**100**	99.73
OA	92.97	99.13	99.56	98.97	99.01	**99.98**	99.94
AA	91.39	98.01	97.19	98.55	98.58	**99.17**	99.00
κ	0.9045	0.9779	0.9688	0.9839	0.9745	**0.9876**	0.9842

grass, bare earth, sidewalk, etc., are present and one can observe that the classification result obtained using the proposed method is better when compared to same with 2T-CNN method [45]. Similarly, the blue color box represents the different details of classes such as highway, road, paved parking slot, rail, etc. The classification accuracy of the proposed method for this region is also close to that of the ground truth than that of 2T-CNN method [45].

The quantitative parameters are evaluated and the same are displayed in Table 8. Here, it shows that the proposed method obtains 87.19 % testing accuracy and kappa of 0.8684 value. These measures are better among the other existing HSI classification methods. It is also observed that the proposed method classifies well on similar type of classes such as healthy grass (96.52%), artificial turf (94.49%) and evergreen trees (97.21%). In addition to that, we have also compared our quantitative results with the winner of the GRSS Image Fusion Contest-2018 on HSI classification [14]. They obtained 77.39% accuracy and a kappa value of 0.7300 which proves that the proposed method achieves better performance on Houston dataset too. Due to the new arrival of such a complex dataset with similar kind of classes, we could compare our results with two recently proposed state-of-the-art methods (2T-CNN [45] and TBSSL [14]). The other quantitative measures such as precision, recall rate, f1-score and accuracy of individual classes are also depicted in Table 8.

(a) Original Image

(b) Ground truth

(c) 2T-CNN
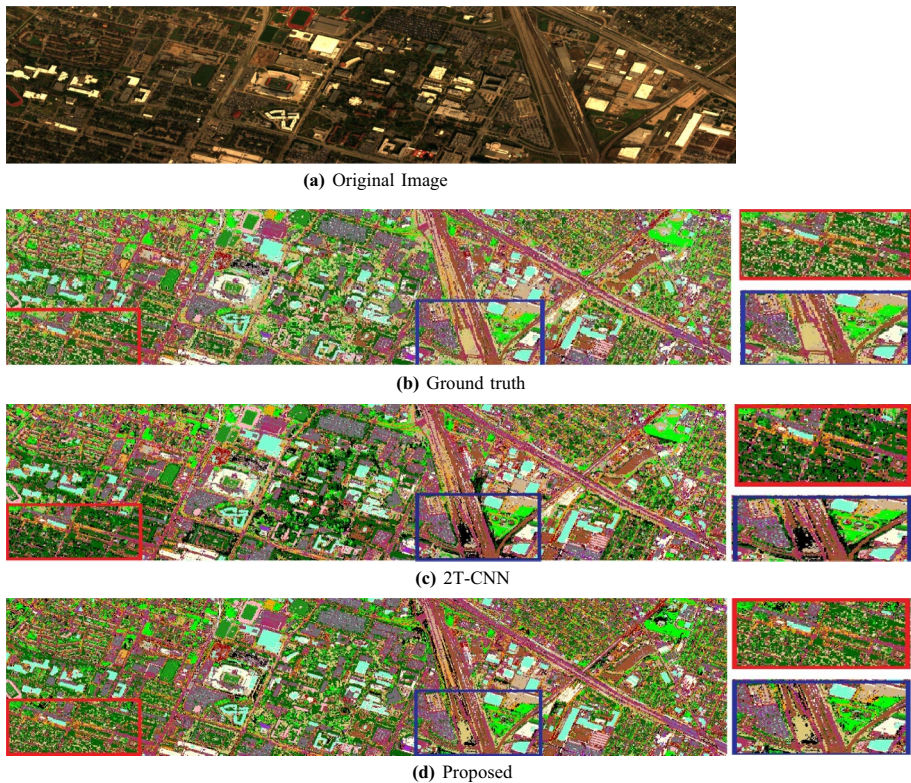
(d) Proposed

Fig. 5 The classification map obtained using the proposed method along with existing methods on Houston dataset [26]. The red and blue boxes consist of zoomed versions of regions shown with those colors in the HSI classification maps

Table 8 The different quantitative measures obtained using the proposed and other existing methods for Houston dataset. Here, the highest values are mentioned in bold font texts

Classes	Samples	2T-CNN Accuracy(%)	TBSSL Accuracy(%)	Proposed method Accuracy(%)	Precision	Recall	f1-score
Healthy grass	43072	82.47	79.52	**96.52**	0.92	0.96	0.94
Stressed grass	36804	85.11	64.00	**93.84**	0.92	0.92	0.92
Stadium seat	32564	71.56	**95.57**	75.15	0.75	0.72	0.74
Paved parking lot	71367	91.85	**92.05**	69.64	0.91	0.91	0.91
Unpaved parking lot	8953	**83.67**	74.36	78.69	0.83	0.76	0.79
Deciduous tree	49446	82.63	82.67	**94.28**	0.95	0.94	0.94
Bare earth	39114	**92.59**	84.83	85.93	0.83	0.84	0.84
Train	100206	79.48	**99.15**	90.84	0.84	0.90	0.87
Evergreen tree	102229	94.16	75.51	**97.21**	0.98	0.97	0.97
Artificial turf	44117	93.89	78.99	**94.49**	0.95	0.94	0.94
Road	13876	**75.96**	53.58	66.91	0.70	0.62	0.66
Highway	7465	56.52	66.01	**66.13**	0.74	0.61	0.67
Railway	70500	86.35	88.29	**93.15**	0.92	0.92	0.92
Car	16046	64.83	**71.23**	66.87	0.81	0.63	0.71
Residential building	17915	**85.35**	74.64	81.34	0.89	0.78	0.83
Non-residential building	32671	69.56	79.92	**87.98**	0.68	0.76	0.72
Major throughfare	9741	55.38	**61.22**	56.17	0.78	0.52	0.63
Sidewalk	53603	**96.79**	75.97	91.62	0.84	0.90	0.87
Crosswalk	31259	52.69	55.91	**56.24**	0.61	0.42	0.56
Unclassified	2523	50.36	51.65	**53.75**	0.59	0.46	0.52
OA	752212	77.56	77.39	**87.92**	0.88	0.88	0.88
AA	-	76.34	76.15	**87.19**			
κ	-	0.7623	0.7600	**0.8684**			

It is observed from the different experiments that the hybrid concept of an autoencoder and CNN in the proposed architecture comparatively provides better extraction of robust and deep features from the HSI data and also helps to improve the classification efficiency of HSI data. Moreover, the architecture in the proposed method is less complex than that of other existing methods which reduces the training time required to train the overall network.

5 Conclusion

In this paper, we propose a shallow CNN architecture with pre-processing autoencoder module for HSI classification. The task of autoencoder is to enhance the features present in the HS image and due to this, the CNN model extracts the features efficiently from the hyperspectral data. This pre-processing procedure leads to obtain optimized weights in the initial layer of the CNN model. This step in the proposed approach also reduces

the overall complexity when compared to other existing CNN methods. The experiments have been conducted on four different HSI datasets and they show that the proposed method performs better when compared to the same with the other existing HS image classification methods.

Declarations

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

References

1. http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, Dec 2018
2. Aptoula E, Ozdemir MC, Yanikoglu B (2016) Deep learning with attribute profiles for hyperspectral image classification. IEEE Geosci Remote Sens Lett 13(12):1970–1974
3. Baumgardner MF, Biehl LL, Landgrebe DA (2015) 220 band Aviris hyperspectral image data set: June 12, 1992 Indian pine test site 3. Purdue University Research Repository 10, R7RX991C
4. Benediktsson JA, Palmason JA, Sveinsson JR (2005) Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans Geosci Remote Sens 43(3):480–491
5. Camps-Valls G, Tuia D, Bruzzone L, Benediktsson JA (2014) Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE Signal Process Mag 31(1):45–54
6. Cao X, Yao J, Xu Z, Meng D (2020) Hyperspectral image classification with convolutional neural network and active learning. IEEE Trans Geosci Remote Sens
7. Chen X, Xiang S, Liu CL, Pan CH (2014) Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci Remote Sens Lett 11(10):1797–1801
8. Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54(10):6232–6251
9. Chen Y, Lin Z, Zhao X, Wang G, Gu Y (2014) Deep learning-based classification of hyperspectral data. IEEE J Sel Top Appl Earth Obs Remote Sens 7(6):2094–2107
10. Chen Y, Nasrabadi NM, Tran TD (2011) Hyperspectral image classification using dictionary-based sparse representation. IEEE Trans Geosci Remote Sens 49(10):3973–3985
11. Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data: Principles and practices. CRC Press
12. Ding C, Xu C, Tao D (2015) Multi-task pose-invariant face recognition. IEEE Trans Image Process 24(3):980–993
13. Duan P, Kang X, Li S, Ghamisi P, Benediktsson JA (2019) Fusion of multiple edge-preserving operations for hyperspectral image classification. IEEE Trans Geosci Remote Sens 57(12):10336–10349
14. Fang S, Quan D, Wang S, Zhang L, Zhou L (2018) A two-branch network with semi-supervised learning for hyperspectral classification. IGARSS - 2018 IEEE International Geoscience and Remote Sensing Symposium. pp 3860–3863
15. Gao L, Li J, Khodadadzadeh M, Plaza A, Zhang B, He Z, Yan H (2015) Subspace-based support vector machines for hyperspectral image classification. IEEE Geosci Remote Sens Lett 12(2):349–353
16. Ham J, Chen Y, Crawford MM, Ghosh J (2005) Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans Geosci Remote Sens 43(3):492–501
17. Hao Q, Li S, Kang X (2020) Multilabel sample augmentation-based hyperspectral image classification. IEEE Trans Geosci Remote Sens 58(6):4263–4278
18. Haut JM, Paoletti ME, Plaza J, Li J, Plaza A (2018) Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. IEEE Trans Geosci Remote Sens 99:1–22
19. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp 1026–1034
20. Hong D, Wu X, Ghamisi P, Chanussot J, Yokoya N, Zhu XX (2020) Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification. IEEE Trans Geosci Remote Sens 58(6):3791–3808

21. Hu W, Huang Y, Wei L, Zhang F, Li H (2015) Deep convolutional neural networks for hyperspectral image classification. J Sens 2015

22. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 1725–1732

23. Kingma DP, Ba JL (2014) Adam: A method for stochastic optimization. Proc. 3rd Int. Conf. Learn. Representations. pp. 1–15

24. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: A convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113

25. Le QV (2013) Building high-level features using large scale unsupervised learning. Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on pp. 8595–8598

26. Le Saux B, Yokoya N, Hansch R, Prasad S (2018) 2018 ieee grss data fusion contest: Multimodal land use classification [technical committees]. IEEE Geosci Remote Sens Mag 6(1):52–54

27. Li J, Bioucas-Dias JM, Plaza A (2012) Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. IEEE Trans Geosci Remote Sens 50(3):809–823

28. Li J, Bioucas-Dias JM, Plaza A (2013) Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. IEEE Geosci Remote Sens Lett 10(2):318–322

29. Li W, Chen C, Su H, Du Q (2015) Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans Geosci Remote Sens 53(7):3681–3693

30. Li W, Wu G, Zhang F, Du Q (2017) Hyperspectral image classification using deep pixel-pair features. IEEE Trans on Geoscience and Remote Sens 55(2):844–853

31. Lin M, Chen Q, Yan S (2013) Network in network. arXiv preprint arXiv:1312.4400

32. Liu W, Mei T, Zhang Y, Che C, Luo J (2015) Multi-task deep visual-semantic embedding for video thumbnail selection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3707–3715

33. Lore KG, Akintayo A, Sarkar S (2017) Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recogn 61:650–662

34. Luus FP, Salmon BP, Van den Bergh F, Maharaj BTJ (2015) Multiview deep learning for land-use classification. IEEE Geosci Remote Sens Lett 12(12):2448–2452

35. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models. In: Proc. icml, vol. 30. p. 3

36. Makantasis K, Karantzalos K, Doulamis A, Doulamis N (2015) Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, pp. 4959–4962

37. Mei X, Pan E, Ma Y, Dai X, Huang J, Fan F, Du Q, Zheng H, Ma J (2019) Spectral-spatial attention networks for hyperspectral image classification. Remote Sens 11(8):963

38. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814

39. Pal M, Foody GM (2010) Feature selection for classification of hyperspectral data by svm. IEEE Trans Geosci Remote Sens 48(5):2297–2307

40. Parkhi OM, Vedaldi A, Zisserman A et al (2015) Deep face recognition. In: BMVC 1. p 6

41. Patel H, Upla K (2018) Aecnn: Autoencoder with convolutional neural network for hyperspectral image classification. Workshop on Computer Vision Applications (WCVA) in IEEE Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)

42. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149

43. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) Hybridsn: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification. IEEE Geosci Remote Sens Lett 17(2):277–281

44. Windrim L, Melkumyan A, Murphy RJ, Chlingaryan A, Ramakrishnan R (2018) Pretraining for hyperspectral convolutional neural network classification. IEEE Trans Geosci Remote Sens

45. Xu X, Li W, Ran Q, Du Q, Gao L, Zhang B (2018) Multisource remote sensing data classification based on convolutional neural network. IEEE Trans Geosci Remote Sens 56(2):937–949

46. Yan D, Chu Y, Li L, Liu D (2018) Hyperspectral remote sensing image classification with information discriminative extreme learning machine. Multimed Tools Appl 77(5):5803–5818

47. Yu S, Jia S, Xu C (2017) Convolutional neural networks for hyperspectral image classification. Neurocomputing 219:88–98

48. Yue J, Zhao W, Mao S, Liu H (2015) Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens Lett 6(6):468–477

49. Zhao W, Du S (2016) Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Trans Geosci Remote Sens 54(8):4544–4554

50. Zhao W, Guo Z, Yue J, Zhang X, Luo L (2015) On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. Int J Remote Sens 36(13):3368–3379
51. Zhu C, Peng Y (2015) A boosted multi-task model for pedestrian detection with occlusion handling. IEEE Trans Image Process 24(12):5619–5629

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