



## Ensemble of features for efficient classification of high-resolution remote sensing image

Gladima Nisia T & Rajesh S

**To cite this article:** Gladima Nisia T & Rajesh S (2022) Ensemble of features for efficient classification of high-resolution remote sensing image, European Journal of Remote Sensing, 55:1, 326-337, DOI: [10.1080/22797254.2022.2075794](https://doi.org/10.1080/22797254.2022.2075794)

**To link to this article:** <https://doi.org/10.1080/22797254.2022.2075794>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 24 May 2022.



Submit your article to this journal



Article views: 36



View related articles



View Crossmark data

## Ensemble of features for efficient classification of high-resolution remote sensing image

Gladima Nisia T  <sup>a</sup> and Rajesh S  <sup>b</sup>

<sup>a</sup>Department of CSE, AAA College of Engineering and Technology, Sivakasi, India; <sup>b</sup>Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi, India

### ABSTRACT

Extracting feature is one of the important methods in classification of high-resolution remote sensing image. A good feature set can result in an efficient classification process. Recent trend moves in extracting the features from the image using neural networks with no human intervention. Our approach uses the deep convolutional neural network for extracting deep features. To still rise the efficiency of the extracted features, the proposed system combines the deep features with other features like Gabor features and novel reformed local binary pattern features. The features are combined and sent for classification. Then, the classification process is done to classify the images. The proposed system introduces two novel ideas, in its feature extraction implementation, namely (1) initialisation of filter values for the CNN and (2) change in local binary pattern feature extraction process. The experimental results are carried out with LISS IV Madurai image, and evaluation is done for the verification of the results. It is found that the system proposed produces good results when compared with other existing methods.

### ARTICLE HISTORY

Received 17 July 2021  
Revised 18 February 2022  
Accepted 6 May 2022

### KEYWORDS

Classification; DCNN; feature extraction; LISS IV madurai image; RLBP; remote sensing image

## Introduction

The remote sensing (RS) image contains lot of hidden details such as land cover and land mapping of the earth surface. The process of classifying each pixel in the RS image thus helps in revealing various useful details. The feature extraction process plays an important role in the classification process. There are lot of feature extraction techniques used before classifying RS images. In earlier days, those classification was done depending upon the low-level feature extraction process which extracts shape, context or local features based on the methods, namely histogram of oriented gradients and scale-invariant feature transform (SIFT; Fan et al., 2015; Yang & Newsam, 2010). The development in the technology slowly drove the feature extraction process from manually selected features to automatic feature learning by the so-called deep learning techniques (LeCun et al., 2015).

As another step towards improvement in feature learning, the perceptron was introduced. Zhouhan Lin et al. classified the image using a model which utilised the autoencoders for extracting the features in an unsupervised manner (Lin et al., 2013). The stacked autoencoders were used in extracting features from RS images, and those features were used for classification (Chen et al., 2014). At some point, the perceptron was replaced by deep learning concepts because of their limitations such as disregarding spatial information and need of many parameters. Meanwhile, the CNN had evolved as one of the efficient neural networks for

feature extraction because of its various benefits (Chen et al., 2014; Yue et al., 2015; Zhang et al., 2016b; Zhao et al., 2015). The main benefit of CNN compared to its predecessors is its automatic feature detection capability without any human direction. The network is computationally efficient with the usage of convolution and pooling operations. It allows parameter sharing thus enabling the model to run on any device. The CNN is used in many applications, namely image classification (He et al., 2015; Krizhevsky et al., 2017), target detection (Wanli Ouyang et al., 2015; Zhang et al., 2014), face recognition (Sun et al., 2014), pedestrian detection (Ouyang & Wang, 2013), hyper spectral image (HSI) denoising (Shi et al., 2021b), change detection in RS images (Shi et al., 2021a), building footprint information (Guo et al., 2021), RS image classification (Chen et al., 2014) and so on.

There are many different CNN architectures proposed at different times. Earlier in 1998, LeNet was invented by Yann LeCun for optical character recognition (LeCun et al., 1998). The architecture was tiny and simple. Even though it was a major breakthrough at that time, it does not perform well with RGB images. Then after a decade gap in 2012, AlexNet was introduced (Krizhevsky et al., 2017). It consists of eight layers with ReLu activation function. Compared to later models, it took additional time to attain higher accuracy results. Then, VGG was introduced to show the result of deep architecture on

accuracy for image classification or recognition tasks (Simonyan & Zisserman, 2015). The network works based on grouping of multiple convolution layers with small kernels resulting in the reduced number of features at the output. The network includes multiple  $3 \times 3$  conv layer, ReLu,  $2 \times 2$  max-pooling layer, fully connected (FC) and softmax layer. Although it had many advantages, it experienced vanishing gradient problem and some other problems due to huge computational requirements, in terms of both memory and time. Also, due to the large width of convolutional layers, it becomes inefficient. So, GoogLeNet/ Inception devised a module called inception module that constructs a sparse CNN with a normal dense construction (Szegedy et al., 2015). Then in the following year, ResNet was introduced as a very deep network with 152-layer network (He et al., 2016). The key idea of the network is adding a shortcut that bypasses two or more stacked convolutional layers, and at the end, they were added to the output of the stacked convolutions. Some of these ResNet model help in image enhancement and classification.

The transfer learning concepts are applied for the CNN-based classification of RS images. For instance, scene classification for different datasets (Pires de Lima & Marfurt, 2020; Jason Yosinski et al., 2014) and high-resolution RS scene classification (Hu et al., 2015) was done through transfer learning. The classification accuracy of CNN is increased with transfer learning. The main advantage is that pretraining a model on a larger and more general dataset improves the performance of the model for the classification of the smaller datasets in comparison with the same model trained with randomly initialised weights. Meanwhile, many neural network architectures were introduced for various purposes. The system for scene classification was introduced using SceneNet. Here, the network architecture coding and searching were achieved using an evolutionary algorithm (Ma et al., 2021). Another system with multi-view deep neural networks was introduced for HSI classification using deep neural network-based relevant latent representation learning. The system uses both autoencoders and semi-supervised graph convolutional network (Sellami & Tabbone, 2022). Similarly, a multitask deep learning method was introduced for HSI classification with unknown classes (Liu et al., 2020). The semantic information modulated deep subpixel mapping network was also developed for mixed pixel problem in RS images He et al., 2020).

In the system proposed by Hu (Hu et al., 2015), in order to abstract the features of HSI based on spectral perspective, a five-layer deep CNN was used, thus leading to an improvised classification performance. The work by Yue includes principal component analysis (PCA), deep convolutional neural network (DCNN) and logistic regression classification of RS

image (Yue et al., 2015). Another different combination with CNN was dependent on the blend of morphological profiles for feature extraction of HSI (Aptoula et al., 2016). The system proposed the method which combines the neural network and an object-oriented classification for RS image classification. Objects were detected using multi-resolution segmentation algorithm (Zhao W et al., 2017).

Some works were proposed which combines the deep features obtained from CNN with other features. The features from DCNN were combined with Gabor features and tested for the input image (Chen et al., 2017). Then as the next step, the deep features were combined with LBP features (Juefei-Xu et al., 2017). The deep features were also combined with SIFT feature (Yue et al., 2018) for the image retrieval. Another work (Tan et al., 2019) combines the grey-level co-occurrence matrix features with the deep features for classification of the images. For the feature extraction process, is it enough of considering only the deep features? Or is it enough that the deep features are combined with other features as mentioned in these existing methods? Still there is a chance for accuracy improvement because even though these existing works perform classification with deep features, there arise a small gap of not considering the other important features like texture, edge, scale and orientation in the same model. When these features are considered, the classification accuracy is increased and our system tries to fill this research gap.

In this proposed method, the system mainly concentrates on the extraction and combination of different features, namely deep features, Gabor features and reformed local binary pattern (RLBP) features. The deep feature extraction is performed using DCNN. Since our system deals with RS, the texture details of the image can't be neglected. So, the texture features are extracted using RLBP and Gabor features. Also, to enhance the resistance of deep features to the orientation and scale changes, the Gabor features are used. In the literature, the Gabor feature is used as the input of CNN. Similarly, the LBP feature is also used as the input of CNN. But our proposed system claims that the performance of CNN can be greatly enhanced by two-stage textural features consideration based on both Gabor feature and RLBP features. The experimental results are provided to explain that the combination of features in our proposed feature extraction method is best suited for our dataset. The Major process of our system is as follows:

- The proposed system pre-processes the input image with an image enhancement technique. Thus, the image is enhanced before it is sent for the next step. By doing so, the clarity of the input is improved.

- The proposed system uses multiple fused features intended for efficient classification. The Gabor and RLBP features are extracted. Both of these features are combined and given as input to DCNN. Thus, the feature set obtained consists of Gabor features, RLBP features and deep features.
- The features extracted are then classified using CNN, and the results are obtained.

The rest of the paper is arranged as follows. The earlier works associated with our system is explained in [section 2](#). The methodology proposed is described in [section 3](#). The experimental results and the discussion about the performance of the system are explained in [section 4](#), and finally, conclusion is provided in [section 5](#).

## Related work

Certain earlier methods which are related to our proposed system is discussed in this section. For extraction of spatial features, the Gabor filter was applied in different directions and scales (Mirzapour & Ghassemian, [2015](#)). Similarly, the LBP method was used to extract spatial features (Song et al., [2010](#)). Both the methods proved their improved efficiency impact on classification accuracy of the input images. But the main drawback of this shallow spatial features was the diminished accuracies of complex scenes and diverse land covers. In order to overcome the issue, CNN was employed.

Some proposed systems introduced the method of integrating the spectral features with shallow spatial features and used this combination as an input for deep neural networks. The Gabor features with different scales and directional changes were combined with PCA spectral features for the input of deep model (Chen et al., [2017](#)). Another study proposed the approach of combining the features extracted from Gabor filters with deep features from CNN. Even if the training samples were limited, the CNN was able to extract features and update weight (Chen et al.,

[2017](#); Luan et al., [2018](#)). The classification results of the previous works are improved when the Gabor features are added to spectral features. The work by Pratiksha Hattikatti (Pratiksha Hattikatti [2017](#)) combined the LBP texture features with the deep features extracted by CNN for classification of the image. This also resulted in the improved classification accuracy. In our proposed method, the extracted deep features are combined with other features namely Gabor feature and RLBP to improve the classification accuracy.

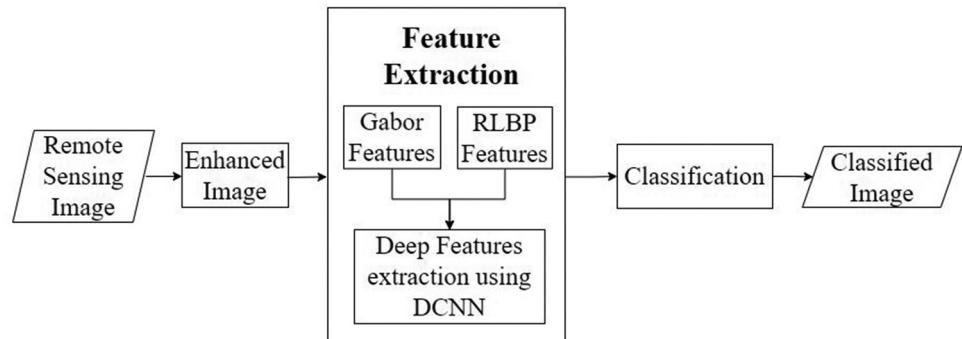
## Proposed work

### Overview

RS plays a major role in the trending research due to its application areas. Classification of the RS image results in revelling a lot of useful information. Thus, the proposed system tries to produce best classified image of the RS image. This section is separated into the following sub-sections: (1) Image Enhancement, (2) Feature Extraction and (3) Classification. In order to classify the RS images, the feature extraction is a crucial step. Earlier to the extraction, if the RS input image is enhanced using some enhancement techniques, then it will still more produce greater classification accuracy (Sharma et al., [2018](#)). The main idea of the proposed system is to create multiple set of features using different methodologies and combine those features into a single set of features. Since, the features extracted from DCNN is combined with other extra features, the classification accuracy is enhanced. The architecture of our system proposed is as shown in [Figure 1](#).

### Image enhancement

There is a lot of difference between the image which is simply classified and the image which is classified after enhancement. Before an image is classified, the enhancement of the image plays a major role. There are lot of methods existing to enhance the image. The earlier proposed method tries to apply bilateral filter to obtain the detail image (Kaplan et al., [2017](#)). But our



**Figure 1.** System architecture of the proposed system.

proposed system tries out an efficient method of enhancing the image. The proposed method applies the combination of filter, namely Gaussian filter, bilateral filter and median filter. The three main objectives for the image enhancement are as follows: (1) increase contrast, (2) emphasise edges and (3) preserve colour. The input RS image is decomposed into L number of detail image with L detail layer. To get the detail image, the first step is to construct the approximation image using Equation (1).

$$A_1 = \text{FILT}(I) \quad (1)$$

Here, the  $A_1$  represents the approximation image and I the original input image. The function  $\text{FILT}$  represents the filters applied. Using the approximation image, the further detail layer is exploited by finding the difference between the input and approximation image. Equation (2) represents the derivation for the detail layer.

$$D_1 = I - A_1 \quad (2)$$

The same idea is implemented to higher layers for decomposition. The same filters are applied to the approximation layers. The higher level of detail image is obtained using Equations (3) and (4).

$$A_l = \text{FILT}(A_{l-1}) \quad (3)$$

$$D_l = A_l - A_{l-1} \quad (4)$$

The enhanced image is obtained by adding the final level approximation layer output to every detail layer output. The detail layer outputs are weighted to give additional importance. The enhanced image is formulated as in Equation (5).

$$\text{Enhanced\_}I = A_l + \sum_{l=1}^L (w_l \cdot D_l) \quad (5)$$

The weight applied,  $w_l$  value is 2.

### Feature extraction

The proposed method uses three steps for extracting features, namely (i) feature extraction using Gabor filter, (ii) feature extraction using RLBP method and (iii) deep feature extraction using DCNN. The methods are explained as below:

### Feature extraction using gabor filter

Gabor filter is a multi-scale and multi-directional filter. They are used to capture spatial domain's frequency and orientation representation. To extract informative features, different frequencies and different directions are applied. Gabor filter with parameters, namely  $f$  &  $\theta$ , indicates frequency and orientation, respectively, and they are defined using Gaussian kernel function. It analyses any specific frequency content in particular direction, especially in a localised area (i.e.) around point or region. The two-dimensional Gabor filter is given by the following Equations (6) and (7):

$$G_c[i, j] = Be^{\frac{-(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i \cos\theta + j \sin\theta)) \quad (6)$$

$$G_s[i, j] = Ce^{\frac{-(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i \cos\theta + j \sin\theta)) \quad (7)$$

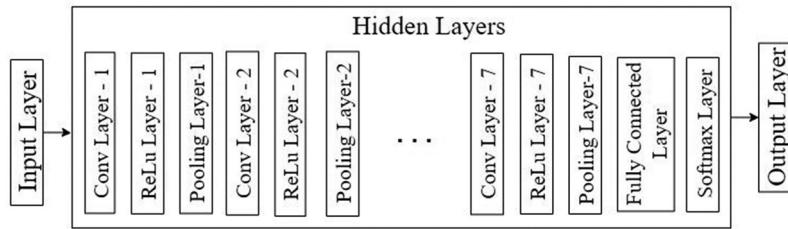
Here,  $G_c$  and  $G_s$  are gaussian function with cos and sin functions, respectively. B and C denotes the normalising factors. The texture variation in particular direction is obtained by varying  $\theta$ . Also, "i" and "j" represent the vertical and horizontal dimensions of the Gaussian kernel which convolute around the image, and  $\sigma$  represents the standard deviation of the Gaussian distribution.

### Feature extraction using RLBP method

The proposed system also extracts features using a new method which works similar to LBP method. Pratiksha Hattikatti (2017) used LBP method for feature extraction and combines those features with CNN features for classification. The proposed method makes small modification to the LBP method and tries to extract feature from the image using the following steps: (1) For a single pixel, consider all the neighbour values of that pixel, as shown in Figure 2 (a), and construct a neighbour pixel matrix (NPM) which is a  $3 \times 3$  matrix. Now the NPM consists of 9 values. (2) Find the mean value of the NPM. (3) Compare each value in the NPM with the mean value and construct another binary NPM (BNPM), as in Figure 2(b). The BNPM is formed by replacing "1" if the pixel exceeds mean and by replacing "0" if the pixel does not exceed mean. (4) Construct a reference matrix (RM) suggested by Pratiksha Hattikatti (2017). The matrix is as shown in Figure 2(c). The matrix contains  $2^x$  values, where  $x = 0, 1, 2 \dots 7$ . (5) Now, Hadamard product is computed

<table border="1"> <tr><td>7</td><td>4</td><td>3</td></tr> <tr><td>6</td><td>X</td><td>9</td></tr> <tr><td>1</td><td>3</td><td>2</td></tr> </table>	7	4	3	6	X	9	1	3	2	<table border="1"> <tr><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>X</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	1	1	0	1	X	1	0	0	0	<table border="1"> <tr><td>1</td><td>2</td><td>4</td></tr> <tr><td>128</td><td>X</td><td>8</td></tr> <tr><td>64</td><td>32</td><td>16</td></tr> </table>	1	2	4	128	X	8	64	32	16	<table border="1"> <tr><td>1</td><td>2</td><td>0</td></tr> <tr><td>128</td><td>X</td><td>8</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	1	2	0	128	X	8	0	0	0
7	4	3																																					
6	X	9																																					
1	3	2																																					
1	1	0																																					
1	X	1																																					
0	0	0																																					
1	2	4																																					
128	X	8																																					
64	32	16																																					
1	2	0																																					
128	X	8																																					
0	0	0																																					
(a)	(b)	(c)	(d)																																				

Figure 2. DCNN model of the proposed system.



**Figure 3.** (a) NPM matrix (b) BNPM matrix (c) RM matrix (d) resultant matrix.

between two matrices, namely BNPM and RM, and the resultant matrix is shown in Figure 2(d). Unlike LBP, the RLBP works different in construction of BNPM matrix by considering the mean value of the pixels. In LBP, the central pixel is used for comparison and is directly used as threshold. So, it is sensitive to noise Sudha Sree (2015). So, the proposed system tries to consider the neighbouring pixels also in addition to the central pixel. So that, the texture feature extraction becomes more accurate for RS images.

The values present in Figure 2(d) are added together and a single value is obtained, which represents the feature value for that particular pixel. Figure 2 shows a sample explanation for single pixel and its neighbouring pixels. In the similar manner, the calculation is done for the whole image.

#### Deep feature extraction using DCNN

**Convolutional neural network.** The deep features are extracted using CNN. The CNN is a deep neural network, which can effectively extract deep features from the image without any human intervention. The CNN usually initialises random filter values. The first layer in CNN is convolution. The purpose of convolution layer is extracting features from input RS image and obtaining output feature maps by applying convolution operation. The input image is convolved with matrix consisting of filter values. Each pixel value on the input gives an output value after convolution. The necessary details from the image are extracted. The resultant matrix can now be referred as convolution feature maps. The filters help in extracting features. After convolution, the pooling is done. The purpose of the pooling layer is reducing the parameters and compressing data for preserving important features. The dimensionality of the convolved features is reduced by applying pooling. It is of different types, namely max, average, sum and so on. Here, the system goes with max pooling. Our system tries to extract efficient features by increasing the CNN layers to extract efficient features.

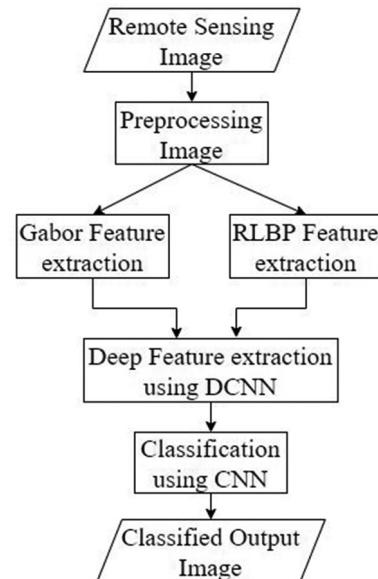
**Initialising filters.** The CNN usually runs with the random initialised filter values. But the proposed system introduces novelty and improves the efficiency of the overall system by manually initialising

the filter values. The filters such as mean, disc, log, prewitt filter for both vertical and horizontal edges, Gaussian, Laplacian, average, Sobel and Scharr for detecting both vertical edges and horizontal edges are used in our DCNN.

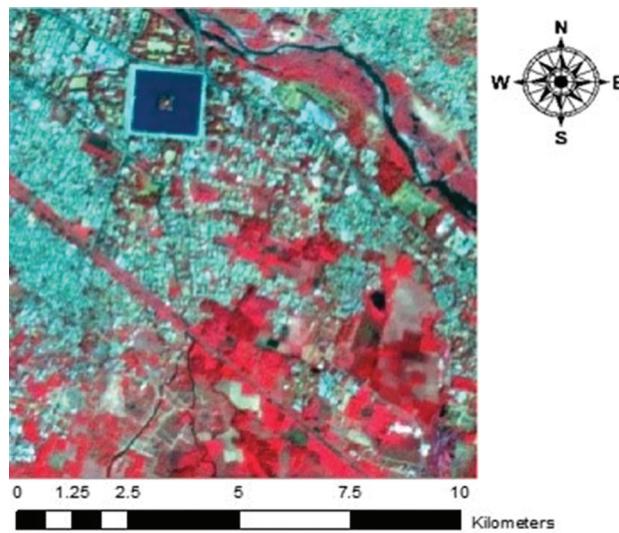
**Deep feature extraction using DCNN.** The proposed system considers the extracted Gabor features and RLBP features as input to the DCNN. The combined input features are thus fed into the convolutional layer of the first layer. The DCNN layer consists of seven layers. Each layer comprises convolutional layer followed by ReLu layer, which is then followed by pooling layer. The DCNN model is as shown in Figure 3. The final pooling result represents the deep feature.

#### Combined feature set

The proposed system provides Gabor and RLBP features as input to DCNN model. So, the feature extraction process results in the combined feature set with the benefit of deep feature, Gabor feature and RLBP feature. The feature here is directly sent for the classification step. The schematic view of the feature combination is as shown in Figure 4.



**Figure 4.** Schematic view of the functionality of the proposed method.



**Figure 5.** LISS IV image of Madurai city.

### Classification

The classification of the RS image is done using the feature subset obtained. The proposed system doesn't apply any different classifying techniques rather classifies the RS image using the CNN itself. The DCNN architecture contains three classification layers, namely FC layer, softmax layer and classification layer. The FC layer is used as a bridge to connect two layers which are preceding and following the FC layer. It connects neuron in one layer to every neuron in other layer. The next is softmax layer, and the purpose of this layer is to assign probability values to each class. The number of neurons present in the softmax layer will be equal to the number of neurons in the output layer. Our system uses Gaussian activation function in this layer and is represented in Equation (8)

$$f(x) = e^{-x^2} \quad (8)$$

The final layer of the network is output layer. Here, the neuron with the highest activation would be the model's predicted class. The results of the proposed system are discussed in detail in Section 4.

### Experimental results and discussions

#### Dataset description

The effectiveness of the proposed system is proved using the LISS IV image of Madurai region and is shown for the reference in Figure 5. Indian RS Satellite P6 version (IRS P6 Satellite) is utilised for LISS IV data acquisition. The spatial resolution of LISS IV image is 5.8 m. The spectral resolution

of red band is 0.62–0.68  $\mu\text{m}$ , green band is 0.52–0.59  $\mu\text{m}$  and near infrared band is 0.76–0.86  $\mu\text{m}$ . The LISS IV image is given in Figure 5. The input image shows the city view of Madurai located in the Tamil Nadu state belonging to India. The land cover of the city is approximately 23.5 km  $\times$  23.5 km. The city is found with longitude of  $78^{\circ}04'47''\text{E}$  to  $78^{\circ}11'23''\text{E}$  and latitude of  $9^{\circ}50'59''\text{N}$  to  $9^{\circ}57'36''\text{N}$ . The land cover includes mainly of urban area, vegetation, saline land, waste land and water area.

#### Architecture of the proposed system

Our proposed system is implemented using MATLAB 2020a. The system specification of the implementation includes Windows 10 with Intel(R) Core(TM) i5 1.80 GHz CPU and 8 GB RAM. As the first step, the image is enhanced. The Gabor and the RLBP features are extracted. The Gabor filer is implemented with 0.2 MHz frequency and four orientations such as  $\theta = 0, 45, 90$  and  $135$ . The Gabor and RLBP features are given as input to the DCNN model. Then, deep features are extracted using the DCNN with the pre-defined filter values. The layers of the CNN are convolution, ReLu, max pool, FC, softmax and classification. The CNN contains seven layers. The convolution weight matrix is of  $3 \times 3$  pixel size with stride and padding of "1". The pooling layer uses filter of size  $2 \times 2$  pixel with stride "1" and padding "2". The pooling layer is then trailed by ReLu and classification layers. The DCNN specification is as shown in Table 1 along with its output

**Table 1.** DCNN specification of the proposed system (FS represents filter size, NF represents number of filters, S represents stride and P represents pooling).

Layer	Specification	Feature map size
Input Layer	Size: [7,7,3]	7 x 7 x 3
Convolution Layer – 1	FS: [3 3], NF: 20, S: 1 x 1, P: 1 x 1	7 x 7 x 20
ReLU Layer – 1	-	7 x 7 x 20
Pooling Layer – 1	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 20
Convolution Layer – 2	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 2	-	7 x 7 x 40
Pooling Layer – 2	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
Convolution Layer – 3	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 3	-	7 x 7 x 40
Pooling Layer – 3	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
Convolution Layer – 4	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 4	-	7 x 7 x 40
Pooling Layer – 4	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
Convolution Layer – 5	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 5	-	7 x 7 x 40
Pooling Layer – 5	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
Convolution Layer – 6	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 6	-	7 x 7 x 40
Pooling Layer – 6	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
Convolution Layer – 7	FS: [2 2], NF: 40, S: 1 x 1, P: 2 x 2	7 x 7 x 40
ReLU Layer – 7	-	7 x 7 x 40
Pooling Layer – 7	FS: [2 2], S: 2 x 2, P: 1 x 1	4 x 4 x 40
FC Layer	-	1 x 1 x 5
Softmax Layer	-	1 x 1 x 5
Output Layer	-	1 x 1 x 5

**Table 2.** Training options of DCNN.

Name	Value
Optimizer	Stochastic Gradient Descent with Momentum
Momentum	0.900
Batch size	32
Epochs	1000
Learning Rate	0.0001

feature size. The options for training deep neural network are given in **Table 2**. The deep features are thus obtained using the DCNN.

In DCNN, the train set refers to the subset of samples which are used for training, and the test set refers to the subset of samples which are used for testing the CNN model. Here, for training, the size of samples in training set is 7 x 7 pixel. Here, 250 samples (50 samples for each class) are considered for training and 500 samples are considered for testing. The sampling strategy followed here is random sampling. The 500 sample pixels chosen are 131, 216, 66, 59 and 4 pixels of urban, vegetation, water body, waste land and saline land samples, respectively. These 500 ground control points (GCPs) were used to make the reference data set for the assessment using e-Trex venture global positioning system device. Some of the GCPs are collected during the field survey of Madurai city.

To analyse the performance of extracted features, the different combinations of features are tested with the test sample.

## Experimental results

The input image is applied to the combination of filter to produce the enhanced version of the original image. The bilateral filter is used for noise removal, whereas it preserves the edges as well. The median filter is used for noise removal. The classification is improved when it handles the enhanced input. The input image is decomposed to L layers of detail image. The proposed system decomposed to 1000 layers, i.e. L =1000. The value of  $w_l$  is 2 in Equation 5. The results of the input image after applying image enhancement technique are shown in [Figure 6](#).

After the image enhancement, the pre-processed image is sent for feature extraction process. As a first step, the Gabor filter is used for extracting features. The Gabor filer is implemented with 0.2 MHz frequency and four orientations such as  $\theta = 0, 45, 90$  and  $135$ . Likewise, the RLBP is used for extracting features. Both of these features are sent as input to the DCNN for extraction of deep features. The depth of the CNN is chosen after certain trail runs. The results of which is shown as graph in [Figure 7](#). It is obvious from the graph results that as the layer depth increases, the classification accuracy is also increased. But, at some point, it reaches its peak value and then starts falling down. The highest classification accuracy is obtained when using seven layers in DCNN. So, the proposed work suggests using the seven-layer CNN architecture for its implementation. Using the layer specification mentioned in [Table 1](#), the proposed system is implemented. Number of Epoch should be properly chosen for training the DCNN. Based on the number of epochs also, the variation in classification accuracies is observed. The trail run with different number of epochs, and their corresponding classification accuracy is shown as a graph in [Figure 8](#). When there are 800 epochs for single run, the classification accuracy reaches its peak values and then starts to maintain it stability. So, it is recommended to specify 800 epochs for single run. Thus, the deep features are obtained with seven-layer CNN which could run for 800 and above epochs. The classified image of this proposed system is shown in [Figure 9\(E\)](#).

The classification results of different combination of feature set are shown in [Figure 9](#). The pre-processed image is given as input to the CNN model and the

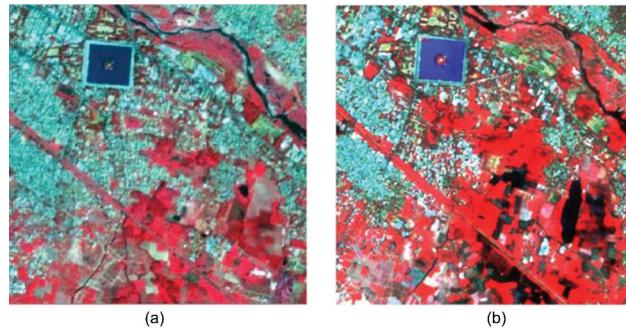


Figure 6 (a) Original LISS IV Madurai image. (b) Enhanced LISS IV Madurai Image.



Figure 7. Classification accuracies vs CNN layers.

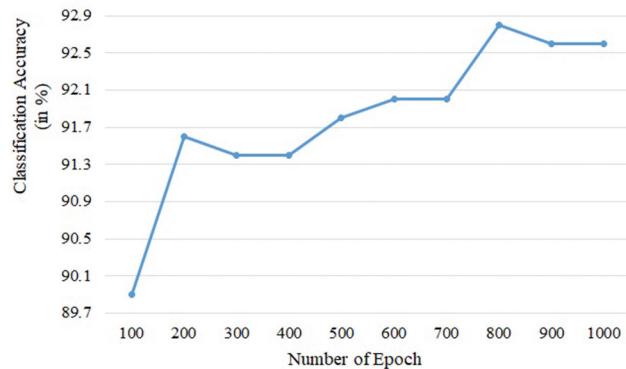


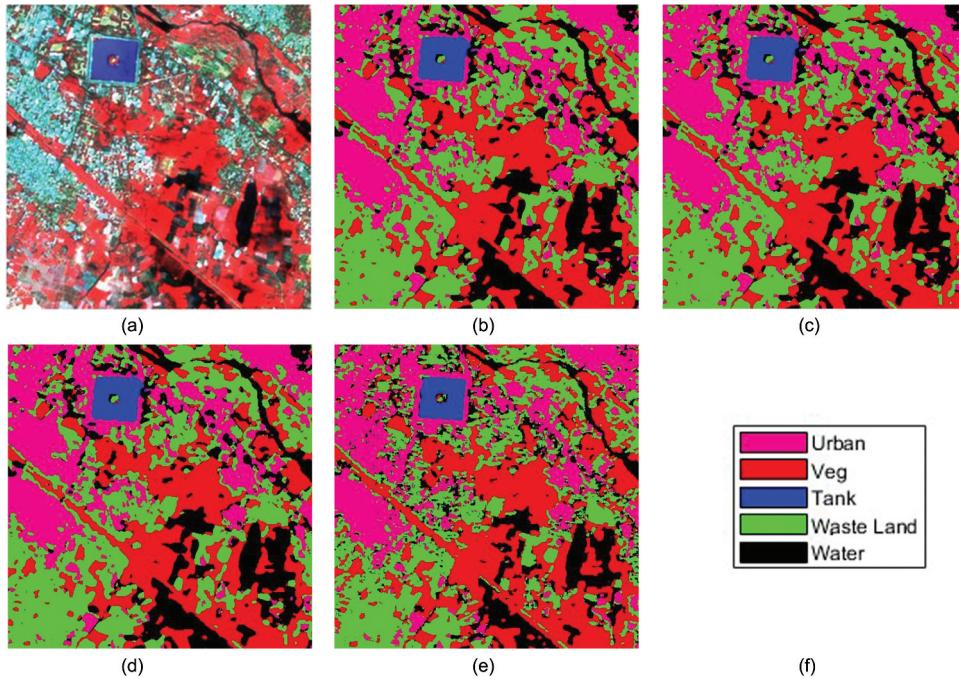
Figure 8. Classification accuracies vs number of epochs.

features are extracted. The classification results obtained are shown in Figure 9(b). When Gabor features are combined with the deep features obtained from CNN, a new feature set is obtained. Using those combined features, the classification is executed, and the results are shown in Figure 9(c). Similarly, the pre-processed image is used for extracting features with the help of RLBP. The RLBP features are combined with the already available deep features from DCNN. The classification results of which are shown in Figure 9(d). And finally, the classification result of our proposed system is shown in Figure 9(e). To show the local details of the classified image clearly, Figure 10 is added with the small area of the image zoomed in.

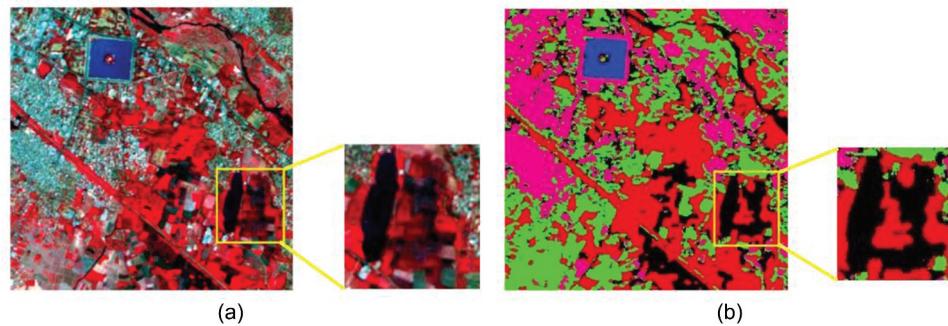
### Evaluation of results

The confusion matrix is used for expressing the accuracy of classification. The accuracy indices through which the proposed system is going to be evaluated are: user accuracy (UA), producer accuracy (PA), kappa coefficient (KC) and overall accuracy (OA). The accuracies are calculated using Equations (5)–(7). The UA, PA and KC are calculated using the formulas as given in Equations (5) and (6).

$$UA = \frac{\text{Number of correctly classified pixel}}{\text{Total no of pixels belong to a class}} \quad (9)$$



**Figure 9.** Output results of the proposed work for LISS IV images. (a) Original LISS IV Image. (b) Classified image with deep features. (c) Classified image with deep features and Gabor. (d) Classified image with deep features and LBP. (e) Classified image of our proposed method



**Figure 10** Output results of the proposed work for LISS IV images with small area zoomed in (a) original LISS IV image and (b) classified image of our proposed method.

$$PA = \frac{\text{Number of correctly classified pixel in each class}}{\text{Number of training pixels used for that class}} \quad (10)$$

KC is calculated with the help of field visit data. It ranges from values  $[-1, 1]$ . The value close to 1 represents the efficient classification and -1 represents worst classification. KC and OA is calculated using Equation (7) and (8).

$$KC = \frac{\text{Number of correctly classified pixels (with respect to GCP)}}{\text{Total number of pixels}} \quad (11)$$

$$OA = \frac{\text{Number of correctly classified areas}}{\text{Total number of reference areas}} \quad (12)$$

**Table 3.** Overall class-wise accuracy for LISS IV Madurai image using different feature extraction methods.

	CNN	CNN + Gabor	CNN+ LBP	Proposed Method
Urban	91.6	96.1	91.6	96.9
Vegetation	89.8	89.8	89.8	95.8
Water body	93.9	93.9	93.9	96.9
Waste Land	93.2	93.2	94.9	94.9
Saline Land	50	75	75	83

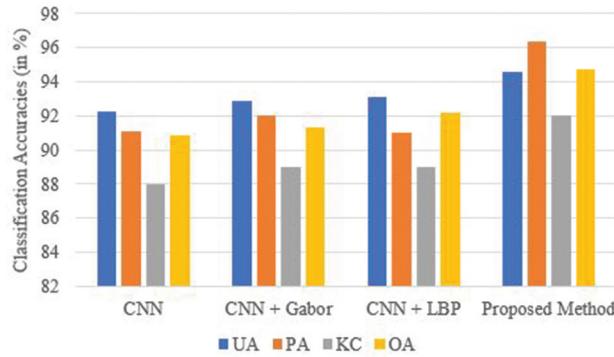


Figure 11. Comparison of UA, PA, KC and OA for different combinations of feature extraction methods.

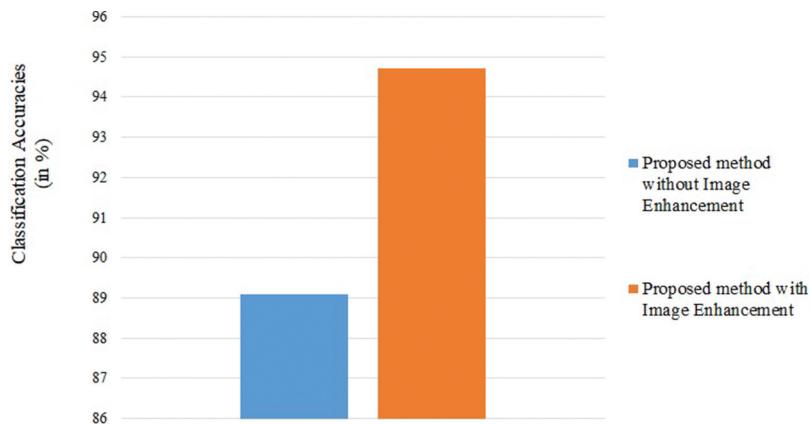


Figure 12. Comparison of classification accuracy of our proposed method with and without image enhancement.

Table 4. Classification accuracy based on different measures, namely UA, PA, KC and OA.

	UA	PA	KC	OA
CNN	92.28	91.1	0.88	90.9
CNN + Gabor	92.9	92.0	0.89	91.3
CNN + LBP	93.1	91.0	0.89	92.2
Proposed Method	94.6	96.4	0.92	94.7

Based on the above metrics, the classification evaluation is done and results are listed in Table 3 and 4. The correctness of the classified image is evaluated based on the direct field visit. To test the results, 500 samples

are taken for examination. Table 3 represents the class-wise accuracies using different feature extraction methods. Table 4 represents the classification accuracies based on different accuracy evaluation metrics for

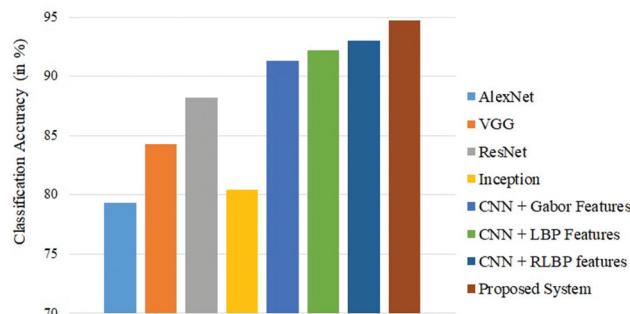


Figure 13. Comparison of classification accuracy of our proposed method with other earlier methods.

different feature extraction methods. The graphical representation to clearly understand the results is also given in Figure 11. And it clearly understood that the proposed method beats the performance of the other methods employed.

In order to prove the effectiveness of the image enhancement in increasing the classification accuracies, the accuracies of the method with and without Image enhancement are shown as a graph in Figure 12. From the graph, it can be interpreted clearly that the enhancement process before the feature extraction surely increases the classification accuracies. Also, the proposed system is compared with some of the existing methods like AlexNet, VGG, ResNet, Inception, CNN + Gabor and CNN + LBP. The comparison is done with respect to their overall classification accuracies and is shown in Figure 13. The graph results prove that our proposed method outperforms other state-of-the-art methods.

## Conclusion

The proposed system presented a new method of combining three different types of features namely deep features, Gabor feature and RLBP feature. The combined feature set is found to be very efficient for classification. A small drawback of obtaining many features while using three feature extraction method is prolonged training time for training the neural network model. But the greatest advantage of the system proposed is that it succeeds in finding out even the small land covers which is evident from the classification results. In future, it can still be improved by applying different classification methodologies and applying feature selection/reduction techniques.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Gladima Nisia T  <http://orcid.org/0000-0002-9752-5529>  
Rajesh S  <http://orcid.org/0000-0001-6435-157X>

## References

- Aptoula, E., Ozdemir, M. C., & Yanikoglu, B. (2016). Deep learning with attribute profiles for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1970–1974. <https://doi.org/10.1109/LGRS.2016.2619354>
- Chen, X., Xiang, S., Liu, C.-L., & Pan, C.-H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1797–1801. <https://doi.org/10.1109/LGRS.2014.2309695>
- Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2094–2107. <https://doi.org/10.1109/JSTARS.2014.2329330>
- Chen, Y., Zhu, L., Ghamisi, P., Jia, X., Li, G., Tang, L. (2017). Hyperspectral images classification with gabor filtering and convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2355–2359. <https://doi.org/10.1109/LGRS.2017.2764915>
- Fan, H., Xia, G.-S., Jingwen, H., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11), 14680–14707. <https://doi.org/10.3390/rs71114680>
- Guo, H., Shi, Q., Marinoni, A., Du, B., & Zhang, L. (2021). Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images. *Remote Sensing of Environment*, 264,(2021), 112589. <https://doi.org/10.1016/j.rse.2021.112589>
- He, K., X., S., R., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- He, K., X., Z., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE international conference on computer vision and pattern recognition (CVPR)*, 770–778. Las Vegas, NV, USA. arXiv:1512.03385.
- He, D., Shi, Q., Liu, X., Zhong, Y., & Zhang, X. (2020). Deep subpixel mapping based on semantic information modulated network for urban land use mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12), 10628– 10646, <https://doi.org/10.1109/TGRS.2021.3050824>
- Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015, Art.(258619), 1–12. <http://dx.doi.org/10.1155/2015/258619>
- Juefei-Xu, F., Naresh Boddeti, V., & Savvides, M. (2017). Local binary convolutional neural networks. *IEEE conference on computer vision and pattern recognition* Honolulu, HI, USA (IEEE), pp. 4284–4293.
- Kaplan, N. H., Erer, I., & Gulmus, N. (2017). Remote sensing image enhancement via bilateral filtering. *8th international conference on recent advances in space technologies (RAST)* (IEEE) Istanbul, Turkey, PP. No. 139–142.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lin, Z., Chen, Y., Zhao, X., & Wang, G. (2013). Spectral-spatial classification of hyperspectral image using autoencoders. *9th international conference on information, communications & signal processing (ICICS)* Tainan (IEEE).

- Liu, S., Shi, Q., & Zhang, L. (2020). Few-shot hyperspectral image classification with unknown classes using multi-task deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6), 5085–5102. <https://doi.org/10.1109/TGRS.2020.3018879>
- Luan, S., Chen, C., Zhang, B., Han, J., & Liu, J. (2018). Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9), 4357–4366. <https://doi.org/10.1109/TIP.2018.2835143>
- Ma, A., Wan, Y., Zhong, Y., Wang, J., & Zhang, L. (2021). SceneNet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172(2021), 171–188. <https://doi.org/10.1016/j.isprsjprs.2020.11.025>
- Mirzapour, F., & Ghassemian, H. (2015). Improving hyperspectral image classification by combining spectral, texture, and shape features. *International Journal of Remote Sensing*, 36(4), 1070–1096. <https://doi.org/10.1080/01431161.2015.1007251>
- Ouyang, W., & Wang, X. (2013). Joint deep learning for pedestrian detection. *Proceedings of the 2013 IEEE international conference on computer vision*. Sydney, NSW, Australia: IEEE. <https://doi.org/10.1109/ICCV.2013.257>
- Pires de Lima, R., & Marfurt, K. (2020). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing*, 12(86), 1–20. <https://doi.org/10.3390/rs12010086>
- Pratiksha Hattikatti (2017). Texture based interstitial lung disease detection using convolutional neural network. *2017 international conference on big data, iot and data science (Bid)* Pune, India, pp 18–22, Vishwakarma Institute of Technology.
- Sellami, A., & Tabbone, S. (2022). Deep neural networks-based relevant latent representation learning for hyperspectral image classification. *Pattern Recognition*, 121(C), 108224. <https://doi.org/10.1016/j.patcog.2021.108224>
- Sharma, V., Diba, A., Neven, D., Brown, M. S., Van Gool, L., & Stiefelhagen, R. (2018). Classification-driven dynamic image enhancement. *IEEE/CVF conference on computer vision and pattern recognition*. Salt Lake City, UT, USA: IEEE. <https://doi.org/10.1109/CVPR.2018.00424>
- Shi, Q., Liu, M., Shengchen, L., Liu, X., Wang, F., & Zhang, L. (2021a). A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE transactions on geoscience and remote sensing*, 60, 5604816. <https://doi.org/10.1109/TGRS.2021.3085870>
- Shi, Q., Tang, X., Yan, T., Liu, R., & Zhang, L. (2021b). Hyperspectral image denoising using a 3-D attention denoising network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12), 10348–10363. <https://doi.org/10.1109/TGRS.2020.3045273>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *IEEE international conference on learning representation (ICLR)*. Kuala Lumpur, Malaysia: IEEE.
- Song, C., Yang, F., & Li, P. (2010). Rotation invariant texture measured by local binary pattern for remote sensing image classification. *2010 second international workshop on education technology and computer science*. Wuhan, China: IEEE., <https://doi.org/10.1109/ETCS.2010.37>
- Sudha Sree, Ch. (2015). Survey on extraction of texture based features using local binary pattern. *International Journal of Engineering Research & Technology (IJERT)*, 4(7), 334–338. <https://www.ijert.org/research/survey-on-extraction-of-texture-based-features-using-local-binary-pattern-IJERTV4IS070328.pdf>
- Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. *Proceedings of the 27th international conference on neural information processing systems*, Montreal, QC, Canada (Cambridge, MA, United States: MIT Press), pp. 1988–1996.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *IEEE international conference on computer vision and pattern recognition (CVPR)*. Boston, MA: IEEE.
- Tan, J., Gao, Y., Cao, W., Pomeroy, M., Zhang, S., Huo, Y., Lihong, L., & Liang, Z. (2019). GLCM-CNN: Gray level co-occurrence matrix based CNN model for polyp diagnosis. *IEEE EMBS international conference on biomedical & health informatics (BHI)*. Chicago, IL, USA: IEEE. <https://doi.org/10.1109/BHI.2019.8834585>
- Wanli, O., Xiaogang, W., Xingyu, Z., Shi, Q., Ping, L., Yonglong, T., Hongsheng, L., Shuo, Y., Zhe, W., Chen-Change, L., Xiaou, T. (2015). Deepid-net: Deformable deep convolutional neural networks for object detection. *Proceedings of the IEEE conference on computer vision. Pattern Recog*. Boston, MA, USA, pp. 2403–2412.
- Yang, Y., & Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of ACM GIS* San Jose, CA, USA: Association for Computing Machinery, pp. 270–279.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27, 3320–3328. <https://arxiv.org/abs/1411.1792>
- Yue, J., Zhao, W., Mao, S., & Liu, H. (2015). Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6), 468–477. <https://doi.org/10.1080/2150704X.2015.1047045>
- Yue, L., Zhou, W., Tian, Q., Sun, S., & Houqiang, L. (2018). Retrieval oriented deep feature learning with complementary supervision mining. *IEEE Transactions on Image Processing*, 27(10), 4945–4957. <https://doi.org/10.1109/TIP.2018.2845120>
- Zhang, N., Donahue, J., Girshick, R., & Darrel, T. (2014). Part-based RCNNs for fine-grained category detection. *European conference computer vision*, Zurich, Switzerland: Springer, Cham.
- Zhang, F., Du, B., Zhang, L., & Xu, M. (2016b). Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(9), 5553–5563. <https://doi.org/10.1109/TGRS.2016.2569141>
- 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. *International Journal of Remote Sensing*, 36(13), 3368–3379. <https://doi.org/10.1080/2150704X.2015.1062157>
- Zhao, W., Du, S., Emery, W. J. (2017). Object-based convolutional neural network for high-resolution imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sens*, 10(7), 3386–3396. <https://doi.org/10.1109/JSTARS.2017.2680324>