Intercontinental Geoinformation Days

igd.mersin.edu.tr

Analysis of the effect of training sample size on the performance of 2D CNN models

Elif Ozlem Yilmaz[1], Taskin Kavzoglu*[1]

[1]Gebze Technical University, Faculty of Engineering, Department of Geomatics Engineering, Kocaeli, Turkey

Keywords
Remote sensing
CNN
Deep Learning
Image Classification
Sample Size

ABSTRACT

Hyperspectral remote sensing plays a significant role in the research of Earth observation owing to rich spectral information. Convolutional Neural Networks have been commonly used in hyperspectral image classification with the rapid development of deep learning algorithms. In this study, the effect of sample size on the performance of 2D CNN models was analyzed using freely available Pavia hyperspectral data for a 9-class classification problem. Thematic maps were produced with different number of samples and the accuracies of the thematic maps were compared. The results were verified for the effectiveness of different number of samples considering accuracy metrics (overall accuracy, F-score and Kappa coefficient). As a result, overall accuracies of 86.42, 91.84, 94.20 and 95.36% were produced for Deep Learning models using 50, 100, 200 and 400 samples, respectively.

1. INTRODUCTION

Deep learning, which is defined as a sub-branch of machine learning, has become a popular application in remote sensing in recent years, as these algorithms handle complex problems with higher accuracy, especially for image classification. Compared to traditional classification methods, deep learning models have achieved higher accuracies in classifying hyperspectral datasets (Paoletti et al. 2019). Since the architecture of deep learning models is flexible, this has a positive impact on the learning ability of these architectures. Another reason for the widespread use of the deep learning approach is the development of Graphics Processing Unit (GPU) hardware of computers (Paoletti et al. 2020). The concept of deep learning was originated from the artificial neural networks (ANNs). ANNs have been long employed for many problems with varying levels of success. Their black-box nature, optimal parameter selection, initial network size and pruning strategies have limited their use, particularly in remote sensing studies (Kavzoglu and Mather 1999). ANNs are data-dependent models focusing on training data characteristics, not the abstract values estimated from the samples as in the case of statistical methods (Cetin 2004). Therefore, training data must be representative or should be processed via refining to be representative (Kavzoglu 2009).The ANNs are formed by combining input, hidden, and output layers (Wang and Raj 2017). The connections between neurons is usually provided by the backpropagation method (Kavzoglu and Mather 2003). The term "deep" is used in Deep Learning because the increase in the number of hidden layers indicates that the model is getting deeper.

In image classification applications, spectral and spatial properties obtained from satellite images are jointly evaluated by deep learning methods. Thus, considering such properties of the data increases the accuracy of the thematic maps (Zhao and Du 2016). Image classification using deep learning methods may be divided into three parts: (i) dataset preparation, (ii) neural network model training, and (iii) classification using the trained model (Zhang et al. 2016). Moreover, deep learning algorithms learn the relationship between input data and labeled data using feature maps. In the literature, the neural network architecture, namely Deep Belief Network, Recurrent Neural Network, Autoencoder, and Convolutional Neural Network, have been commonly used for image processing such as segmentation, detection and classification (Du and Li 2018; Khan et al 2017; Merchant 2020; Sildir et al. 2020; Wang et al. 2020; Wu and Prasad 2017).

The purpose of this study is to analyze the performance of 2D CNN model using hyperspectral dataset by considering the different number of samples (i.e. 50, 100, 200 and 400). To meet this objective, the result of thematic maps generated by 2D CNN model was

* Corresponding Author

(eoyilmaz@gtu.edu.tr) ORCID ID 0000-0002-6853-2148
(kavzoglu@gtu.edu.tr) ORCID ID 0000-0002-9779-3443

Cite this study

Yilmaz E O & Kavzoglu T (2021). Analysis of the effect of training sample size on the performance of 2D CNN models. 2nd Intercontinental Geoinformation Days (IGD), 241-244, Mersin, Turkey

assessed based on overall accuracy, Kappa coefficient and F-score value.

2. CONVOLUTIONAL NEURAL NETWORKS

The Convolutional Neural Network (CNN) model has been widely used as the architecture of Deep Learning model (Vali et al. 2020). The main difference of the architecture of CNN models from other neural networks is that CNN includes convolution filters (Khan et al. 2018). The model of CNN has a great capacity to extract the features of the big data, such as hyperspectral imagery.

The architecture of the CNN model, a feed-forward neural network, consists of several combinations of dense, pooling, flattening, convolutional, and dropout layer (Yılmaz 2020). The convolutional layer is used for feature map extraction from a large dataset. The dense layer connects neurons in the previous and next layer. The pooling layer reduces the large data but keeps only the substantial information. The flattening layer converts multidimensional properties into a one-dimensional vector. Moreover, the dropout layer is used to prevent overfitting in the deep learning models (Srivastava et al. 2014). In addition to this architecture, other training parameters namely optimization, activation and learning rate also used in CNN models during the training stage (Yılmaz 2020).

3. STUDY AREA AND DATASET

In this study, a well-known Pavia University hyperspectral dataset was used to test the influence of using different numbers of training samples in the training stage. The dataset was obtained by a Reflective Optics Spectrographic Image System (ROSIS) sensor, which has 610x340 pixels with 1.3 meters spatial resolution (Fig. 1). The original dataset includes 115 spectral bands, but 12 bands were removed because they comprised noisy data.

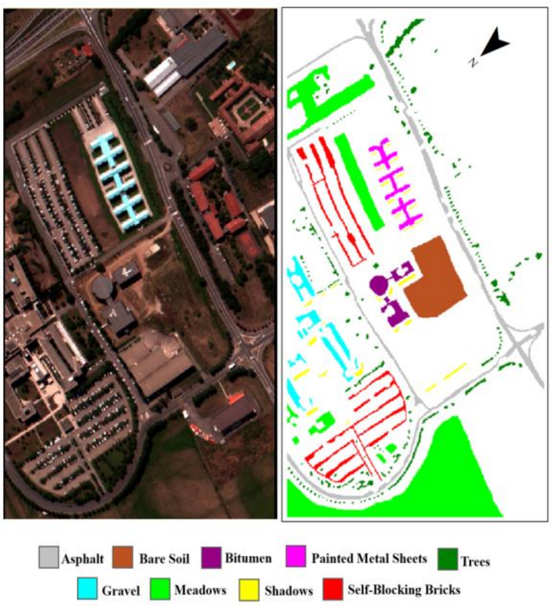

Figure 1. The study area and ground reference data for Pavia University hyperspectral imagery.

The reference dataset of Pavia image was generated for nine major land use/cover (LULC) classes and 42,776 labeled samples were available. The classes of LULC, namely asphalt, meadows, gravel, trees, painted metal sheets, bare soil, bitumen, self-blocking bricks, and, shadows, are given Table 1.

Table 1. The number of samples from reference data.

LULC Classes	Number of Samples
Asphalt	6,631
Meadows	18,649
Gravel	2,099
Trees	3,064
Painted metal sheets	1,345
Bare soil	5,029
Bitumen	1,330
Self-blocking bricks	3,682
Shadows	947
Total	42,776

Before classifying the Pavia University hyperspectral dataset using 2D CNN model, the first step was to randomly divide the dataset into three subsets: the first (training dataset) with 75% of the samples, the second (validation dataset) with 15% of the samples to evaluate the overall accuracy of the thematic maps and the third (test dataset) with %10. Thus, 50, 100, 200, and 400 samples for each reference class were randomly selected from the training dataset. It should be noted that the application of classification was performed using Jupyter Notebook with Python Language.

4. RESULTS

In this study, LULC classification was performed using the 2D CNN model. In order to build the user-defined model, input, batch, convolutional, flattening, dropout and dense layers were generated (Fig. 2.). It should be noted that a patch size of 5x5 was chosen for all the applications of 2D CNN models. Moreover, the training dataset was set according to the determined patch size. In order to train the deep learning model quickly, the training dataset was normalized before the model training phase. The unit parameters of the dense layer were chosen as 64 and 128 considering a trial-and-error approach. Moreover, the activation functions of ReLU and Softmax were employed in the processing stage.

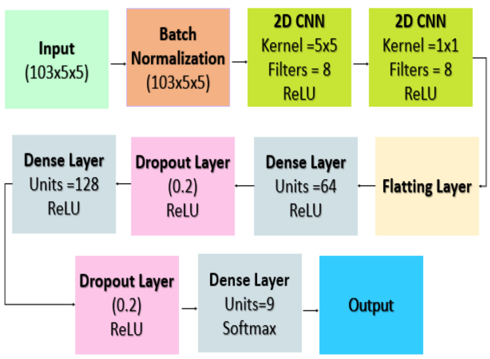

Figure 2. 2D CNN model adopted in this study.

The produced thematic maps representing 9 LULC classes were generated with 2D CNN model using 50, 100, 200 and 400 samples (Fig. 3.).

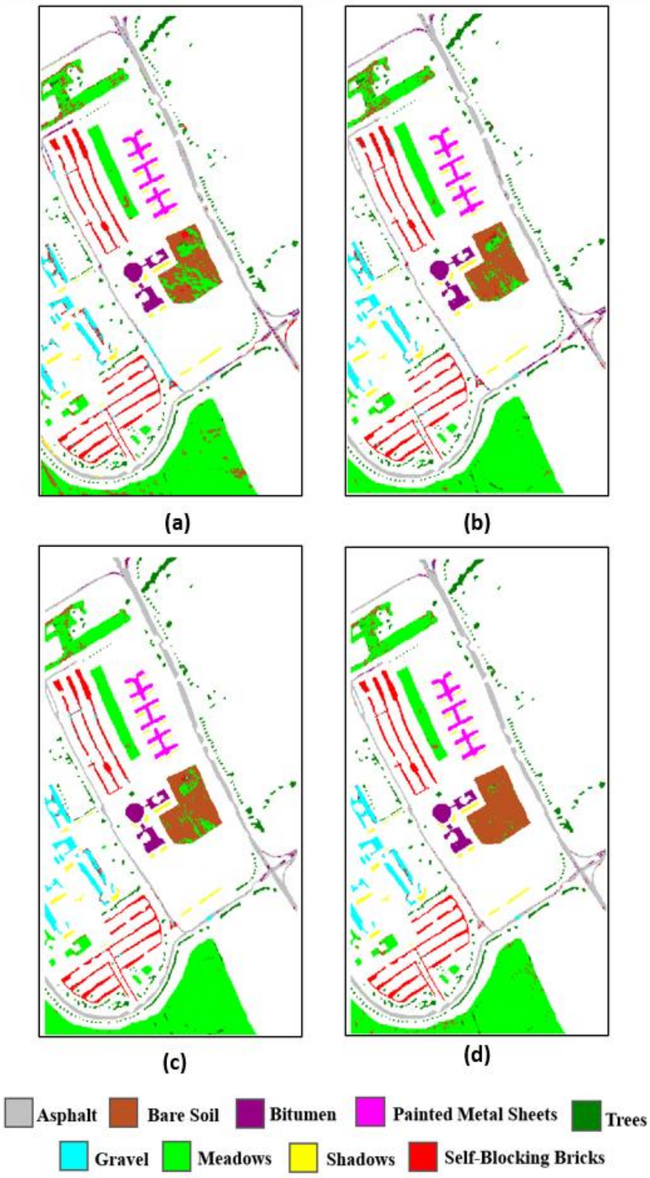

Figure 3. Thematic maps produced with 2D CNN model by using (a) 50 samples, (b) 100 samples, (c) 200 samples, (d) 400 samples.

To analyze the accuracy assessment of the thematic maps, the overall accuracies and Kappa coefficients were calculated by using the reference dataset. The overall accuracies of the thematic maps generated using 50, 100, 200 and 400 samples were estimated to be 86.42, 91.84, 94.20 and 95.36 respectively. Also, Kappa coefficients were calculated as 0.82, 0.89, 0.92 and 0.94, respectively (Table 2). Furthermore, F-score values were produced to evaluate the estimated accuracy of each LULC class. As can be seen in the table, the LULC classes with the highest F-score value (1.00) was calculated for the painted metal sheets for all number of samples combination and shadow classes, except for the case of 50 samples. The LULC classes with the lowest F-Score values belong to the bare soil (0.68), gravel (0.77) and bitumen (0.79) classes for 50 samples. From the visual analysis of the thematic maps produced, it was observed that the gravel class was mixed with the self-blocking brick class for 50 samples. It was also observed that the meadow and bare soil classes were mixed together due to similar spectral characteristics. The thematic map created with 100 samples had the lowest F-score values (0.82) with the bitumen and bare soil classes. Moreover, it was observed that the bare soil class was mixed with meadows class in the thematic map produced with both 200 and 400 samples.

Training times for different sizes of training samples were recorded and shown in Table 2. In the models using 50, 100 and 200 samples, the training processes lasted 500.5, 500.4 and 500.3 seconds, respectively. However, the slowest training period of the model took about 930 seconds for 400 samples. The reason could be related to the large sample size of the dataset employed during the training.

Table 2. Accuracy assessment for the thematic maps produced with 2D CNN model.

LULC Classes	Number of Samples			
	50	100	200	400
Asphalt	0.87	0.93	0.97	0.96
Meadows	0.91	0.94	0.95	0.96
Gravel	0.77	0.87	0.91	0.95
Trees	0.93	0.95	0.95	0.97
Painted metal sheets	1.00	1.00	1.00	1.00
Bare soil	0.68	0.82	0.85	0.89
Bitumen	0.79	0.82	0.91	0.91
Self-blocking bricks	0.87	0.91	0.94	0.97
Shadows	0.93	1.00	1.00	1.00
Overall Acc. (%)	86.42	91.84	94.20	95.36
Kappa Coef.	0.82	0.89	0.92	0.94
Training Time (sec)	500.5	500.4	500.3	937.2

5. CONCLUSION

The deep learning methods have been widely used in remote sensing applications, including object detection, image segmentation and image classification. Superior performances have been reported in the literature for the classification of hyperspectral images with deep learning algorithms. In this study, Pavia University hyperspectral image dataset was classified using 2D CNN models with 50, 100, 200 and 400 samples. In addition, thematic maps were generated by the deep learning model using a 5x5 patch size, then accuracy measurements were conducted. The 2D CNN model with 400 samples resulted in a significant increase in overall accuracies (~9%), particularly compared to 50 samples per class. The findings in this study revealed that samples per class employed in the training stage of 2D CNNs can have significant impact on the achieved accuracy. Poor performances were observed for the cases where limited training data were available for training.

REFERENCE

Du J & Li Z (2018). A hyperspectral target detection framework with subtraction pixel pair features. IEEE Access, 6, 45562–45577.

https://doi.org/10.1109/ACCESS.2018.2865963

Kavzoglu T & Mather P M (1999). Pruning artificial neural networks: An example using land cover classification of multi-sensor images. International Journal of Remote Sensing, 20(14), 2761–2785. https://doi.org/10.1080/014311699211796

Kavzoglu T & Mather P M (2003). The use of backpropagating Artificial Neural Networks in land cover classification. International Journal of Remote Sensing, 24(23), 4907–4938. https://doi.org/10.1080/0143116031000114851

Cetin M, Kavzoglu T & Musaoglu N (2004) Classification of multi-spectral, multi-temporal and multi-sensor images using Principal Components Analysis and Artificial Neural Networks: Beykoz case. Proceedings of 20th Congress of the International Society of Photogrammetry and Remote Sensing, Istanbul, 12-23 July 2004.

Kavzoglu T (2009). Increasing the accuracy of Neural Network classification using refined training data. Environmental Modelling and Software, 24(7), 850–858. https://doi.org/10.1016/j.envsoft.2008.11.012

Khan M J, Yousaf A, Javed N, Nadeem S & Khurshid K (2017). Automatic target detection in satellite images using Deep Learning. Journal of Space Technology, 7(1), 44–49.

Khan S, Rahmani H, Shah S A A & Bennamoun M (2018). A guide to Convolutional Neural Networks for computer vision. Synthesis Lectures on Computer Vision, 8(1), 1–207. https://doi.org/10.2200/s00822ed1v01y201712cov015

Merchant M A (2020). Classifying open water features using optical satellite imagery and an object-oriented convolutional neural network. Remote Sensing Letters, 11(12), 1127–1136. https://doi.org/10.1080/2150704X.2020.1825869

Mosavi A, Ardabili S, & Várkonyi-Kóczy A R (2020). List of Deep Learning Models. In Lecture Notes in Networks and Systems, 101, 202–214. https://doi.org/10.1007/978-3-030-36841-8_20

Paoletti M E, Haut J M, Plaza J, & Plaza A. (2019). Deep learning classifiers for hyperspectral imaging: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 158, 279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006

Paoletti M E, Haut J M, Tao X, Miguel J P & Plaza A. (2020). A new GPU implementation of support vector machines for fast hyperspectral image classification. Remote Sensing, 12(8), 1257. https://doi.org/10.3390/RS12081257

Sildir H, Aydin E, & Kavzoglu T (2020). Design of Feedforward Neural Networks in the classification of hyperspectral imagery using superstructural optimization. Remote Sensing, 12(6), 956. https://doi.org/10.3390/rs12060956

Srivastava N, Hinton G E, Krizhevsky A, Sutskever I & Salakhutdinov R (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 399–421. https://doi.org/10.1016/B978-1-4832-1446-7.50035-2

Vali A, Comai S & Matteucci M (2020). Deep Learning for Land use and land cover classification based on hyperspectral and multispectral Earth observation data: A review. Remote Sensing, 12(15), 2495. https://doi.org/10.3390/rs12152495

Wang H & Raj B (2017). On the Origin of Deep Learning. Computer Science - Machine learning, 1–72. Retrieved from http://arxiv.org/abs/1702.07800

Wang J, Zheng Y, Wang M, Shen Q & Huang J (2020). Object-scale adaptive Convolutional Neural Networks for high-spatial resolution remote sensing image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 283–299. https://doi.org/10.1109/JSTARS.2020.3041859

Wu H & Prasad S (2017). Convolutional Recurrent Neural Networks for hyperspectral data classification. Remote Sensing, 9(3), 298. https://doi.org/10.3390/rs9030298

Yılmaz E Ö (2020). Uzaktan algılanmış verilerin derin öğrenme yöntemiyle sınıflandırılması. MS Thesis, Gebze Technical University, Kocaeli (in Turkish).

Zhang L, Zhang L & Du B (2016). Deep Learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine, 4(2), 22–40. https://doi.org/10.1109/MGRS.2016.2540798

Zhao W & Du S (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 113 (March 2016), 155–165. https://doi.org/10.1016/j.isprsjprs.2016.01.004