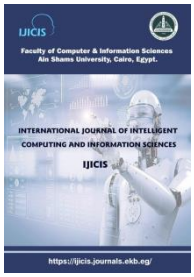
IJICIS, Vol.21, No.3, 13-28
DOI: 10.21608/ijicis.2021.79070.1099

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



ENHANCED PIXEL BASED URBAN AREA CLASSIFICATION OF SATELLITE IMAGES USING CONVOLUTIONAL NEURAL NETWORK

Noureldin Laban*

Data Reception and Analysis Division, National Authority for Remote Sensing and Space Science, Cairo, Egypt
nourlaban@narss.sci.eg

Bassam Abdellatif

Data Reception and Analysis Division, National Authority for Remote Sensing and Space Science, Cairo, Egypt
bassam.abdellatif@narss.sci.eg

Hala M. Ebied

Scientific Computing department Computer and Information Science, Ain Shams University, Cairo, Egypt
halam@cis.asu.edu.eg

Howida A. Shedeed

Scientific Computing department Computer and Information Science, Ain Shams University, Cairo, Egypt
dr_howida@cis.asu.edu.eg

Mohamed F. Tolba

Scientific Computing department Computer and Information Science, Ain Shams University, Cairo, Egypt
fahmytolba@cis.asu.edu.eg

Received 2021- 6-4; Revised 2021-8-8; Accepted 2021-8-11

Abstract: *Recent years have witnessed a great development in the use of deep learning in the applied fields in general, including the improvement of remote sensing. Satellite imagery classification has played a prominent role in various development processes. This paper presents a new improvement in automatic urban classification using One Dimension Convolutional Neural Network (1DCNN) architecture. The suggested approach has three enhancement processes. First, select training boxes for different classes and create many pixels with variable class signatures. This makes the training process dependent on the broadband of signature for the classes. Second, modified 1D convolution was used to re-encode pixel values to increase distinguish power. Third, adding a new median filter layer at the end of network architecture to remove pixels like noise to make the resulting map smoother. An image of Greater Cairo is used and the different urban classes are defined within it. The proposed method was compared to other methods based on pixels. The proposed method proved to be numerically and visually superior.*

Keywords: *Satellite images, image classification, semantic segmentation, deep learning, 1DCNN, data augmentation, neural net training, Urban area classification*

* Corresponding author: Noureldin Laban
Data Reception and Analysis Division, National Authority for Remote Sensing and Space Science, Cairo, Egypt
E-mail address: nourlaban@narss.sci.eg

1. Introduction

The use of satellite imagery has become essential in our present life. Several Earth observations satellites have been launched which send exceptionally enormous amounts of data daily [1]. These data have contributed to sustainable development processes, whether agricultural, geological, urban, or other aspects of development[2]. It became necessary to deal with this data automatically, whether it is processing or interpretation. This led to a great development in machine learning technologies that are used in processing and understanding remote sensing data [3].

The new thinking has contributed to dealing with data processing units and the production of the matrix processing processor, or what is called the Graphic Processing Unit (GPU), which increased the computing power when working with convolutional calculations on exceptionally large matrices [4]. Before that, the development of building artificial neural networks and converting them into a Convolutional Neural Network (CNN) capable of dealing directly with image matrices. At the same time, the open-source community is being organized to build customized software libraries for deep learning networks and the expansion of participation in them on the one hand. On the other hand, datasets are more challenging and powerful in education and training [5]. All this was done through open international competitions, which led to a global movement in this field and is still ongoing [6].

CNN's basic idea is based on a series of convolutions and pooling with activation functions in the presence of a set of Convolutions filters containing many learning weights. Weights are adjusted during the training process through backpropagation using a loss function. This process requires many computations, especially for adding and multiplying matrices, which is what GPUs provide. The most used CNN models are AlexNet [7], VGGNet [8], InceptionNet [9], and ResNet [10]. All models are a basic part of most deep learning frame works [11]. There are also many improvements that have been added for CNN frameworks to improve the learning process. These improvements include batch normalization[12], data augmentation [13], dropouts [14]. All these improvements make CNNs a state-of-art technique, skipping all previous methods.

The Convolutional Neural Network was originally created to handle 2D images using 2D convolution filters [15]. When the filter size is m x m where m is usually 3, 5, or 7. When m is a value of 1, this means re-encoding the input image. It produces images of the same size but with a different encoding. This technique is usually used for less difficult problems that require faster execution. On the other hand, they can be relied upon when dealing with pixel-based semantic segmentation [16]. The structure of 1DCNN is simple. It is consisting of the few numbers of convolution layers followed by one or two fully connected layers as MultiLayers Perceptrons (MLP) [17]. This simple structure is suitable for simple pixels-based classification.

Classification of satellite images using different machine learning technologies has become a vital topic over the past years, especially with the huge boom in deep learning techniques [3], [18]. Satellite images differ from traditional images in several things, the most important of which is high dimensionality so that there are many bands other than red, green, and blue. Also, Satellite images have richer content than traditional images. On the other hand, few training samples for supervised learning classification are available for satellite images [19]. There were many ways to solve these challenges using shallow or deep learning [20].

There are many traditional methods used in space image manipulation software such as k-Nearest Neighbors (k-NNs), Support Vector Machines (SVMs), and Random Forests (RFs) [21]. These

methods are primarily used for their ease of use with good accuracy in dealing with pixel-level in images classification. Random Forests (RFs) develop a multiple of decision trees then combine the results through voting between these trees [22]. Support Vector Machines (SVMs) aim to get the best separator hyperplane between two classes using its kernel function. The main problem with SVM in large training data is failing to find an ideal solution in a reasonable time.

Satellite image classification methods are divided into two types. The first type is pixel-based methods, in which the pixels are classified pixel by pixel for the entire image and this type is trained using a pixel dataset. The second type is object-based methods, where the pre-selected objects are created to perform the classification process for them, and their training relies on the pre-classified images. The first type is the most commonly used, especially with remote sensing packages, as it is easier and cheaper to collect and label training data. Although, the traditional methods that depend on pixels to train have become weak in obtaining the variety of data in the presence of data scarcity. As for image-based classification, it requires massive amounts of data to figure out what needs more effort and cost. So the question is, is the improvement in efficiency equivalent to the cost and additional effort required? So the question was how do we use a deep learning approach and use a pixel-based approach where we can take advantage of deep learning and at the same time get the benefits of a pixel-based approach with ease and cost. That is what we presented.

Our main contributions in this paper are: first to introduce a new data collection method for geospatial data to process satellite images represented by polygon-based training data, which is an intermediate method between pixel-based method and image-based method, characterized by an increasing number of total pixels with diverse representation for the target class. Second, to introduce a simple convolutional neural network based on a 1x1 convolutional filter to re-encode the input data and also add a new intermediate filter to the output layer which improves the resulting semantic map.

The remainder of this paper is organized as follows: in section 2, we provided the details of Building geospatial dataset for training and testing our proposed model. In section 3, it includes the enhanced 1DCNN architecture. The evaluation methods are explained in section 4. In section 5, we clarified the practical experiments, including the study area, the different settings for the experiments, the specifications of the training data and their types, and finally the different comparisons between the proposed method and the other methods, numerically and visually. Section 6 concludes the paper.

2. Building spatial ground truth dataset

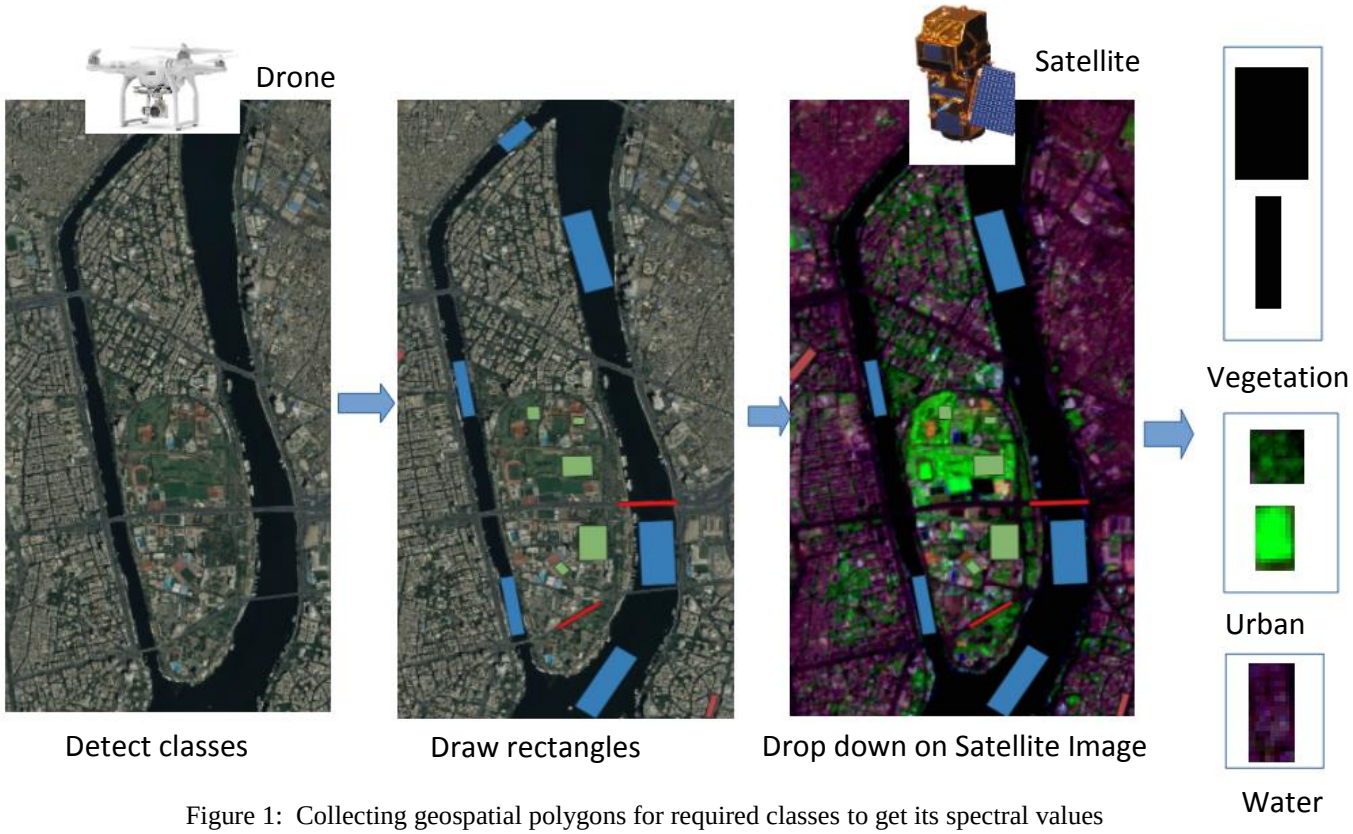


Figure 1: Collecting geospatial polygons for required classes to get its spectral values

Working with satellite imagery, the best way to define a training dataset is to use a geospatial polygon around the classes to be identified. We only relied on rectangular shapes as their pixels then used them to train our model. Error! Reference source not found. illustrates the steps of collecting and registering data for the different classes. Firstly, a group of experts move to the work field using either the naked eye or a drone to draw the polygons of each class using one of a mobile application. Secondly, geospatial polygons are placed on the satellite images scanned on the same day and subsequently cut out the parts of the image with specifying the class for each part to be used after that in building the training images. The process of collecting ground truth data is accompanied by satellite images so simple classes can be predicted from them.

After examining the different rectangles for each class of classes to be classified by popular GIS platform. We take the pixels in each rectangle and put the class label on it as shown in Figure 2. We have many pixels per class that expresses the diversity in the pixel signature of this class, which improves the classification ability of the model used.

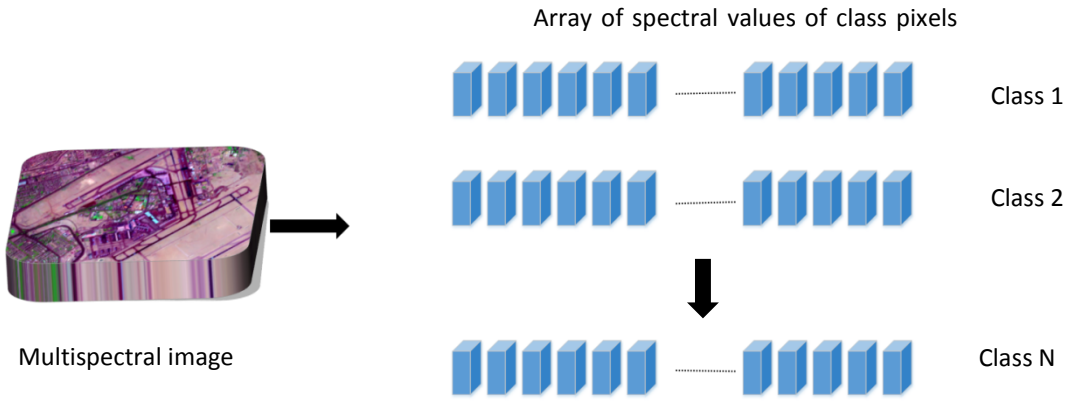


Figure 2: create pixel dataset for each class

3. Enhanced 1-DCNN Architecture

The proposed approach is formed of a simple convolution neural work architecture. It is formed of two convolution layers, one fully connected layer, and the final median filter layer as post processing step. The detailed structure of the proposed architecture shows in

Figure 3.

After specifying training pixels of each class using rectangle polygons, these pixels are stacked into one-dimension vector with its class labels. Each pixel consists of 10 spectral values. The input layer is fed to 1D convolution with 64 convolution filters with 1x1 (1-Dimension) with one step stride and the same padding. ReLU (Rectified Linear Unit) as an activation function is applied to generate the output of the first convolution layer, the vector depth is 64. The second convolution layer is formed of 56 convolution filters with 1x1, 1 stride, and the same padding with the same activation function. The output of the second convolution layer is a vector with depth of 56. The output of the second convolution layer is fed to the first fully connected layer with 160 neurons. Then then second fully connected layer with 160 neurons. The output of the fully connected layers is fed to the softmax layer with cross-entropy as a loss function to get the first output semantic map. A new median filter layer is added to remove noise-like values to produce the final output classification map.

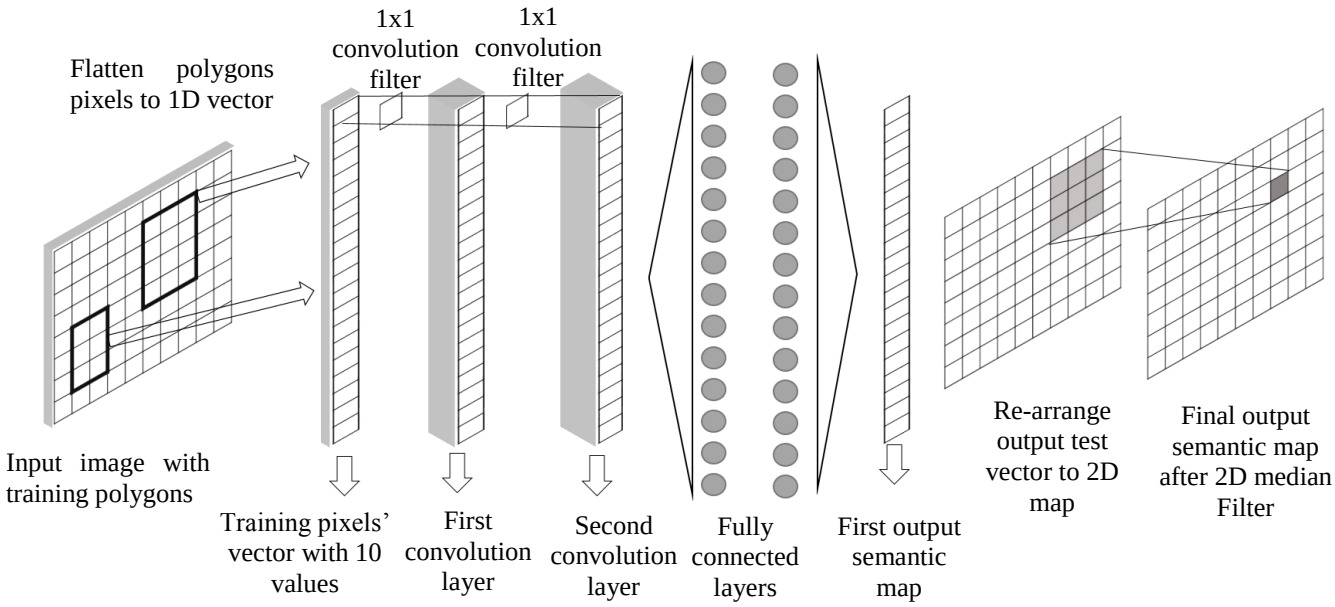


Figure 3: Proposed architecture of enhanced 1DCNN

3.1. One-Dimension Convolution Layers

A one-dimensional convolution layer is based on a 1x1 pixel filter. With multiple filters of this type, a new multi-band image is produced by the number of filters used. The convolution process becomes a re-coding of pixel values in a larger range, as the number of filters used may reach up to 64. Figure 4 illustrates the process of one dimension convolution. The purpose of the training process becomes to obtain the best values for these filters to achieve the highest possible result.

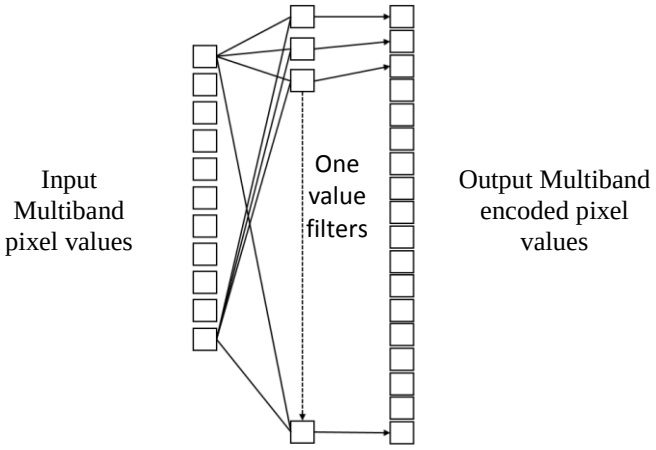


Figure 4:1x1 Convolution filter - One Dimension Convolution Process

Fully connected Layer

A fully connected layer is designed to perform the required classification between convoluted layers and classification maps. It takes auto-generated features from convolution layers. Therefore, it contains a larger number of parameters and is more difficult in the training process. Therefore, when designing it takes care to be a few in the number of neurons as possible. The designed Fully connected layer is formed of two layers. Each one of them is formed of 160 neurons. On top of these layers, there is a softmax layer to calculate the cross-entropy loss.

3.2. Median Filter Layer

We have added a new layer to the network structure which is formed of a median filter. Its filter is designed to remove noise like value from the resulted classification map calculated by the softmax layer. This filter makes the resulted image much smoother. It gets the most dominant value in the region of filter either 3x3, 5x5, or 7x7. Figure 5 shows the process of removing noise-like values by using the dominant value in its neighborhoods.

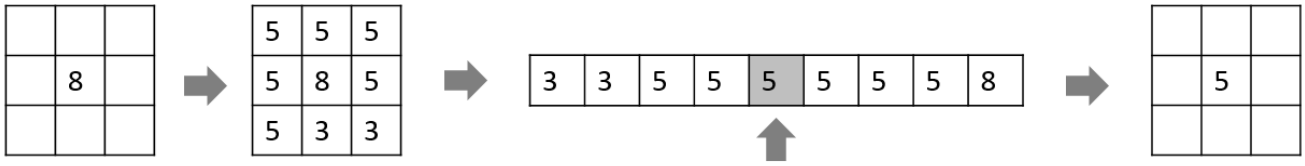


Figure 5: Removal of noise like value using median filter

4. Evaluation Metrics

We rely on rectangular areas to train the proposed method, and to determine their accuracy and compare them to other methods, we find the total pixels contented in each rectangular area in the test set. Hence, the number of predicted pixels and ground truth pixels of all kinds are used in calculating accuracy. We have used evaluation metrics that is popular in the corresponding business, namely, recall and precision and F1-score. Equations 1 to 3 specify the method for calculating each metric.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

Also, we depend on macro accuracy rather than the weighted accuracy. Macro accuracy is calculated by averaging the individual accuracy of each class as there are unbalanced distribution of classes.

5. Experimental Results

To determine the efficacy of the proposed method, we made several experiments compared with other methods and changed a large number of parameters to get the best one. All of these experiences are carried out using an area of significant diversity in urban classes.

5.1. Study Area and Material

Our study area includes the region of Great Cairo satellite images collected by the Sentinel-2 satellite with a resampled spatial resolution a 10-meter starting at July 2020 as in Figure 6. Ten bands are selected from Sentinel-2 satellite image bands to form spectral values of pixel. Radiometric and geometric correction are applied to the resulted mosaic. Mosaic width is 6324 pixels and height is 3330 pixels. The area of mosaic is about 2100 square kilometers.

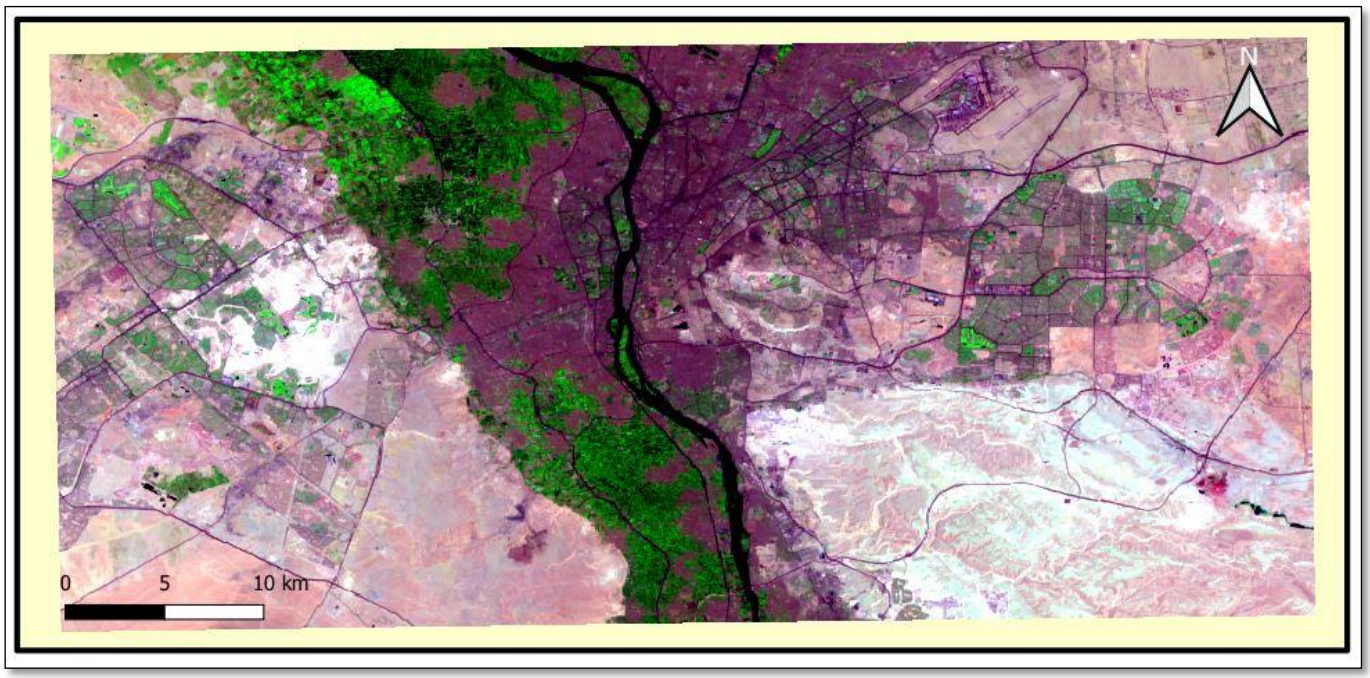


Figure 6: Great Cairo sentinel 2 Satellite

5.2. Implementation Details

We have implemented our Enhanced 1DCNN framework using Keras based on TensorFlow as a backend which is pre-installed on Linux operating system with NVIDIA Tesla K80 GPU and 12GB RAM located in Google Cloud Platform (GCP). We used a learning rate equal to 0.0002 using 200,000 iterations with patch size equal to 64. Meanwhile, we used ADAM optimizer to update weights of the Enhanced 1DCNN model to reduce back propagation losses.

5.3. Dataset Specifications

We have four classes in urban areas plus a class under construction. This is in addition to four general classes: water, desert, roads, and vegetation. The total number of classes is nine classes.

The urban area includes four classes, starting from the most crowded and least green areas to the least crowded and greenest areas. Class 1 represents the urban blocks like Bulaq El-Dakrour, Kafr Tahmas, and Saft Al-Laban where housing is unplanned and green spaces are very few. Class 2 includes areas such as Mohandessin, Nasr City, and Heliopolis, where the urban blocks are more organized and the green spaces are more than class 1. As for class 3 many areas of the new cities, such as Sixth of October, New Cairo, where the buildings are organized and more green spaces. The class 4 is more organized, greener, than other areas, such as some New Cairo and Al-Sheikh Zayed areas such as Al-Rehab and closed residential compounds. Figure 7 shows examples of different classes. Table 1 illustrates the details of each class number of pixels and percentage in training and testing datasets. The percentage of each class is propositional to its occurrences in the total image.

Table 1: Details of each class number of pixels and its percentages in training and testing datasets

	Training Pixels		Testing Pixels	
Class	Number of Pixels	Percentage	Number of Pixels	Percentage
Under Construction	52366	6.7%	22425	3.8%
Urban1	24233	3.1%	18133	3.0%
Urban2	16339	2.1%	7020	1.2%
Urban3	21616	2.8%	15746	2.6%
Urban4	5273	0.7%	3473	0.6%
Roads	22558	2.9%	12681	2.1%
Desert	560868	71.5%	445164	74.8%
Vegetation	56871	7.3%	33308	5.6%
Water	23971	3.1%	37402	6.3%
Total Number	784095		595352	

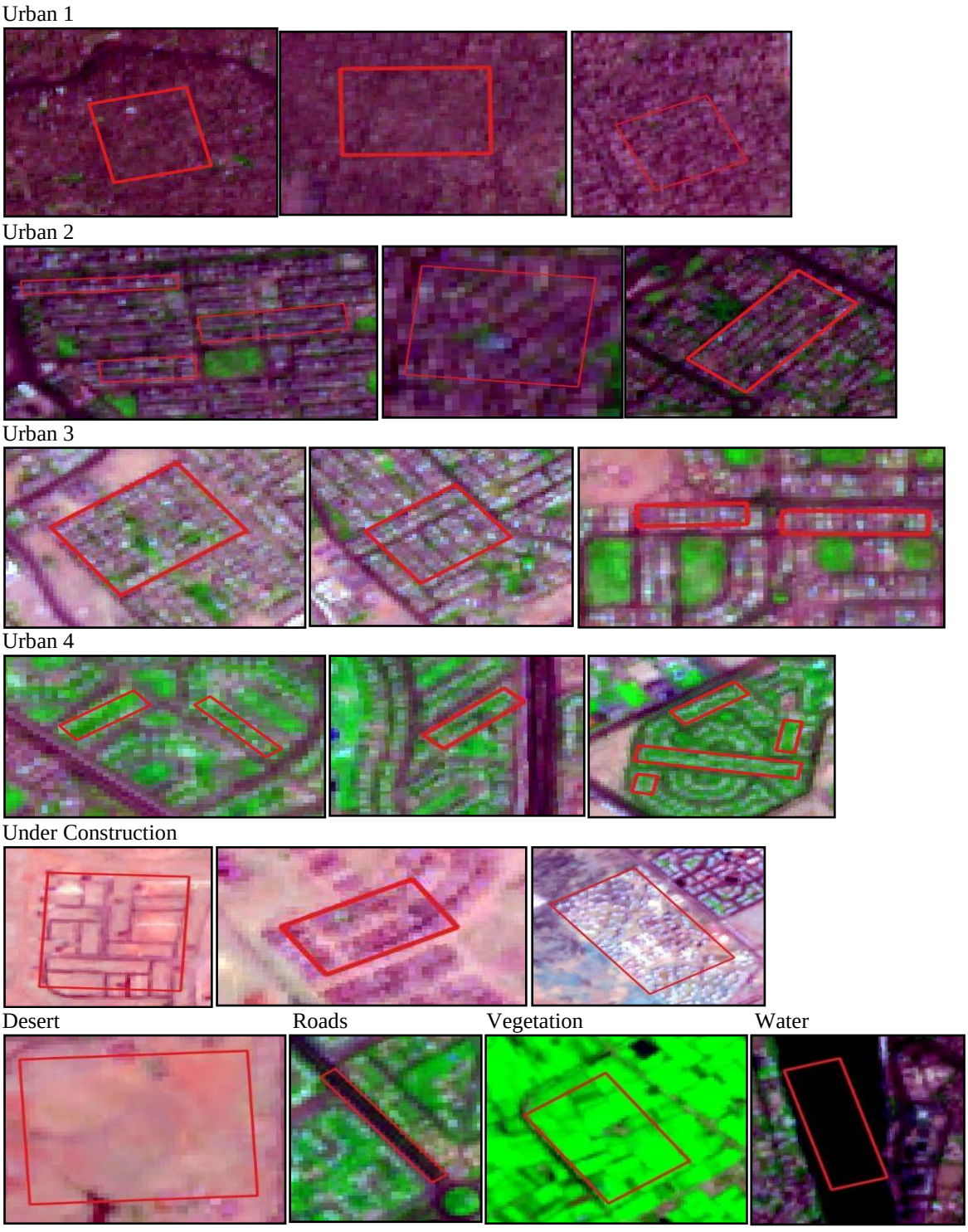


Figure 7: Different rectangles examples for urban classes and other classes

5.4. Accuracies Comparison

The use of Enhanced 1DCNN has shown remarkable progress in the results we obtained compared to other methods either statistical methods as Random Forests (RFs) or Non-statistical methods as k-Nearest Neighborhoods (kNNs). These methods are the most used methods in satellite image processing platforms as ERDAS Imagine®, ENVI® and Trimble eCognition® Suite. Figure 8, Error! Reference source not found., and Figure 10 illustrate the precision, recall, and F1-score, respectively. Figures shows the comparison between the proposed methods and traditional methods of pixel-based methods namely kNNs and RFs. The k neighbors for kNNs is 3. RFs parameters are number of jobs and number of estimators. Number of jobs is 5 and number of estimators is 45. The enhanced DCNN shows a remarkable improvement for almost all classes and the overall accuracies is macro or weighted accuracies.

Figure 11 shows a visual comparison between proposed method and the common pixel-based methods. Two different regions were used for comparison. The first is East Cairo and the other is Downtown. This possibility contains many classes. The proposed method showed results more smoothly and with higher accuracy than other methods. Figure 12 shows the result of classification of Greater Cairo.

5.5. Results Discussion

The proposed method is based on pixel-based image classification, meaning it is based primarily on pixel classification. In the final stage, the image is also classified pixel by pixel. This depends on the ability of the classifier to make the correlation between the input pixel signature and the corresponding semantic class. The main benefit of pixel-based methods in the field of remote sensing is the speed and accuracy in collecting field data. The proposed method took advantage of data collection using polygons, which increased the number of collected data, and at the same time, it was converted to pixels. Other methods of Deep learning rely on image-based classification, which is the most difficult and most expensive.

Regarding to other pixel-based methods, it depends primarily on the pixel signature itself represented here by 10 spectral values. While the proposed method is based on these spectral values, it has relied on three strategies to find out the correlation between these spectral values and their semantic classes. The first is re-coding these values using 1x1 convolution filters, the second is the use of a median filter, which removes the unwanted pixels, and the third is the use of fully connected layers, which was able to know the invisible relationship between the input and output. All this in the end led to the superiority of the proposed method over the rest of the pixel-based methods.

From the review detail of the results, we see that the basic classes were easier to distinguish between the overlapping urban classes, and the easiest of them was water. Between them, we find the distinction between the different urban classes, which differ according to the overlap between them and the other classes, where urban1 was the highest in distinction, urging concrete structures and the least green areas, followed by urban 4, where green areas and organized form, while we find urban 2 and urban 3 is the least where Overlap with other categories increase.

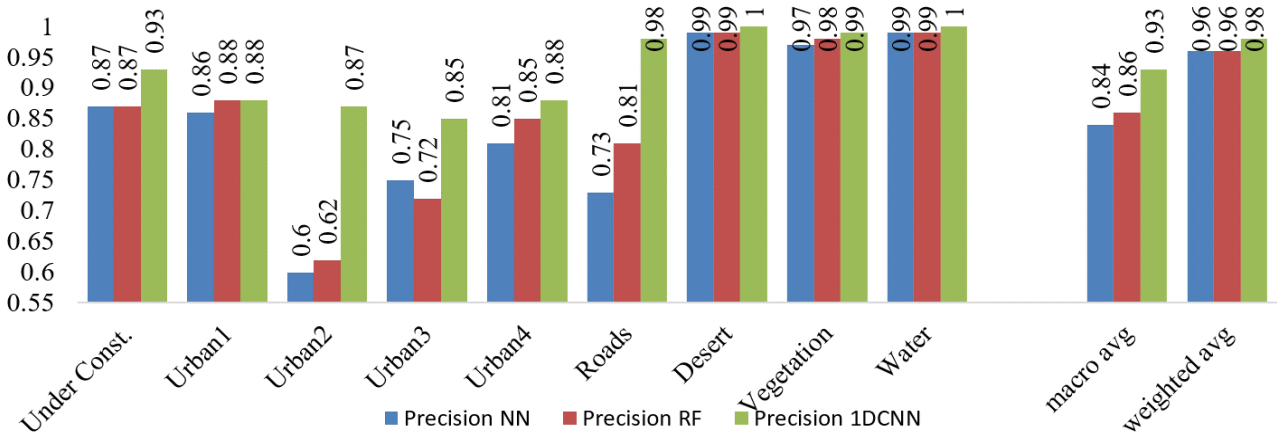


Figure 8: Precision

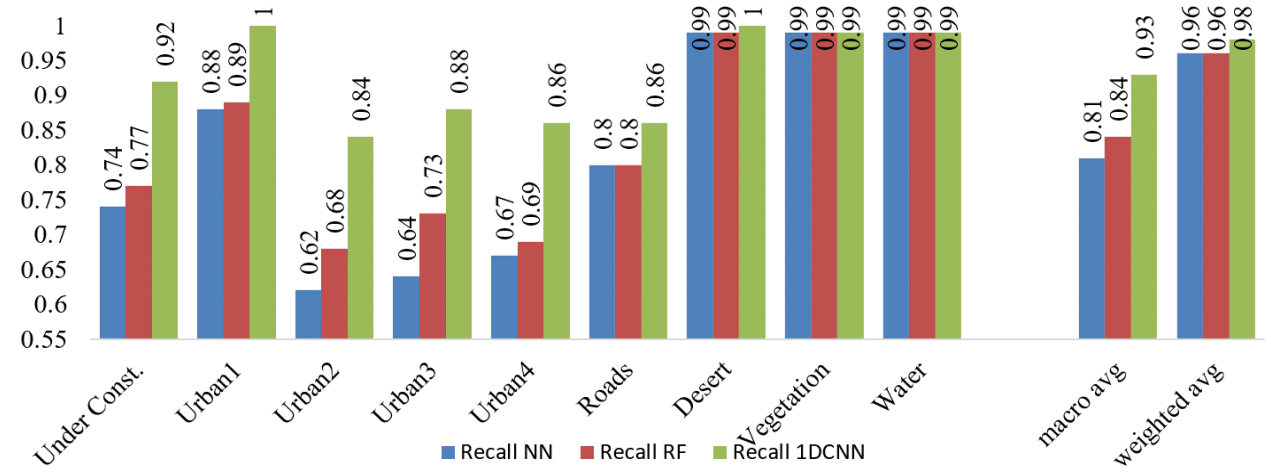


Figure 9: Recall

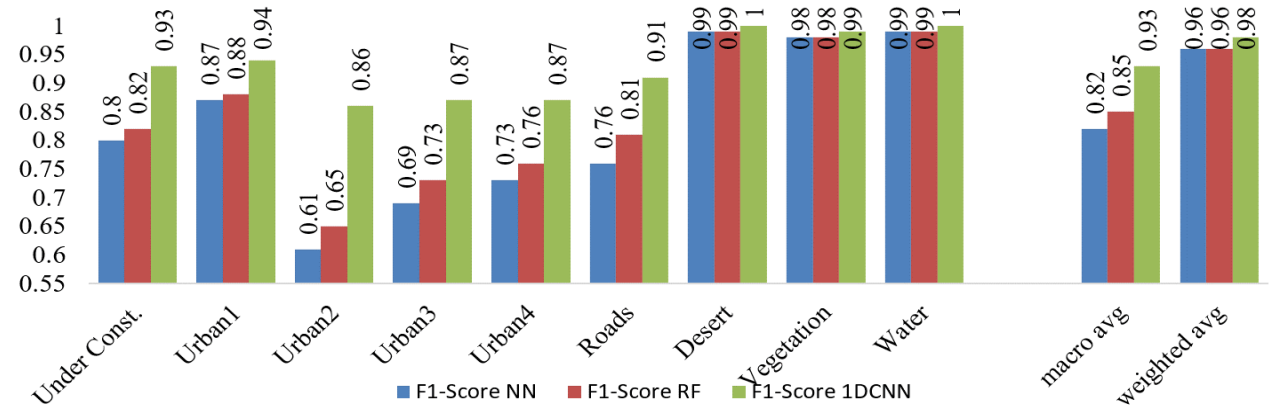


Figure 10: F1-score

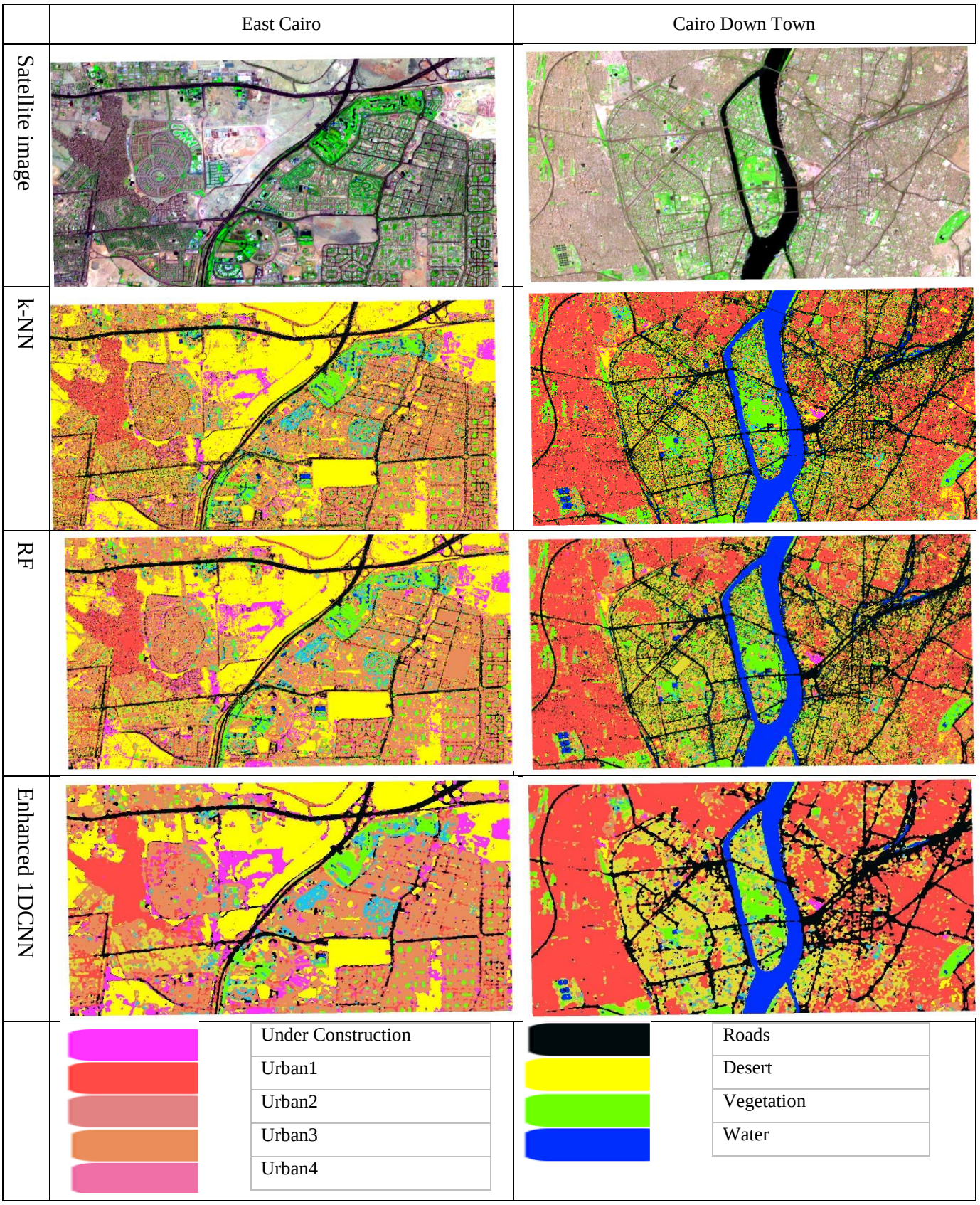


Figure 11: Classification maps comparisons relevant to satellite images

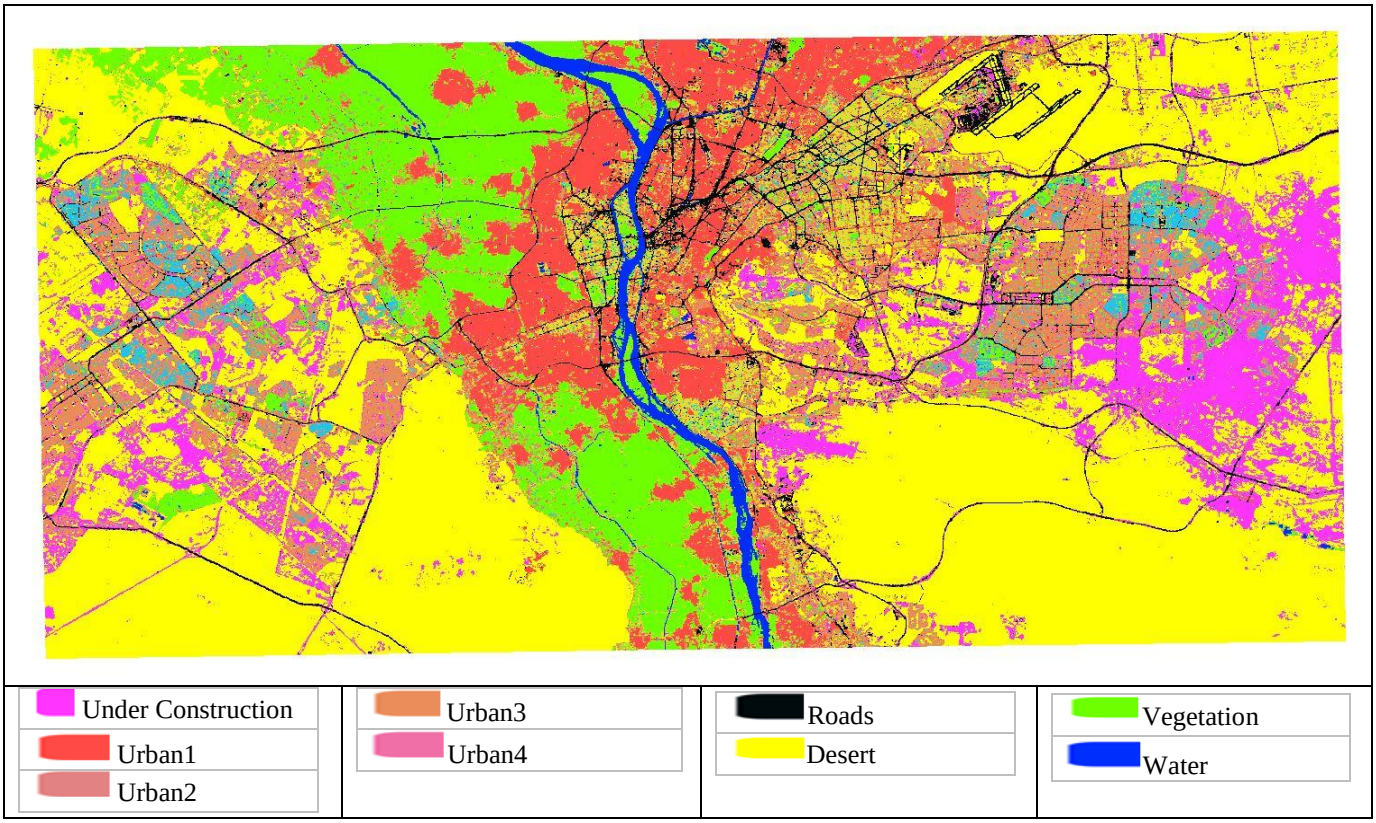


Figure 12: Overall classification map of Greater Cairo using proposed Enhanced DCNN.

6. Conclusion

In this paper, we proposed an enhanced convolutional neural network based on three enhancement processes. First, using a geospatial polygon to gather training data for training enabled us to collect the largest amount of training data in an easy and straightforward way using GIS software. Second, using a 1-dimension convolutional neural network had a great ability to learn based on a pixel-based approach. Third, using a median filter for the results removed the single wrong points caused by the pixel-based approach and made the classification results smoother. We use the proposed method to classify the satellite image of Sentinel 2 for Greater Cairo into five different urban classes with other common classes as desert, vegetation, and water. The proposed method showed a significant increase in terms of precision, recall, and f1- score compared to other pixel-based methods.

References

[1] Q. Liu, M. Kampffmeyer, R. Jenssen, and A. B. Salberg, "Dense dilated convolutions merging network for land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6309–6320, 2020, doi: 10.1109/TGRS.2020.2976658.

[2] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, no. March 2019, pp. 94–114, 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[3] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57,

no. 9, pp. 6690–6709, 2019, doi: 10.1109/TGRS.2019.2907932.

[4] K. C. Ukaoha and E. C. Igodan, "Architecture Optimization Model for the Deep Neural Network," *Int. J. Intell. Comput. Inf. Sci.*, vol. 19, no. 2, pp. 1–16, 2019, doi: 10.21608/ijicis.2019.96101.

[5] A. Al-furas, M. AL-dosuky, and T. Hamza, "Improving Feature Maps in Early Layers of Convolutional Neural Networks Using Otsu Method," *Int. J. Intell. Comput. Inf. Sci.*, vol. 16, no. 2, pp. 37–45, 2018, doi: 10.21608/ijicis.2018.10905.

[6] G. Hamed, M. Marey, S. Amin, and M. Tolba, "Comparative Study and Analysis of Recent Computer Aided Diagnosis Systems for Masses Detection in Mammograms," *Int. J. Intell. Comput. Inf. Sci.*, vol. 21, no. 1, pp. 33–48, 2021, doi: 10.21608/ijicis.2021.56425.1050.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS'12 Proc. 25th Int. Conf.*, vol. 1, pp. 1–9, 2012, doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[9] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017, pp. 4278–4284.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[11] N. Numan, S. Abuelenin, and M. Rashad, "Prediction of Lung Cancer Using Artificial Neural Network," *Int. J. Intell. Comput. Inf. Sci.*, vol. 16, no. 2, pp. 1–19, 2018, doi: 10.21608/ijicis.2018.10013.

[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *32nd International Conference on Machine Learning*, Jun. 2015, pp. 448–456, Accessed: Aug. 21, 2017. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html.

[13] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[14] G. Hinton, "Dropout : A Simple Way to Prevent Neural Networks from Overfitting," vol. 15, pp. 1929–1958, 2014.

[15] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, 2021, doi: 10.1016/j.ymssp.2020.107398.

[16] N. Laban, B. Abdellatif, H. M. Ebeid, and H. A. Shedeed, "Seasonal Multi-temporal Pixel Based Crop Types and Land Cover Classification for Satellite Images Using Convolutional Neural Networks," no. 2017, 2018.

[17] R. Mohammed, O. Nomir, I. I. Khalifa, and T. Hamza, "a System for Acute Leukemia Cells Segmentation and Classification," *Int. J. Intell. Comput. Inf. Sci.*, vol. 16, no. 4, pp. 79–87, 2016, doi: 10.21608/ijicis.2016.19829.

[18] G. Cheng, J. Han, and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Feb. 2017, doi: 10.1109/JPROC.2017.2675998.

[19] B. Cui, X. Chen, and Y. Lu, "Semantic Segmentation of Remote Sensing Images Using Transfer

Learning and Deep Convolutional Neural Network with Dense Connection," *IEEE Access*, vol. 8, pp. 116744–116755, 2020, doi: 10.1109/ACCESS.2020.3003914.

[20] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sens.*, vol. 12, no. 2, pp. 1–25, 2020, doi: 10.3390/rs12020207.

[21] C. Yao, X. Luo, Y. Zhao, W. Zeng, and X. Chen, "A review on image classification of remote sensing using deep learning," *2017 3rd IEEE Int. Conf. Comput. Commun. ICCC 2017*, vol. 2018-January, pp. 1947–1955, 2018, doi: 10.1109/CompComm.2017.8322878.

[22] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep Recurrent Neural Networks for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017, doi: 10.1109/TGRS.2016.2636241.