Seasonal Multi-temporal Pixel Based Crop Types and Land Cover Classification for Satellite Images Using Convolutional Neural Networks

Noureldin Laban and Bassam Abdellatif
Data Reception and Analysis Division
National Authority for Remote Sensing
and Space Sciences
Cairo, Egypt
Email: Nourlaban, bassam.abdellatif@narss.sci.eg

Hala M.Ebeid, Howida A.Shedeed, and Mohamed F.Tolba
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
Email: halam, dr_howida, fahmytolba@cis.asu.edu.eg

***Abstract*—Nowadays, Satellite images have become a major source of data for many aspects of development. Land and crops classification using satellite images is a recent important subject. From the other side, Deep Convolutional Neural Networks (DCNNs) is a powerful technique for understanding images. This paper describes a pixel based crops and land cover classification originating from one source satellite imagery represented by Sentinel satellite and based on several dates for the same agricultural season. We propose a DCNN architecture based on multi-temporal data that was fed to a one-dimension (1-D DCNN). The proposed architecture is compared with other methods of satellite image classification algorithms; such as Random Forest (RF), Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN). Experiments are conducted for the mutual experiment of major crops and land cover classification for Al-Fayoum governorate in Egypt. The 1-D DCNN achieves about 89% accuracy using 10 spectral bands of Sentinel-2 satellite imagery for the area of interest. The proposed architecture although it outperforms other methods, needs further research to optimize the memory usage.**

***Index Terms*—Artificial Intelligence, Crop Classification, Convolutional Neural Networks,Remote Sensing (RS), Egypt,Sentinel-2 , Satellite Images, TensorFlow.**

I. Introduction

Over the last decade, there has been a revolution in remote sensing data gathering technologies. There are many different types of satellites that have been launched over the last period of time [1]. This generates a enormous challenge to classify this massive amount of data with different characteristics [2]. A crop classification is one of the most important and vital process in land cover classification, as reliable crop classification is important for the exploration of agricultural land utilization in development and environmental projects [3]–[5].

There are many classification algorithms that have been applied to crop classification [6] [7]. The most popular and effective methods for land cover classification are; Deep learning and ensemble based methods [1], [8]. Support Vector Machines (SVMs) are suitable for applications which used for satellite image classification , clearly visible, for the fact that they require small training datasets, on which they can handle a robust generalization [3], [9]. In recent years, Random Forests (RFs) have been used effectively in classification of satellite images [1], [6], [10], [11]. Convolutional Neural Networks (CNNs) have achieved record breaking results in machine learning particularly image classification [12]. Its principal scheme is to simulate the human vision process to solve problems of big data. This is done by using all the available data and providing the semantic information as a final product [1].

Using satellite or airborne imagery for land cover mapping and crops classification has exponentially increased over the past decade, due to the improved data availability and accessibility besides increasing the computational power [6]. Crops and land cover classification are widely used by government departments and private sectors worldwide specially with satellite images [4].

There are two approaches to satellite image classification: Pixel based and object based. Pixel based crop classification is based on the spectral values for each pixel. Object based crop classification is based on logical objects that are the output of a segmentation algorithm. Convolutional Neural Networks (CNNs) have different schemes of implementation. Through the last few years there have been many improvements to CNNs architectures. Batch Normalization allows having much higher learning rates and less sense about initialization [13]. Overlapping convolution and pooling allows you to choose the slides and the padding which results in the same output size as input [12].

The pixel-based classification allows adjusting the classifier model in the field easily by adding a few points to training data [14] during data selection process. It is generally used for low level classification processes as land cover and land mapping, land slides mapping and change detection [15]. Object-based algorithms try to identify the real objects with specific shapes [14]. Pixel-based is more adequate for land cover and crop classification. It is used widely for many applications [6].

Through the last few years, there has been a big increase in different types of earth scanning satellites, along with the new revolution of deep learning techniques. This leads to open benchmark datasets for satellite image annotation and a huge library of open source deep learning libraries. This progress positively affects the crop type classification and land cover researches.

[1] presented deep learning architecture that targets crop type and land cover classification from multi-temporal and multi-source satellite imagery for a test site in Ukraine based on a multi-level process. [5] introduced a crop sequence based ensemble classification method,with exploring the expert knowledge and TerraSAR-X multi-temporal image based phenological information. [2]overcome the absence of large labeled satellite images datasets by engaging two techniques with CNN: Transfer learning with fine-tuning and data augmentation adapted to satellite images.

[10] combined two different techniques for image recognition: for feature extraction using a pre-trained CNN and for feature classification using the multinomial logistic regression model. [3] have developed a voting system in ensemble learning, which is replaced by a debate base conflict resolution. [16] applied an Automated Cropland Mapping Algorithm (ACMA) using Google Earth Engine cloud computing which has extensive knowledge of the croplands of Africa . [17] classify high resolution image scene by fusing global saliency based, multi-scale, multi-resolution, multi-structure local binary pattern features and the local code bookless model feature.

[18] proposed an end-to-end framework for a pixel-based classification of satellite imagery with Convolutional Neural Networks (CNNs). [19] uses a convolutional neural network model as a deep feature extraction method for hyperspectral image classification by utilizing several convolutional and pooling layers to extract deep features because CNNs are discriminant, non-linear and invariant. [4] uses high spatial and high spectral resolution remote sensing data for an object-based crop mapping framework for miscellaneous areas.

[20] proposed a spectral-spatial classification of the hyperspectral image using a dual-channel convolutional neural network framework. They extract the hierarchical spectral features using one-dimensional CNN. [21] presented another deep learning framework for spectralspatial classification of hyperspectral images. They merged the spatial and spectral features using deep learning architecture including stacked auto- encoders and deep convolutional neural networks. [22] used a deep learning framework for hyperspectral image classification using both spectral and spatial features. They used Logistic Regression (LR), Deep Convolutional Neural Networks (DCNNs) and a hybrid of principal component analysis.

In this paper, we propose a seasonal multi-temporal pixel based crop classification for Satellite images using one-dimension Convolutional Neural Networks(1-D CNN). We use data from sentinel-2 satellite for several dates covering Fayoum region in Egypt. The main problem we faced was the little amount of data for each pixel which is represented in 10 bands of the sentinel-2. We built our one-dimensional CNN model based on zero padding convolution, pooling and batch normalization. We trained our model using 8 different classes of crops and land cover include: ”sugar beet”, ”water”, ”trees”, ”urban”, ”bare land”, ”wheat”, ”clover” and ”background”. We compared our results to other methods.

The remainder of this paper is organized as follows: In Section 2, we propose our methodology, including the overview of architecture and CNN model. The region of study, training data and experimental results compared to other classification methods are presented in Section 3. Finally, Section 4 concludes this paper.

II. PROPOSED ARCHITECTURE

In this section, we describe in detail the fundamental operations of 1D-CNN based classification and involve how to construct the seasonal multi-temporal data to train this network as shown in Figure 1.

A. *Seasonal Multi-Temporal Data Preparation*

Deep convolutional neural networks require a very large training data set to learn its deep structure [20]. However, for satellite images, getting this data set is very expensive, so we have only a limited labeled samples, which may lead to over-fitting. On the other hand, each crop has different stages of growing. So for the same location , there are different spectral signature for the same crop. We built training data set using a sequence of satellite images for the same season as shown in Figure 1. This has two advantages. First; catch different signature of crops in the learning model and second, increase the number of training data set.

B. *Convolutional Neural Networks*

A Deep Convolutional Neural Networks (DCNNs) can achieve record breaking results on a highly challenging dataset by purely using supervised learning [12]. There are many deep learning neural network architectures that have been developed in the last few years with many concepts having been introduced to enhance the performance of DCNNs. Dealing with pixel based crop classification has two main problems: First, the data set has small dimensional feature vectors which represent the spectral reflectance for each pixel. This small length affects the convolution and pooling processes on the CNN architecture. Second problem, we need to achieve an efficient training process with this small vector length. These problems have been solved through the following subsections.

1) Single Dimension Convolution/Pooling with Padding: The Convolution layer is the core building block of a Convolutional Network [23]. To overcome the problem of small vector length, we use the convolution with padding. It is convenient to pad the input volume with zeros around the border. This padding allows to manage the spatial size of the output length and doesn’t affect the spatial size of the input length so the input and output length are the same.

2) Batch Normalization Training: Batch normalization increases learning rates and in the same time, be less careful about initialization. It decreases the co-variate shift, which complicates the training of machine learning systems. Removal of co-variate shift from internal activations of the

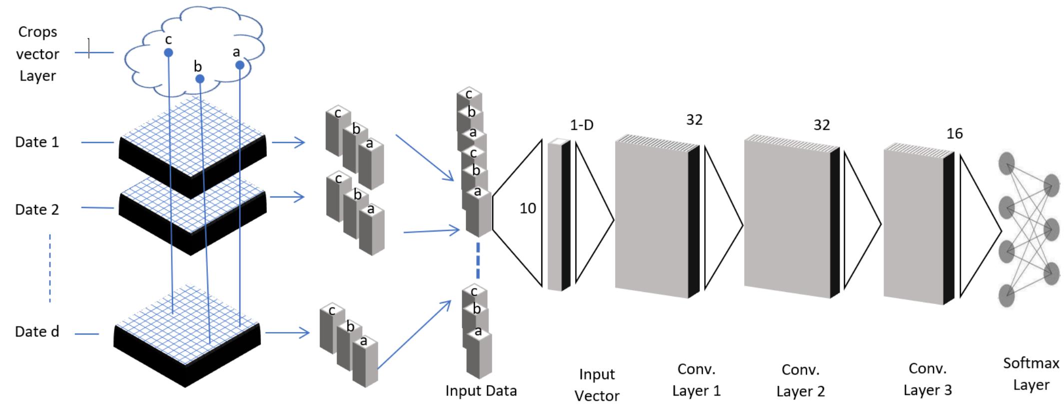


Fig. 1: Proposed Architecture for Seasonal Multi-Temporal Satellite Images Classification.

network may aid in training [13]. It also applies to sub-networks and layers. It regularizes gradients from distraction to outliers and flow towards the common goal within a range of the mini-batch resulting in acceleration of the learning process. Batch normalization is utilized to normalize the features result in the convolutional layer so that the feature weight of each band will be near each other. So, a larger learning rate can be used to accelerate the training operation [24].

3) Validation Dataset: Validation is considered as an assessment of the model predictive performance on an independent validation dataset [25], [26]. We utilize this validation dataset through the training process to measure the over-fitting of our training model. In addition, to demonstrate the outcome of different validation percentages, the training and testing processes were conducted many times, with different training and validation percentages each time. Finally, we have divided our ground truth data into three sub-sets validation set, training set and testing set.

C. Prediction of the whole image classification and vectorized results

After we get the best CNN model, we use it to classify the whole image pixels as shown in Figure 2. We reshape these pixels into two-dimension Matrix of the original image. Then we vectorize the resulted classified image to form the vector version of classification. Using the vector version can easily compute the total area for each crop and land cover. This allows interpreting the classification results and supplying the proper statistics for decision-makers.

III. EXPERIMENTS AND RESULTS

A. Data and Setup

We treat with the problem of crop types and land cover classification for the Fayoum region of Egypt. We use satellite images obtained by the Sentinel-2 satellite with a resampled spatial resolution a 10-meter starting from January to March 2016 (specifically 10/1,20/1,9/2,10/3/2016) as in Figure 3. We select 10 bands from Sentinel-2 satellite image bands to form each pixel. we apply a radiometric and geometric correction to the satellite images. Also, we re-sample all bands of the satellite images to 10-meter spatial resolution.

The study region has been classified into seven classes including; main agricultural crops ('sugar beet', 'bare land', 'water', 'urban', 'wheat', 'trees' and 'clover'). The total area of the region is about 2000 square km with a diversity of different land cover types and agricultural crops.

We have gathered ground truth datasets from January to March 2016 and have split the gathered data into testing and training datasets as in table I. Each training sample is represented as a pixel on the satellite image and is formed of 10-valued reflectance array.

Experiments are carried out to analyze the effect of the proposed model on classification accuracy. All the experiments are performed on the same workstation with 72 core Intel Xeon Phi Processor 7290 ,each core has 2.50 GHz and workstation has 256 GB RAM. We implemented our experiments using Deep 1-D Convolutional Neural Networks (Deep 1-D CNN) with two main characteristics: Zero Padding Convolution and Batch Normalization in Python using Google's TensorFlow deep learning library.

B. Fine-tuning of the CNN model

We fine tune our CNN model using different hyper-parameters . It has seven layers with three convolution layers, three max-pooling layers and one for fully connected layered. The total number of iteration we used is 60000 iterations with 2×10^{-5} as a learning rate. We also used 32 as batch size. Also, we use the Stochastic Gradient Descent (SGD) for computing gradients with ADAM Optimizer.

C. Results

Overall classification accuracies for RF, kNN, SVM and Deep 1-D DCNN were 0.83, 0.85, 0.86 and 0.89 respectively, as shown in Table. I. Table II shows the confusion matrices for all used methods.

The results obtained by the proposed Deep 1-D CNN method against other methods shows a significant improvement for both recall and precision for the region of interest.

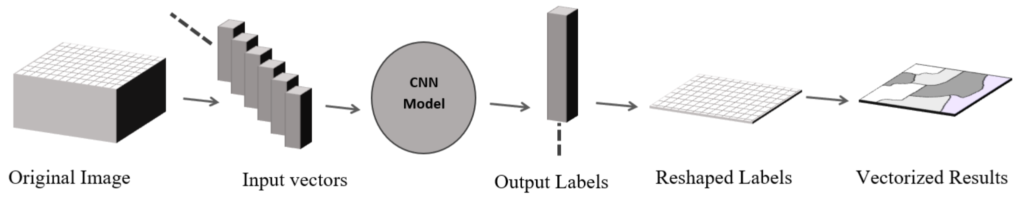


Fig. 2: Prediction of the Real Image Classes

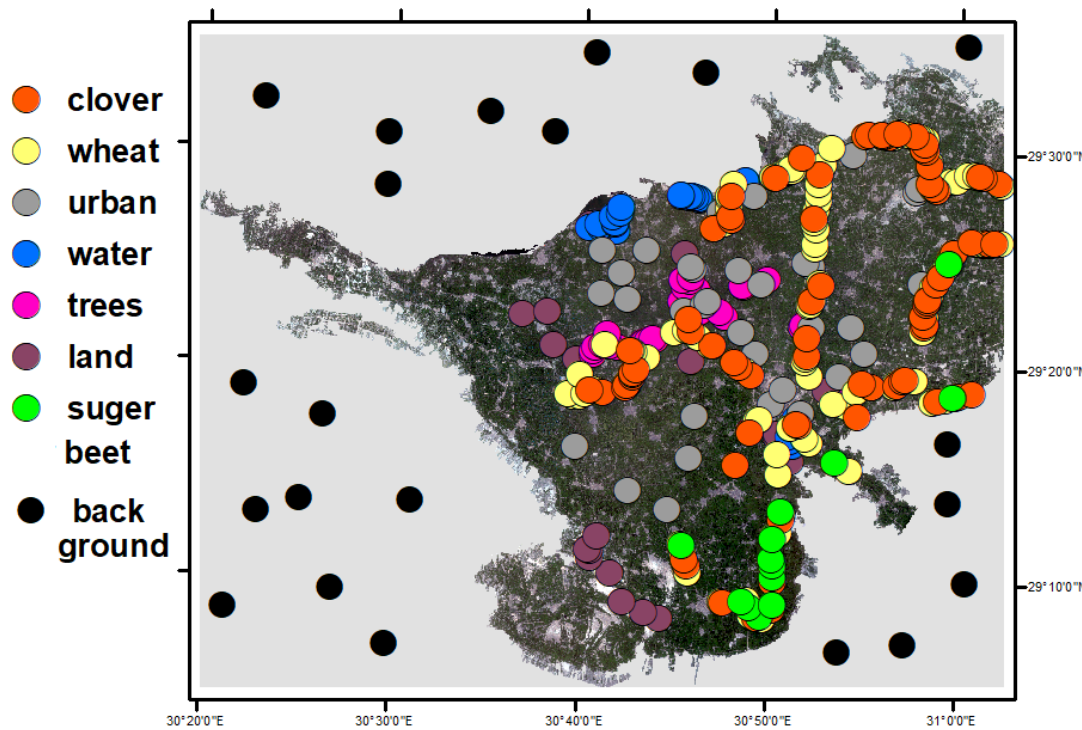


Fig. 3: Fayoum region with a number of samples for training and testing for each class.

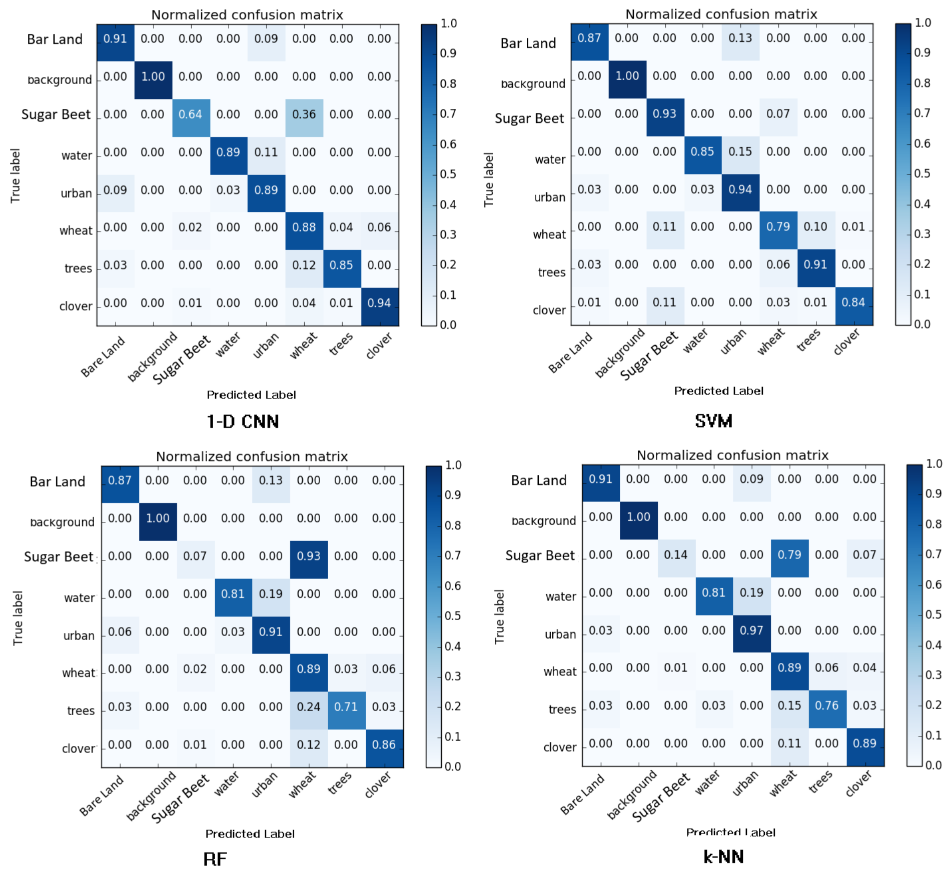
TABLE I: Overall, average, individual class accuracies

Classes	Number of samples			Classification Methods			
	Single input Training	All seasonal Trainning	Testing	k -NN	RF	SVM	1-D CNN
Bare Land	22	88	23	0.91	0.89	0.87	0.87
sugar beet	13	52	14	0.24	0.11	0.53	0.54
Water	27	108	27	0.88	0.9	0.9	0.94
Urban	35	140	35	0.85	0.88	0.88	0.89
Wheat	122	488	122	0.85	0.84	0.86	0.9
Trees	34	136	34	0.78	0.78	0.79	0.81
Clover	80	320	80	0.9	0.86	0.91	0.94
Background	21	84	19	1	1	1	1
Total/ Average Accuracy	354	1416	354	0.85	0.83	0.86	0.89

We use the total F-measure as accuracy metric as it represents the harmonic mean between recall and precision values. Water and background were the most clear and notable classes for almost all classification methods. Crops as Sugar beet, Clover and Wheat classes were the most difficult classes to be distinguished by almost all methods, even though the proposed CNN method has a significant power to distinguish these crops. The most confusion was found between Clover and Wheat crops. Sugar beet has the least recall accuracy.

Some examples of classification results obtained by different methods as shown in Figure 4.

TABLE II: Confusion Matrix for SVM, kNN, RF, and Deep 1-D CNN methods.



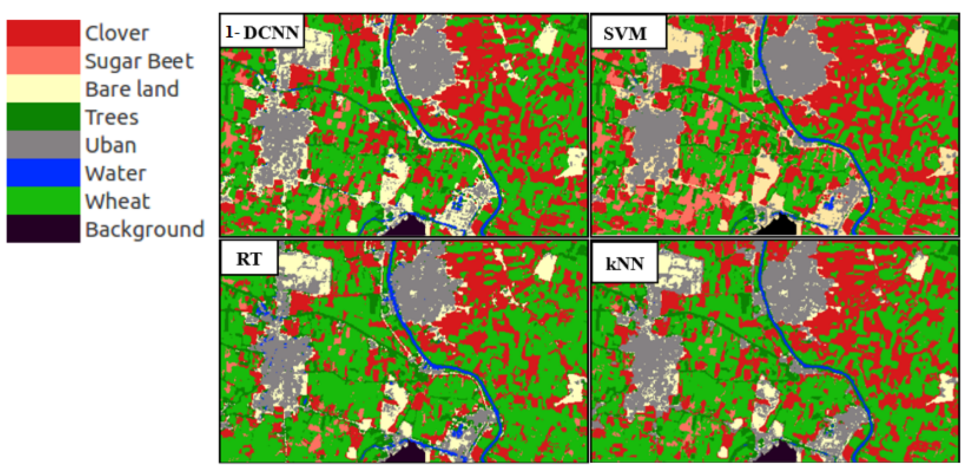


Fig. 4: Examples of the different outputs of different classification methods used.

IV. Conclusion

In this paper, we proposed a seasonal multi-temporal pixel based land cover and crop classification using Deep Convolutional Neural Network (DCNN) architecture. The architecture uses 10 spectral bands of Sentinel-2 satellite imagery over the Fayoum region in Egypt during winter season of 2016. The proposed architecture of Deep 1-DCNN outperforms other methods as SVM, kNN and RF. Also the proposed architecture has performed on average about 89% for major crops (wheat, clover and sugar beets) and land cover classes (Urban, water and bare land). The pixel based approach shows more define and appropriate classification although it needs to optimize the memory usage. For future work, we intend to examine the CNN approach using other satellite images datasets which include very high spatial resolution and hyperspectral information and covering a higher spatial region. In addition to the optimization of the pixel based classification process, in terms of memory management and processing power.

Acknowledgments

This work was supported in part by the GEF/World Bank Project " Regional Co-ordination for Improved Water Resources Management and Capacity Building " alongside The National Authority for Remote, Sensing and Space Science, Egypt.

References

[1] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, May 2017.

[2] 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 Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, April 2017.

[3] S. Contiu and A. Groza, "Improving remote sensing crop classification by argumentation-based conflict resolution in ensemble learning," *Expert Systems with Applications*, vol. 64, no. Supplement C, pp. 269 – 286, 2016.

[4] M. Wu, W. Huang, Z. Niu, Y. Wang, C. Wang, W. Li, P. Hao, and B. Yu, "Fine crop mapping by combining high spectral and high spatial resolution remote sensing data in complex heterogeneous areas," *Computers and Electronics in Agriculture*, vol. 139, no. Supplement C, pp. 1 – 9, 2017.

[5] B. K. Kenduiywo, D. Bargiel, and U. Soergel, "Higher order dynamic conditional random fields ensemble for crop type classification in radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4638–4654, Aug 2017.

[6] R. Khatami, G. Mountrakis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sensing of Environment*, vol. 177, pp. 89 – 100, 2016.

[7] N. Laban, B. Abdellatif, H. M. Ebeid, H. A. Shedeed, and M. F. Tolba, "Improving land-cover and crop-types classification of sentinel-2 satellite images," in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*, A. E. Hassanien, M. F. Tolba, M. Elhoseny, and M. Mostafa, Eds. Cham: Springer International Publishing, 2018, pp. 449–458.

[8] E. Ferreira, A. d. A. Arajo, and J. A. d. Santos, "A boosting-based approach for remote sensing multimodal image classification," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2016, pp. 416–423.

[9] X. Huang and L. Zhang, "An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 257–272, Jan 2013.

[10] G. Xu, X. Zhu, D. Fu, J. Dong, and X. Xiao, "Automatic land cover classification of geotagged field photos by deep learning," *Environmental Modelling & Software*, vol. 91, pp. 127 – 134, 2017.

[11] I. H. Ikasari, V. Ayumi, M. I. Fanany, and S. Mulyono, "Multiple regularizations deep learning for paddy growth stages classification from landsat-8," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct 2016, pp. 512–517.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448–456.

[14] W. Strothmann, A. Ruckelshausen, J. Hertzberg, C. Scholz, and F. Langsenkamp, "Plant classification with in-field-labeling for crop/weed discrimination using spectral features and 3d surface features from a multi-wavelength laser line profile system," *Computers and Electronics in Agriculture*, vol. 134, no. Supplement C, pp. 79 – 93, 2017.

[15] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11 – 28, 2016.

[16] J. Xiong, P. S. Thenkabail, M. K. Gumma, P. Teluguntla, J. Poehnelt, R. G. Congalton, K. Yadav, and D. Thau, "Automated cropland mapping of continental africa using google earth engine cloud computing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 126, no. Supplement C, pp. 225 – 244, 2017.

[17] X. Bian, C. Chen, L. Tian, and Q. Du, "Fusing local and global features for high-resolution scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2889–2901, June 2017.

[18] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, Feb 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, Oct 2016.

[20] H. Zhang, Y. Li, Y. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network," *Remote Sensing Letters*, vol. 8, no. 5, pp. 438–447, 2017.

[21] J. Yue, S. Mao, and M. Li, "A deep learning framework for hyperspectral image classification using spatial pyramid pooling," *Remote Sensing Letters*, vol. 7, no. 9, pp. 875–884, 2016.

[22] J. Yue, W. Zhao, S. Mao, and H. Liu, "Spectral-spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sensing Letters*, vol. 6, no. 6, pp. 468–477, 2015.

[23] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1537–1541, Oct 2016.

[24] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4520–4533, Aug 2017.

[25] G. B. Humphrey, H. R. Maier, W. Wu, N. J. Mount, G. C. Dandy, R. J. Abrahart, and C. W. Dawson, "Improved validation framework and r-package for artificial neural network models," *Environmental Modelling & Software*, vol. 92, no. Supplement C, pp. 82 – 106, 2017.

[26] C. Grajeda, F. Breitinger, and I. Baggili, "Availability of datasets for digital forensics and what is missing," *Digital Investigation*, vol. 22, no. Supplement, pp. S94 – S105, 2017.