Automatic discriminative feature extraction using Convolutional Neural Network for remote sensing image classification

Akhtar Jamil (1), Bulent Bayram (2)

[1] Department of Computer Engineering, Istanbul Sabahattin Zaim University, Istanbul,34303, Turkey
[2] Department of Geomatic Engineering, Yildiz Technical University, Istanbul, 34220, Turkey
Email: akhtar.jamil@izu.edu.tr; bayram@yildiz.edu.tr

KEY WORDS: Deep learning, convolutional neural networks, spectral-spatial feature, random forest

ABSTRACT: Supervised approaches require selection of training samples from all classes and then a set of highly discriminative feature descriptors are extracted to represent each class. Traditional classification methods employee handcrafted features. Although, such features are effective but usually require prior knowledge and involve lot of laborious work. In addition, the availability of less training samples for multi/hyperspectral data makes the problem even more challenging. An alternative solution could be to employee deep learning-based approach for automatic extraction of highly stable feature patterns from input data. This paper proposes a new method using a deep learning based on Convolutional Neural Network (CNN) for automatic extraction of spectral-spatial features from high-resolution multi-spectral images. The learning framework consisted of a series of convolutions and pooling layers. We evaluated the effectiveness of the proposed method for the problem of land cover classification. The dataset consisted of 10 high-resolution multi-spectral images obtained from Rize district of Turkey. The classification was then performed by applying the random forest classifier. The results indicated that the proposed method was effective and easier to implement and learn.

1. INTRODUCTION

The availability of remote sensing data has made it possible to develop various application in a cost effective manner. Specifically, multi/hyperspectral images are used for various tasks such as land cover classification [1], tree species extraction [2], shoreline extraction [3], object detection [4] etc. They can be used in various domains such as agriculture, military, environmental science [5]. Although multispectral images not provide rich spectral information compared to the hyperspectral images, yet they provide crucial reflectance information which can be used to distinguish between various objects of interest. To obtain such information, various approaches have been proposed. These proposed methods were mostly dominated by object-based and machine learning approaches. However, in the recent times it has been noticed that the deep learning-based approaches have dominated other machine learning methods for various problem including remote sensing data classification. Their success is attributed to their ability to produce better results on large scale data and they have the ability to extract highly discriminative features automatically from input images which increase the classification accuracy of the classifier.

Machine learning approaches, on the other hand, have been used commonly to derive useful information from high resolution remote sensing data. For instance, [6], [7] employed support vector machine for land cover classification. [8] integrated the results of SVM, ANN and RF for classification of tree species and land cover classes from high resolution images. Further, in [9] authors used Pleiades-1A images for extraction of green space areas, while [10] used aerial imagery for land cover classification using machine learning approaches. Similarly, in [11] the performance of multi-class SVMs is compared with maximum likelihood and ANN classifiers for remote sensing data classification. [12] employed DTs for classification of land cover classes and dominant trees from LiDAR data. Further details about widely used supervised machine learning classifiers for classification of remote sensing data can be found in [13].

Without a doubt ML methods have produced promising results for high-resolution image classification [8]. However, their ability is limited due to dependence on the handcrafted feature extraction. These low-level features such as gray level cooccurrence matrix, spectral histograms, histogram of oriented gradients, may not be able to truly represent the classes of interest [14]. Particularly, for remote sensing images, which are more challenging in terms of feature extraction and classification due to presence of heterogeneous objects with similar spectral information, handcrafted features hider the capabilities of the machine learning-based approaches. Guided by such observation, automatic feature extraction using deep learning approaches can be more effective compare to the hand-crafted features in terms of robustness and processing time.

As mentioned that the deep learning based methods can help extract more robust and abstract feature representation that improve the classification accuracy of the classifier [15]. Applying unsupervised feature learning can reduce the time complexity of the feature engineering step and also highly discriminative features can be obtained [16]. Recently, deep learning-based approaches received considerable attention, and the remote sensing community has also adopted deep learning-based model for classification of remote sensing images. For instance, in [14] authors proposed sparse auto-encoders (SAE) model for feature representation and a multi-level method for land-use classification from high resolution images. In [17] compared the effectiveness of three different feature fusion methods including, voting, weighted averages and fuzzy integrals. The main focus of their work was on fusion at classification level and employed three multiple deep convolutional neural networks by training three different approaches: CaggeNet, GoogLeNet and ResNet5. In [18] authors used deep recurrent neural networks, particularly long short term memory (LSTM) for classification of land covers from multitemporal spatial data. The results presented by various researchers show the success of the deep learning based approaches in remote sensing data classification.

In this paper, our primary focus is to eliminate the human involvement for designing handcrafted features by applying deep learning for automatic extraction of highly discriminative features from input images. We designed a network based on CNN to derive features automatically from the high-resolution images. The learned features were then fed into an RF classifier that produced thematic maps for each class. The classification results are then compared with the reference data for evaluation.

2. MATERIALS

Rize is a district in the Trabzon region in south-eastern part of Turkey. We selected it as our study area for obtaining images and further analysis for land cover classification. Generally, this area contains many agricultural regions. However, the city is expending yearly and overall the structure of the city is rapidly changing due to addition of manmade objects. Tracking such changes is challenging but very important for various decision-making processes. Remote sensing techniques provide a cost-effective solution to deal with this challenge. Figure 1 (a) shows the map of turkey and location of Rize district, while (b) shows samples of main land cover classes considered in this study (tea gardens, other trees types, bare land and impervious surfaces).

The images were obtained by UltraCamX digital aerial camera coupled on a plane on March 27-28, 2013. The obtained images were further processed by EMI Group Inc. Turkey for removal of noise and to make some necessary corrections. The final product, which were received in the form of the four band ortho images and their corresponding reference data. The ortho images consisted of four spectral bands namely: red, green, blue and near infrared. We used ten ortho images for evaluation. In addition, we considered four main classes of interest for land cover classification: tea gardens, other trees, bare land and impervious surface. Table 1 summarizes the technical specification of the aerial camera system used to capture the images.

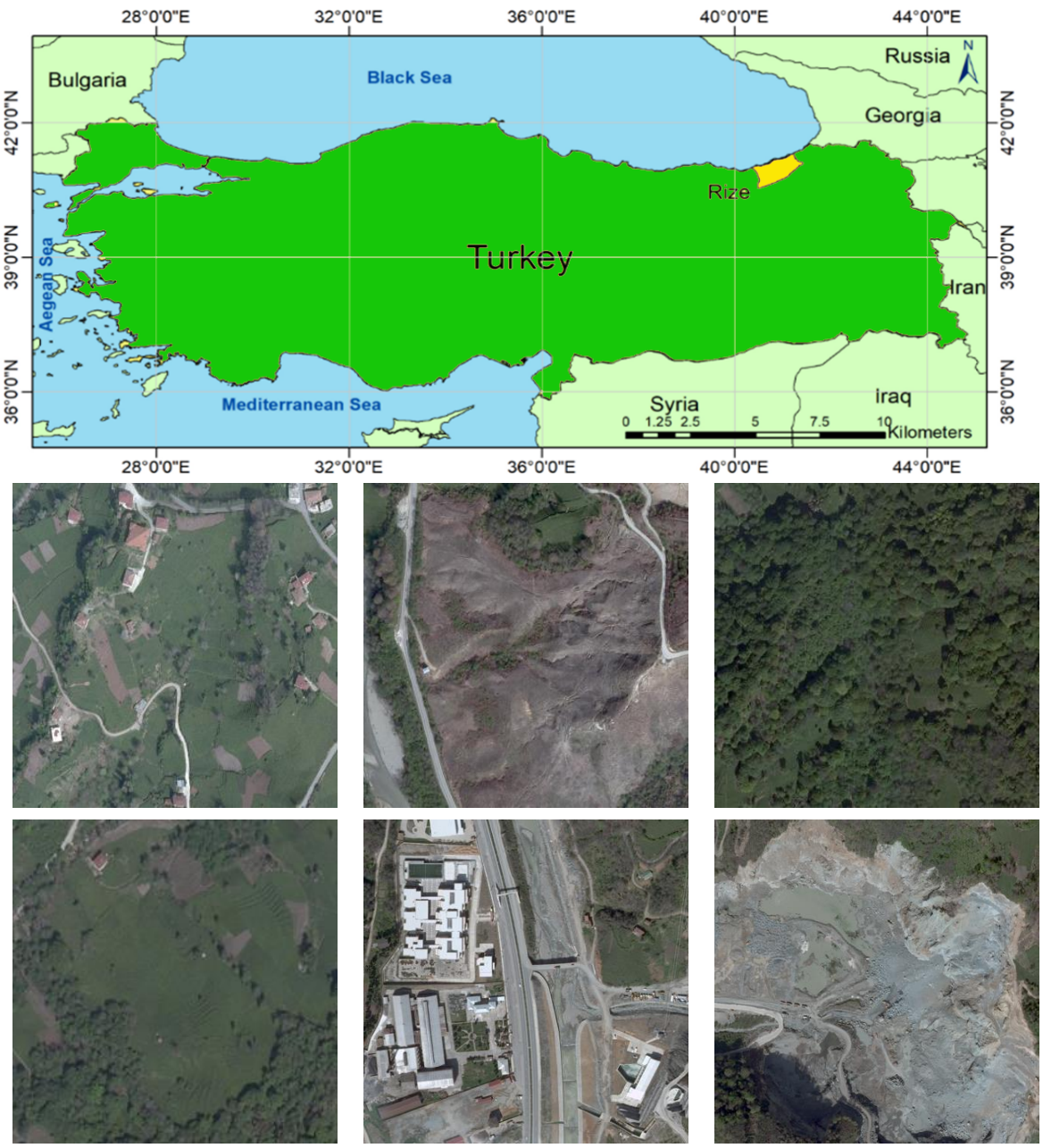


Figure 1 a) Map of Turkey showing location of the study area (Rize district) b) Shows few samples with different land cover classes (tea gardens, other trees, bare land, impervious surface)

Table 1: Technical Specification of UltraCamX

Parameter Name	Value
Ground Sample Distance	30 cm
Capture Dates	13-03-2013
Side Laps	30%
End Laps	70%
Flight Altitude	4200 m
Sensor Type	Aerial
Radiometric resolution	8 bits

3. PROPOSED METODOLOGY

The proposed methodology is a combination of an automatic feature extractor and a supervised classifier for classification of four land cover classes from high resolution imagery data. The workflow of the proposed method is shown in figure 2. As it is shown that CNN is used as feature extractor, therefore we applied sequence of convolutions and maxpooling operations to obtain the feature maps. These feature maps are passed to RF classifier for classification. Finally, we evaluate the proposed method by comparing the output with the ground truth data. The following sections describe the basics of CNN architecture first and its application for feature extraction followed by RF based classification.

3.1 CNN Architecture

Inspired by the human visual system, which can perform detection and recognition tasks very effectively, networks with relatively deeper (two or more) layers can be effectively used for performing similar task with higher accuracy. CNN is a deep learning model which has two special aspects: local connections and shared weights. These feature tend to provide better generalization capabilities [19].

The core component of CNN architecture includes convolution and pooling layers. A deep CNN can be constructed by stacking several such layers. The input data is convoluted with several kernel filters which are then passed to the pooling layer. The pooling layer reduces the feature variance to detect more robust features. Generally, a fully convolutional layer is attached to the end of the architecture followed by a softmax classifier for classification.

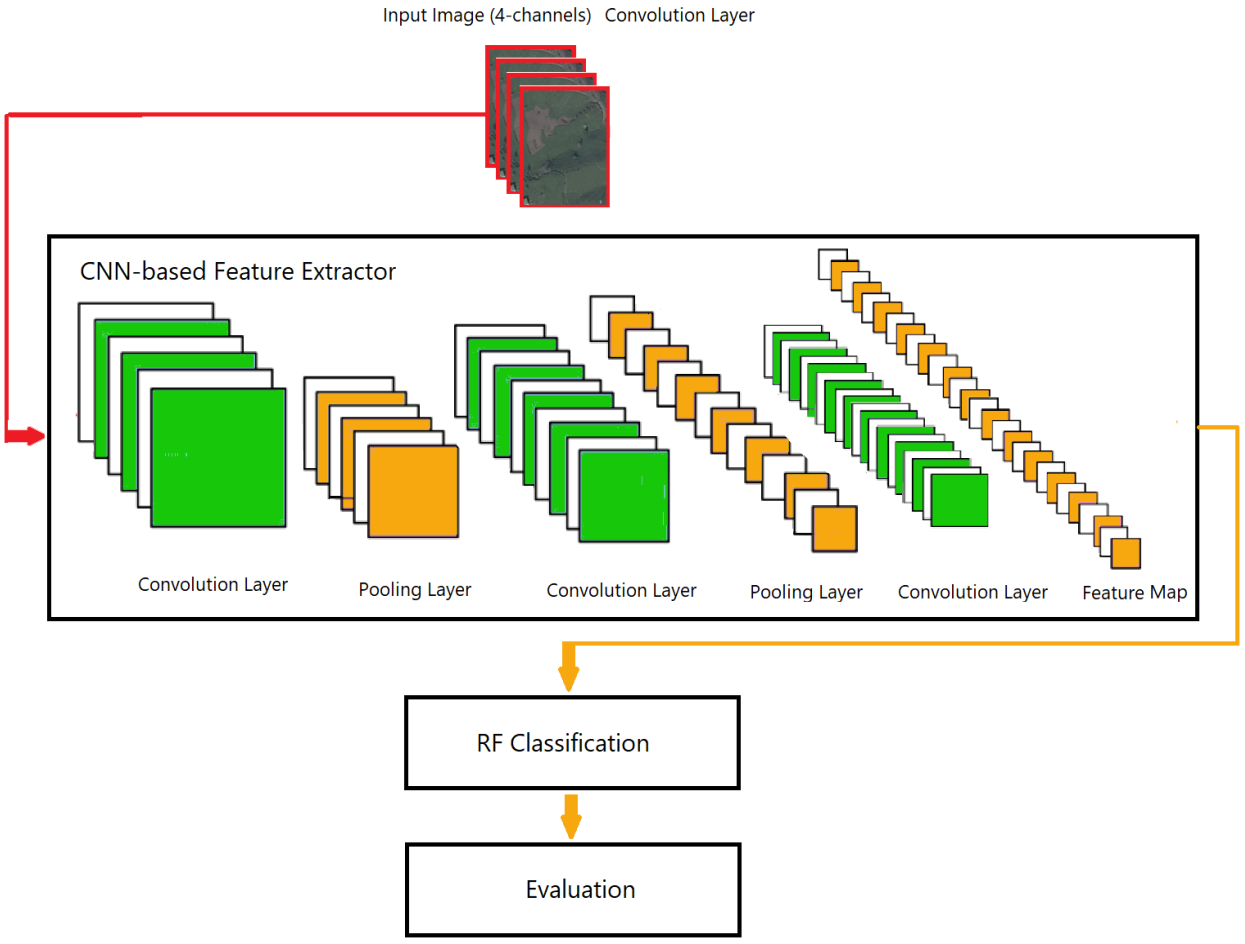


Figure 2 The schematic diagram of the proposed method

3.2 RF for Classification

The RF is a supervised classification algorithm which widely used in various classification tasks. This algorithm was proposed by Leo Breiman and Adèle Cutler in 2001 [20]. The random forest algorithm is based on a set of decision trees. These different trees are characterized by the same number of nodes, but different data. The decisions of these different decision trees will be combined to give a final answer that represents an average response of all these decision trees. The classifier has also shown it efficiency for classification of land covers e.g. [21]

The input images, which contains four spectral bands, are passed though the CNN based feature extractor to automatically derived features from the input image. Along with these features we passed the labels into the RF classifier during training. Finally, the classification maps were generated on the test data which were then compared with the reference data for evaluation. The overall schematic diagram of the proposed method is shown in figure 2.

4. EXPERIMENTAL RESULTS

Here, we validate the proposed method by conducting experiments on the chosen data set. The results were quantitatively compared using overall accuracy and kappa coefficient performance measures. We split the labeled samples into two training and test samples. During the training procedure, 90% of the training samples were used to learn weights and biases of CNN while the remaining 10% of the training samples were used to validate the architecture.

The experiments were performed on a standard PC with a processor clock rate of 2.50 GHz, and a RAM capacity of 16 GB. All coding was written in the MATLAB R2019a environment. In addition, a ground truth data preparation utility software was written in C# windows forms application using Emgu CV library. The utility application produced reference data area in the form of XML format which contained vertices of the polygons of the selected regions.

4.1 Selection of Training samples

The training samples were selected from 5 images (50%) of the total sample from data set. These were manually selected by visual inspection to best describe the each class. A balanced training samples were selected for each class to avoid any possible bias towards higher sample classes. All these samples were resized to match the expected input size of the feature extractor (16 x 16). For each class 290 samples were selected from five randomly selected ortho images. Figure 3 shows a sample ortho image from where the training samples have been selected. To reduce the processing time, we considered the impervious surfaces and bare land as one class which is termed as bare land.

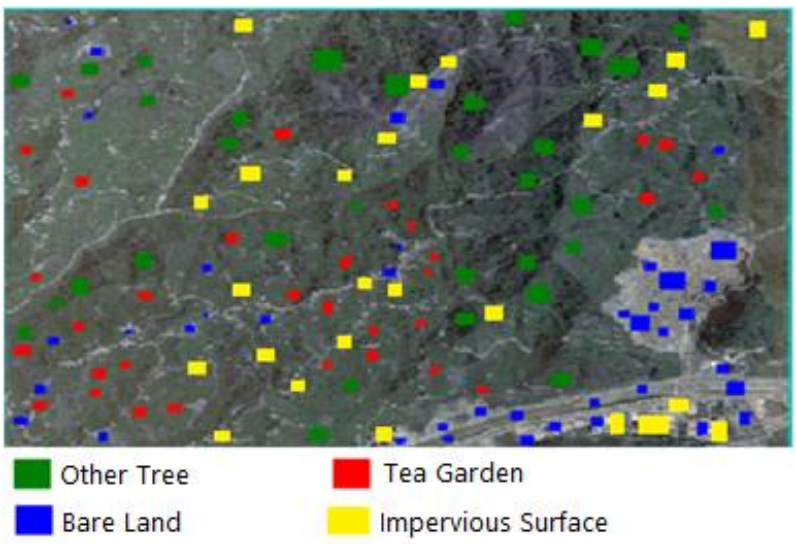


Figure 3. Selection of training samples

4.2 CNN-based Feature Extraction

To capture the spatial context for each pixel, the high-resolution images were divided into 16 x 16 pixels patches around each centered pixel. This patch size produced optimal accuracy for our classification results. A feature vector was obtained for each patch by applying a shallow CNN based feature extractor. This feature extractor was a combination of convolutional layers followed by max pooling layer. Three such intermittent layers were considered in our experiments. As we know that a deeper network requires higher amount of training data, however, to avoid the manual intensive work of preparing large training data, we employed a shallow network of three layers which could extract features that provide enough spatial contextual information for classification. These features were then classified by the random forest classifier into one of the four classes. Table 2 summarizes the architectural setting of the CNN feature extractor.

The size of the input layer of CNN was set to 16 x 16 x 4 since 16 ×16 pixel neighborhoods of current pixel were selected as input to CNN for feature extraction. Here 4 indicates the number of bands in the ortho images. The size of the filter bank for each layer was set to 3 while the number of filters applied at each layer vary. The number of filters applied at first, second and third layers were 4, 8 and 16 respectively. Also, for each maxpooling layer, stride was fixed to 1. We also applied batch normalization after every convolution layer to speed up the learning. The rectified linear unit (ReLU) was used as non-linear activations for each neuron.

Before starting the training process, the initial learning rate was set to 0.01. Stochastic gradient descent with moment (sgdm) optimization algorithm was used for training the networks. Also, the maximum number of epochs was set to 100. We used CNN with L2 regularization for feature extraction. During the training, we employed back-propagation to learn the weights and biases using RF classifier. The output obtained at the ReLU layer was used as feature vector for training and testing the RF classifier.

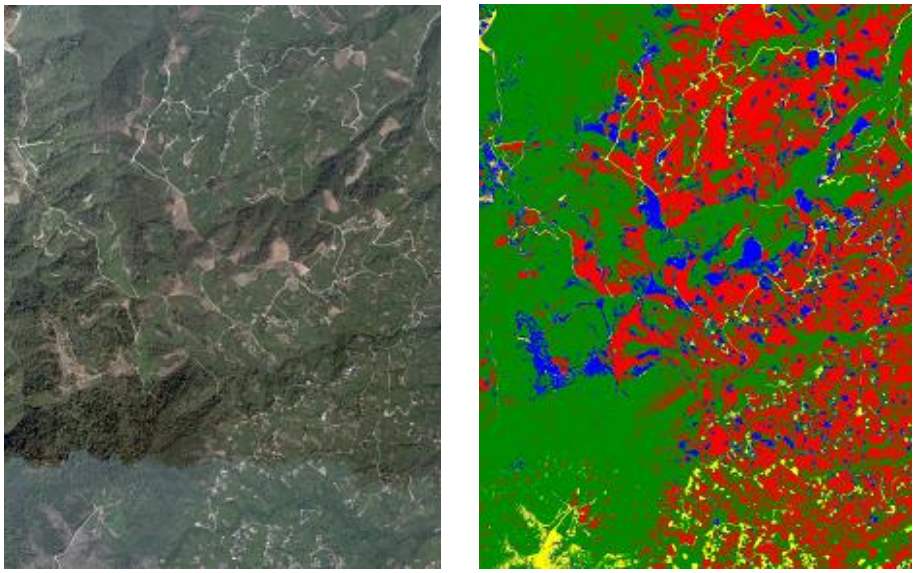
Table 2 CNN Architecture setting

Layer	Convolution	Activation	Max Pooling	Stride
1	3 x 3 x 4	ReLU	2 x 2	1
2	3 x 3 x 8	ReLU	2 x 2	1
3	3 x 3 x 16	ReLU	2 x 2	1

4.3 RF-based classification

Although there are many other parameters that can be finetuned for RF classifier, however, two most important parameters include the number of predictors (NP) and number of random trees (NT). The NP was obtained by cross validation and set to 75 while NT was set to 30.

The feature vectors obtained from the CNN feature extractor along with labels were taken as input to the RF classifier during training phase. Once the classifier was trained, the same set of features were input into the classifier for classification. The classifier retuned the label associated with the feature vector. We generated a classification map for each class in the form of a binary image where white regions represented the class of interest. The classification results obtained from RF for a sample ortho image are shown in figure 4.



(a) (b)

Figure 4. (a) Input image (RGB color space) (b) RF Classification result for four land covers

4.4 Accuracy Assessment

The results were quantitatively compared using overall accuracy and kappa coefficient as performance measures. Table 3 shows the experimental results obtained on our data set. It is shown that the CNN based features and RF classifier produced acceptable classification results for four land cover classes. The highest accuracy was obtained for impervious surface (90.22%) followed by bare land (87.36%). Even though there were some misclassifications between bare land and impervious surface. The overall accuracy for vegetation classes (tea garden and other trees) remained relatively lower than other two classes. The main reason could be the similarity between the two classes. In some situations, it was even difficult for human eye to distinguish between these two classes. In addition, recently planted teagardens visually looked similar to bare land. These were also misclassified into bare land classes as well which also effected the classification accuracy. In summary, the proposed method still exhibits a very good average accuracy for our data set.

Table 3 Classification accuracy for RF classifier

Class	Overall Accuracy (%)	Kappa
Tea Garden	82.40	0.77
Other Trees	83.71	0.76
Bare Land	87.22	0.81
Impervious Surface	90.36	0.85

5. ACKNOWLEDGEMENT

We are thankful to the EMI Group Turkey for providing this data so that we could conduct this study which is a part of TEYDEP Project entitled “Development of Object Based Neural Network Image Processing System Determination of Vegetation and Forestry Boundaries” (Project Nr. 7140512). It was supervised by EMI Group-Turkey, and consulted by Prof. Dr. Bulent Bayram.

6. CONCLUSION

Feature engineering is one of the main challenges of the traditional supervised machine learning based approaches. This is costly in terms of human effort and even sometimes may not produce the desired results. This problem was alleviated by extracting features automatically by applying

CNN. The advantage of such an approach was two-folded: 1) it reduced the human effort as it automatically extracts the features compared to handcrafted feature engineering, 2) as the input data passes down a large number of deep layers, highly abstract and robust features can be obtained that improved the classification accuracy for the classifier. Our feature extraction phase exploited spectral-spatial information to obtain robust features and then RF classifier was applied for classification of the highly discriminative features into one of the land cover classes. Overall good classification results were obtained, however, relatively lesser classification accuracy was obtained for vegetation classes, specifically tea gardens. This was ascribed to preparation of the reference data. In some cases, the reference data showed that recently planted tea gardens (which look apparently similar to bare land) have been marked as tea garden. However, our proposed method is unable to identify this and classified them into bare land class. We believe that this is neither an issue of the proposed method nor there is a fault in the generation of reference data. We believe that if this issue of recently planted trees is resolved then the accuracy can further increase.

References

[1] Y. Chen, Y. Ge, G. B. M. Heuvelink, R. An, and Y. Chen, "Object-Based Superresolution Land-Cover Mapping From Remotely Sensed Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 328–340, Jan. 2018.

[2] A. Jamil and B. Bayram, "The delineation of tea gardens from high resolution digital orthoimages using mean-shift and supervised machine learning methods," *Geocarto Int.*, vol. 0, no. 0, pp. 1–15, Jun. 2019.

[3] N. Demir, B. Bayram, D. Z. Şeker, S. Oy, A. İnce, and S. Bozkurt, "Advanced Lake Shoreline Extraction Approach by Integration of SAR Image and LIDAR Data," *Mar. Geod.*, vol. 42, no. 2, pp. 166–185, Mar. 2019.

[4] C. Chen, W. Gong, Y. Chen, and W. Li, "Object Detection in Remote Sensing Images Based on a Scene-Contextual Feature Pyramid Network," *Remote Sens.*, vol. 11, no. 3, p. 339, Feb. 2019.

[5] C. Li, Y. Wang, X. Zhang, H. Gao, Y. Yang, and J. Wang, "Deep Belief Network for Spectral–Spatial Classification of Hyperspectral Remote Sensor Data," *Sensors*, vol. 19, no. 1, p. 204, Jan. 2019.

[6] R. Seifi Majdar and H. Ghassemian, "A probabilistic SVM approach for hyperspectral image classification using spectral and texture features," *Int. J. Remote Sens.*, vol. 38, no. 15, pp. 4265–4284, 2017.

[7] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

[8] A. Jamil and B. Bayram, "Tree Species Extraction and Land Use/Cover Classification From High-Resolution Digital Orthophoto Maps," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 1, pp. 89–94, Jan. 2018.

[9] Zylshal, S. Sulma, F. Yulianto, J. T. Nugroho, and P. Sofan, "A support vector machine object based image analysis approach on urban green space extraction using Pleiades-1A imagery," *Model. Earth Syst. Environ.*, vol. 2, no. 54, p. 54, 2016.

[10] Y. Qian, W. Zhou, J. Yan, W. Li, and L. Han, "Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery," *Remote Sens.*, vol. 7, no. 1, pp. 153–168, 2015.

[11] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, Mar. 2005.

[12] T. Sasaki, J. Imanishi, K. Ioki, Y. Morimoto, and K. Kitada, "Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data," *Landsc. Ecol. Eng.*, vol. 8, no. 2, pp. 157–171, Jul. 2012.

[13] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: an applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp.

2784–2817, 2018.
[14] C. Lu, X. Yang, Z. Wang, and Z. Li, “Using multi-level fusion of local features for land-use scene classification with high spatial resolution images in urban coastal zones,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 70, no. September 2017, pp. 1–12, 2018.
[15] L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, 2016.
[16] A. Wang, J. Lu, G. Wang, J. Cai, and T.-J. Cham, “Multi-modal Unsupervised Feature Learning for RGB-D Scene Labeling,” in *Lecture Notes in Computer Science*, vol. 8693 LNCS, no. PART 5, 2014, pp. 453–467.
[17] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, “Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.
[18] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, “Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, Oct. 2017.
[19] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.
[20] A. Cutler, D. R. Cutler, and J. R. Stevens, “Random Forests,” no. February 2014, 2011.
[21] J. Stefanski, B. Mack, and B. Waske, “Optimization of Object-Based Image Analysis With Random Forests for Land Cover Mapping,” *Sel. Top. Appl. Earth Obs. Remote Sensing, IEEE J.*, vol. 6, no. 6, pp. 2492–2504, 2013.