

Urban monitoring application in new administrative capital in Egypt using machine learning techniques and Sentinel 2 data

Tamer ElGharbawi^{1,*}, Mohamed Talaat^{2*}, Ahmed Amin³, Ahmed A. ElSharkawy⁴

¹ Civil Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt, email: tgh@eng.suez.edu.eg

² Civil Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt, email: Mohamed_talaat@eng.suez.edu.eg

³ Civil Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt, email: profahmedamin@yahoo.com

⁴ Civil Department, Faculty of Engineering, Port Said University, Port Said, Egypt, email: a.elsharkawy@eng.psu.edu.eg

* Mohamed Talaat, Email: Mohamed.talaat.fathy@gmail.com, DOI: 10.21608/pserj.2022.171168.1200,

ABSTRACT

Recently, Egypt has been undergoing rapid development in the urban and transportation sectors with a focus on the New Administrative Capital. This requires temporal monitoring and assessment for such projects with low cost, high accuracy, and simple application. Therefore, sentinel-2 data was utilized to assess the effectiveness of several methods for urban classification and monitoring. The study discusses three main approaches: histogram threshold, spectral indices, and machine learning techniques. In machine learning, we focus on K-nearest neighbor (KNN), Linear discrimination analysis (LDA), and Random Forest (RF) due to their robustness and simplicity. It was found that the RF technique presents the highest accuracy for our study region with 98.8%. Moreover, this study monitored the effect of such development on air pollution in the New Administrative Capital using sentinel-5P data which showed an increase in the CO and NO₂ levels during the last five years by nearly 11% and 75% respectively, with a noticeable drop during the year 2020 due to the Covid19 lockdown. The classification model was trained and verified using ground truth data obtained on the 10th of Ramadan City. Then, it was applied to monitoring urban development and the associated environmental impact in the New Administrative Capital in Egypt.

Keywords: Random Forest, K-nearest neighbor, Linear discrimination analysis, Sentinel-5P

Received 27-10-2022,

Revised 16-11-2022,

Accepted 16-11-2022

© 2022 by Author(s) and PSERJ.

This is an open-access article licensed under the terms of the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1 INTRODUCTION

Urban and transportation networks monitoring (UTNM) is an integral part of numerous applications, such as urban damage assessment [1], the production of digital maps (Sultanov et al.,2018), land use categorization & verification [2], modernizing Geographic Information System (GIS) databases, and route planning [3].

Provincial planning and city modeling define the proper localization of urban regions and transportation networks. Moreover, the realization of the location, size, and density of constructions can be quite beneficial in estimating the dispersal of a city's population. For these reasons, UTM can be utilized to analyze the size and

location of civilized settlements in different disorganized areas.

Traditional methods such as field surveying are too expensive and time-consuming for UTM. In comparison to traditional approaches, satellite photography delivers critical information about the Earth's surface at cheaper prices, with higher temporal frequency, larger spatial coverage, and faster processing speed. Recent research utilized medium-resolution optical satellite imagery like Landsat-8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) and Sentinel-2 (S2) Multi-Spectral Instrument (MSI) for several applications such as wildfire assessment [4], [5], land use and land cover mapping [6] and vegetation monitoring [7]. Very High Resolution (VHR) satellite sensors such as Worldview-2 and GeoEye are used for mapping forest burn severity across spatial scales in a

Pine Barrens ecosystem [8] and estimating the severity of burnt regions. The Case of the 2011 Evros Wildfire in Greece Geoeye-1 [9] provides sub-meter spatial resolution and unique image quality. However, with longer revisit times and higher costs, these satellites were used most frequently for detailed knowledge of the types and coverage of intra-urban features [10].

Sentinel-2 has a considerably wide spectral range, narrow bandwidths, reliable radiometric characteristics, high spectral resolution data, free data accessibility, and reasonable revisit time. Therefore, it is the most adequate platform for the realization of a system capable of continuous assessment, and monitoring of urban regions and transportation networks. Nevertheless, UTNM using satellite images can be quite challenging due to differences in terms of lighting conditions, quality, and resolution. Additionally, accurate classification results require high-resolution images [11].

Valuable research was presented utilizing medium-resolution satellite images for monitoring the urbanization development rate. Saber et al., (2021) utilized Landsat-8 data and maximum Likelihood classification for urban classification of the new administrative capital in Egypt, with a classification accuracy of 93%. It was not recommended to use Landsat 8 images to detect a change if the results were required with high accuracy. Additionally, Muhammed et al., (2021) used RGB image and machine learning techniques for rooftop extraction with an accuracy of 98%. El-Zeiny et al., (2022) utilized Landsat-8 and supervised/unsupervised classification for monitoring of coast of Port Said Governorate in Egypt with an accuracy of 85.7%.

Moreover, [14] used AVIRIS-NG hyperspectral data and Artificial Neural Network for spectral-spatial urban target detection with an accuracy of 92.4%. Bakhtiari et al., (2017) utilized high resolution for semi-automatic road extraction from digital images with an accuracy of 81%. [16] used world view-2 for road network extraction with an accuracy of 88%.

This paper presents the assessment results of several monitoring approaches for urbanization development rate utilizing the freely available sentinel-2 data with the application to the New Administrative Capital in Egypt, which is considered a priority due to its high development rate and the keen interest of the Egyptian government. The tested approaches were histogram thresholding, spectral indices, and machine learning techniques. Histogram thresholding is the simplest approach to image segmentation; the key parameter in the thresholding process is the selection of the threshold value and the application bands. Spectral indices segmentation is defined as ratios, differencing, rationing differences, and sums, or by forming linear mixtures of spectral band data at a few wavelengths. Another widely used method for extraction of UTNM is Machine-learning techniques, which are normally capable of modeling complex class signatures, accept a variety of

input predictor information, and do not make assumptions about the data distribution (i.e. nonparametric) [17].

Air pollution monitoring results of the New Administrative Capital in Egypt during the last five years (from 2018 to 2022) using sentinel-5P data shows an increase in the carbon monoxide (CO) and nitrogen dioxide (NO₂) levels by nearly 11% and 75% respectively. The main sources of these contaminants are manufacturing, transportation facilities, and power generation plants. Therefore, the emergence of the TROPOMI/Sentinel-5 Precursor satellite, which was launched in October 2017 by The European Commission's Copernicus program, was a pinnacle moment in the continuous monitoring of air pollution. The goal of this research is to evaluate the classification accuracy of the medium-resolution satellite image, as well as to monitor the pollution levels in the New Administrative Capital.

2 STUDY REGION

Figure 1, region (A) illustrates the New Administrative Capital, a large-scale and major project of a novel Egyptian capital city that has been under construction since 2015. It was declared on March 13, 2015, by Egypt's then-housing minister, Mostafa Madbouly, at the Egypt Economic Development Conference. The capital city is one of the economic development programs and projects that are part of a larger initiative known as Egypt Vision 2030.

The novel capital city is located 45 kilometers (28 miles) east of Cairo, just beyond the Second Greater Cairo Ring Road, in a relatively under-developed territory halfway to the Suez seaport metropolis, with coordinates of 30 02' 12" N and 31 28' 24" E. Because it houses ministries, foreign embassies, and significant government offices, this city is expected to become Egypt's new administrative and financial capital. It will also be divided into 21 residential districts and 25 dedicated districts, covering a total area of 700 square kilometers (270 square miles) and housing 6.5 to 7 million people. Moreover, its downtown contains skyscrapers such as Obelisco Capitale, designed in a shape of a Pharaonic obelisk and will tower at the height of 1,000 meters (3,300 ft), becoming the tallest building in the world; and Iconic Tower, which will be the tallest in Africa. A central and innovation park, artificial lakes, 2,000 educational foundations, 1,250 mosques and churches, 663 hospitals, 40,000 hotel rooms, a 90,000-seat stadium, a large theme park four times the size of Disneyland, 90 Km² of solar energy farms, an electric railway link with Cairo, and a new international airport at the site of the Egyptian Air Force's existing Wadi al Jandali Airport are also planned. It will be built in the smart city style. It is estimated that the transfer of presidential residences, government ministries, foreign embassies, and parliament will cost USD 45 billion. The

project's overall cost and timetable have not been revealed. The administrative capital area was chosen because it is a new city, as the development that took place during the past 5 years, especially in buildings and road networks, can be monitored and followed up [18].

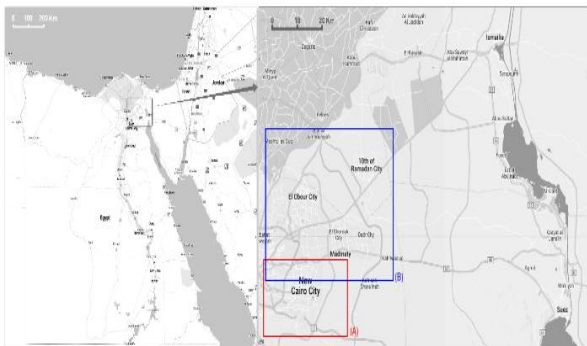


Figure 1: The location of the Study Area was (A) the application area (the New Administrative Capital) and (B) the training area (10th of Ramadan city)

Figure 1B depicts the 10th of Ramadan, which was used as a training and testing model in a city in Egypt's Sharqia Governorate. It is one of the most industrialized of the first generation of new urban communities. It is close to Cairo, which is 46 kilometers (29 miles) away and considered part of Greater Cairo. It is located on the Cairo-Ismailia desert highway. It was established in 1977 by Presidential Decree No. 249. There are numerous industrial zones in the city, with industries such as electronics, food processing, ready-made garments, plastic, paper, textiles, construction materials, steel, medicines, and furniture. It's located at Latitude $30^{\circ}10'08''N$ and Longitude $31^{\circ}35'55''E$ with a total area of 94.76 km². The image was taken from the Google Earth engine that used 10 m resolution imagery for the study area. The study area was chosen rather than any other area because it is an area that contains all the six different classes in a clear way through which we can use the training. It contains the area of cities and road networks, water, desert, rocks, and water [19].

3 MATERIAL AND METHODS

3.1 Dataset

The Sentinels are an ESA-improved network of satellites that provide high-resolution optical images from Sentinel-2A and 2B, ocean and land data appropriate for environmental applications, and air quality data from Sentinel-5P. Sentinel-2 is a European Space Agency (ESA) optical imaging mission for the operational monitoring of land and coastal areas. Sentinel-2 is a constellation of satellites orbiting in a polar sun-synchronous orbit. As shown in table 1, Sentinel-2 will provide a unique combination of universal coverage (290 km), high revisit (5 days with two satellites), high spatial resolution (10m), and multi-

spectral imaging (13 visible and shortwave infrared spectral bands). The Sentinel-5P is a sun-synchronous, near-polar orbiting satellite that flies at an altitude of 817 km above the earth's surface at an ascending point with an equator crossing time of 13:30 (local time), providing daily global coverage. Three alternative data streams could be used to monitor air pollutants using TROPOMI: near-real-time (NRTI), offline (OFFL), and reprocessing (RPRO) [20][21]. The images were obtained using Google Earth Engine, a cloud-based platform for planetary-scale geospatial analysis that applies Google's massive computational capabilities to a variety of high-impact societal issues such as deforestation, drought, disaster, disease, food security, water management, climate monitoring, and environmental protection. It is a one-of-a-kind integrated platform in the field, designed to empower not only traditional remote sensing experts but also a much broader audience who lack the technical capacity to use traditional supercomputers or large-scale commodity cloud computing resources [22]. The Sentinel-2 images were acquired in October of 2021 for model training in the training area. Regarding the application area, we acquired Sentinel-2 data for land use classification over four years starting from 2018 to 2021, and Sentinel 5P data for air pollution monitoring over five years starting from 2018 to the first two months of 2022. The image was acquired in October because there was less cloud cover in this period.

Table 1. Resolution and wavelength for all bands (sentinel-2)

Band	Resolution	Central Wavelength
B1	60 m	443 nm
B2	10 m	490 nm
B3	10 m	560 nm
B4	10 m	665 nm
B5	20 m	705 nm
B6	20 m	740 nm
B7	20 m	783 nm
B8	10 m	842 nm
B8a	20 m	865 nm
B9	60 m	940 nm
B10	60 m	1375 nm
B11	20 m	1610 nm
B12	20 m	2190 nm

3.2 Software

All the processing was done using in-house developed software based on Python and QGIS. All the multispectral images analysis were performed using Python V3.8. In this study there are many different Python-based scientific and numeric libraries were used such as NumPy, computer vision Scikit-image library [23], OpenCV library [24], and machine learning Scikit-learn library [25]. The image processing and machine learning model were generated on Windows 10 laptop Intel Core i7 5th generation, quad-core (2.20 GHz) with 8 GB RAM.

3.3 Methodology

A workflow procedure for analyzing images under study in this research was shown in **Figure 2**. The acquired image for training was processed on Google Earth Engine for spectral correction to get surface reflection then the fog and clouds were masked from the image. After that, the RGB image was extracted from the multispectral image on Python. It was used to select the training pixels as shown in **Figure 3**. The features of selected pixels were extracted to an excel file as a two-dimensional matrix, the first dimension represents the pixel number and the second one represents features (spectral features). Before training, the extracted pixels (250000 pixels less than 0.1% of total image size) were pre-processed to remove outliers using Principle Component Analysis (PCA) [26] to filter the data. PCA is based on the Eigenvector of the variance or the covariance matrix of the merged data set. PCA merits reduce the data redundancy between bands and emphasize the different information in the derived components [26] after filtering the data, every pixel was coded as follows to represent ground truth where label one refers to the desert, label two refers to the green area, label three refers to buildings, label four refers to a rock, label five refers to transport networks, and label six refers to the water area.

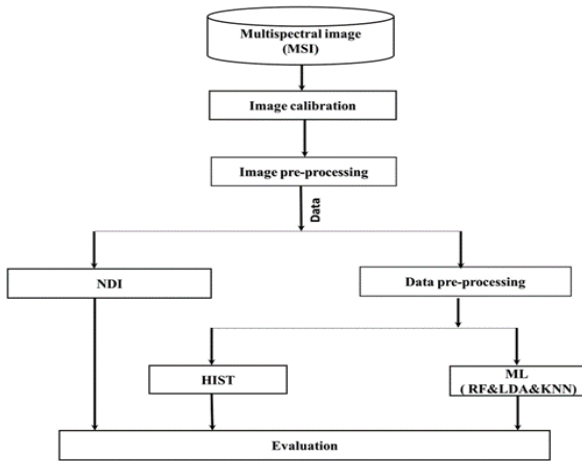


Figure 2: Image processing flowchart



Figure 3: Ground truth for the training area and six classes are differentiated: desert, green, buildings, rocks, roads, and water

3.3.1 Histogram Thresholding

The histogram (HIST) depicts the distribution of each class within one band, indicating that the first class, for example, water, has a reflectivity ranging from how many to how many within the first band, and a value ranging from how many to how many within the second band, and so on for all classes, allowing us to explain the best band that can separate the classes. Histogram distribution for each separate band of the image is first plotted. It is a graph in which the horizontal axis shows the tonal variations and the vertical axis shows the number of pixels in that tone [27]. Band 12 has the greatest segment results, although there are still some misclassifications across classes as shown in **Figure 4**.

Figure 5 displays the segmented results at band 12 using a threshold value of (2000–3500), although there are misclassified pixels between Urban and Transport Networks, which is consistent with the result shown in the histogram at band 12 as shown in **Figure 4**.

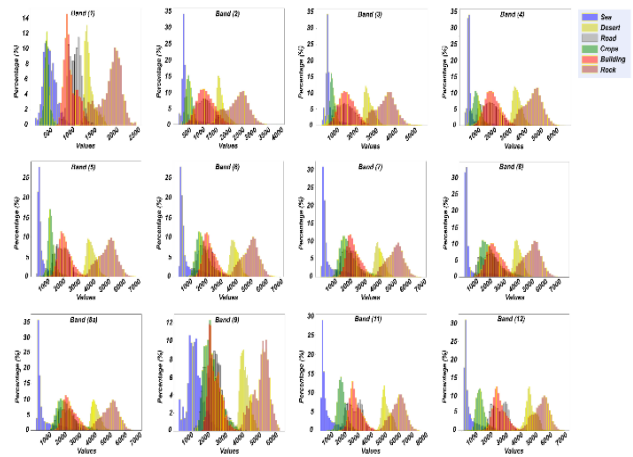


Figure 4: Results of HIST showing band 12 has the best result

All of the classes overlap in all of the bands, where there is overlap between the 6 classes in Band No. 1, and so on in all of the bands, but there is almost an overlap between roads and buildings in Band No. 12, so it was estimated that the best band to separate roads and buildings from the rest of the classes. Although the presence of overlap between the two classes, in Band No. 12 because the histogram was ineffective in the segmentation process.

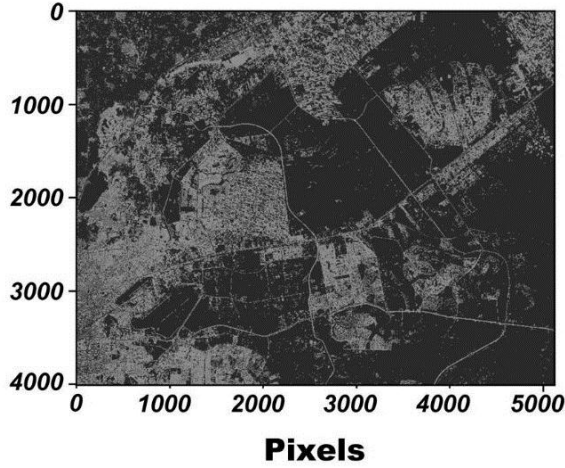


Figure 5: Segmentation results using HIST for transportation networks

3.3.2 Spectral Indices (NDI)

Spectral indices are used to show the relative abundance of characteristics of interest that may be identified by variations in the surface reflectance values of two or more bands [10]. This study creates the NDI to answer the previously mentioned urban issues. It's a linear relationship between all two bands, with the normal indices determined as the subtraction of two bands divided by their total.

$$INDI = \frac{B_n - B_m}{B_n + B_m} \quad (1)$$

Where: B_n is the band at a specific wavelength, and B_m is the band at another wavelength.

As index-based, it was compared between the different combinations where band (1) and band (5) achieved the best results for segmentation. Figure 6 shows the segment results after choosing the threshold value because the NDI was ineffective in the segmentation process.

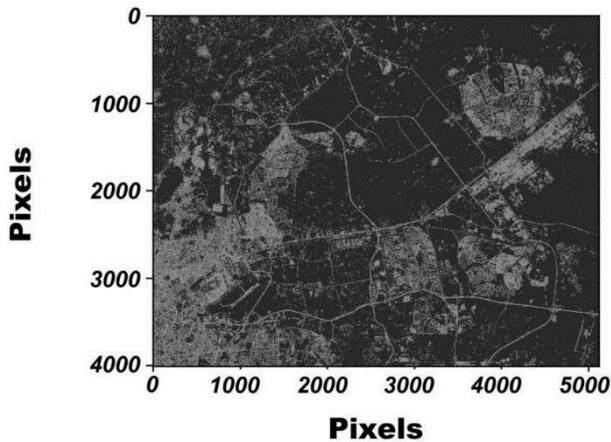


Figure 6: Segmentation results using NDI for transportation networks

3.3.3 Machine Learning Techniques

Different machine learning algorithms named (K-nearest-neighbor (KNN), linear discrimination analysis (LDA), and random forest (RF)) were compared to get the best one for the categorization classes understudy. The preprocessed pixels were used as an independent variable (X-matrix), where the labeled number for every class was used to represent the dependent variable (Y-matrix). Seventy percent of the data was reserved for developing the calibration model, and 30% was used for independent validation of this model. The remaining unlabeled pixels in the image were used to test every model. The model performance is estimated by the similarity of actual and expected values/labels. To improve robust classification models, the confusion matrices were used to evaluate the precision (Pr), sensitivity (Se), specificity (Sp), and overall accuracy of the model (Acc) [28]. These evaluations are based on four diverse standard parameters like True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). They are known as equations (2-5) [29].

$$Acc = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \times 100 \quad (2)$$

$$Pr = \frac{\sum TP}{\sum TP + \sum FP} \times 100 \quad (3)$$

$$Se = \frac{\sum TP}{\sum TP + \sum FN} \times 100 \quad (4)$$

$$Sp = \frac{\sum TN}{\sum TN + \sum FP} \times 100 \quad (5)$$

a) K-NEAREST NEIGHBORS (KNN)

It is a basic yet effective classification method. The main disadvantages of KNN are (1) its low efficiency; a lazy learning mechanism prevents it from being used in many applications. (2) its dependency on the selection of a "good value" for k. The KNN model substitutes the data as the classification foundation. The value of k is computed automatically, and it varies depending on the data. It is best in terms of classification accuracy. The model's architecture decreases the model's reliance on k and speeds up categorization [30]. The algorithm succeeded in differentiating the separate classes, as seen by the confusion matrix in Figure 7a, but there was misclassification between the road and the buildings, and when the entire picture was tested, it was clear. There was also misclassification between these classes (buildings and roads). The confusion matrix, as shown in Figure 7a, depicted what happened during the algorithm when it was applied to the various classes. The relation between the true label for the desert class and the predicted label was 99.3% so there was misclassifying by 7.0% with the rock class. For the green class, the relation between the true label and the predicted label

was 99.2%. so, there was misclassifying by 0.7% with buildings and 0.1% with water. For the building class, the relation between the true label and the predicted label was 91.7%. so, there was misclassifying by 0.2 % with desert, 3.4% with green and 4.7% with the road, and 0.1% with water. For the rock class, the relation between the true label and the predicted label was 89.1%. so, there was misclassifying by 10.8% with desert. For the road class, the relation between the true label and the predicted label was 85.2%. so, there was misclassifying by 14.7% with buildings and 0.1% with deserts. For the water class, the relation between the true label and the predicted label was 90.6%. so, there was misclassifying by 4.5% with green, 3.4% with building, and 1.5% with the road. The recall of the KNN algorithm is 95% as shown in **Table 2**.

b) LINEAR DISCRIMINANT (LDA)

In face recognition, LDA is an effective discriminant analysis technique. It achieved dimensionality reduction by attempting to discover a mapping from an originally high-dimensional space to a low-dimensional space that preserves the most discriminant traits [31]. It was the algorithm that determined the optimal separation line between the various classes (six classes) to separate them. The algorithm succeeded in distinguishing the distinct classes, as shown by the confusion matrix in **Figure 7b**, but there was misclassifying between the road and the buildings. When the total picture was checked, it was evident that there was also misclassifying between these classes (buildings and roads). As shown in **Figure 7b**, the confusion matrix showed what happened during the algorithm, as applied to the different classes where the relation between the true label for the dessert class and the predicted label was 92.5%. So, there was misclassifying by 7.5% with the rock class. For the green class, the relation between the true label and the predicted label was 97.5%. So, there was misclassifying of 2.1% with buildings,0.1% with roads, and 0.4% with water. For the building class, the relation between the true label and the predicted label was 77.3%. So, there was misclassifying by 0.6 % with desert, 1.4% with green and 20.6% with the road, and 0.1% with water. For the rock class, the relation between the true label and the predicted label was 88.1%. So, there was misclassifying by 11.8% with desert. For the road class, the relation between the true label and the predicted label was 78.1%. So, there was misclassifying by 21.1% with buildings and 0.9% with deserts. For the water class, the relation between the true label and the predicted label was 87.7%. So, there was misclassifying by 4.8% with green, 6% with building, and 1.5% with the road. The recall of the LDA algorithm is 89% as shown in **Table 2**.

c) RANDOM FOREST (RF)

Breiman and Cutler created the "random forest" method in 2001. It works by training and producing the class, which is the mode of the classes generated by individual trees, using numerous decision trees. It outperforms single decision trees and is far more efficient than classic machine learning techniques, especially when dealing with enormous datasets. Thousands of explanatory variables can be handled by a random forest. When the R package "random, forest" is installed, it may be used to rank the relevance of variables. It can describe complicated interactions among variables and is useful for displaying the nonlinear effects of variables (Li et al., 2013). It was the best and most accurate in separating the various classes, as seen by the results of the confusion matrix as compared with the other algorithms, as shown in **Figure 7c** the confusion matrix showed what happened during the algorithm, as applied to the different classes where the relation between true label for dessert class and the predicted label was 99%. So, there was misclassifying by 1% with the rock class. For the green class, the relation between the true and predicted labels was 99.3%. So, there was misclassifying by 0.6% with buildings and 0.1% with water. For the building class, the relation between the true label and the predicted label was 96.2%. So, there was misclassifying by 0.5 % with green, 3.1% with, and 0.1% with water. For the rock class, the relation between the true and predicted labels was 97.2%. So, there was misclassifying by 2.8% with desert. For the road class, the relation between the true and predicted labels was 95.3%. So, there was misclassifying by 4.7% with the building. For the water class, the relation between the true label and the predicted label was 95.3%. So, there was misclassifying by 2.3% with green, 1.6% with building, and 0.8% with road. The recall of the RF algorithm is 99% as shown in **Table 2**.

The consequences of validation and calibration of the trained models were tabulated in Table 2, also the confusion matrix of these models was manifested in **Figure 7**. All models achieved high accuracy without over-fitting, whereas RF surpasses other techniques.

Table 2. Accuracy for different machine learning algorithms

ML Technique	Accuracy		F1 score	Precision	Recall
	calibration	Validation			
KNN	96.15%	95.49%	95%	96.2%	95%
LDA	89.06%	88.86%	89%	90%	89%
RF	99.98%	98.80%	99%	99%	99%

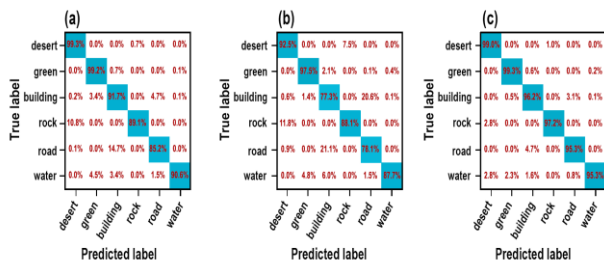


Figure 7: Confusion matrix for different models (a) K-nearest-neighbor, (b) Linear discrimination analysis, and (c) Random forest

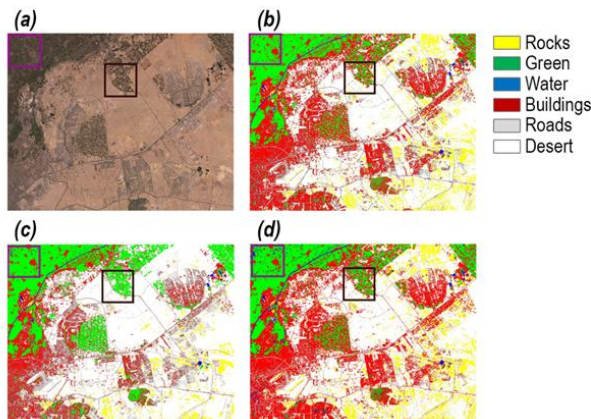


Figure 8: Comparison between three algorithms where the red box shows (a) RGB true color, (b) KNN model, (c) LDA model, and (d) RF model

The RGB image contains some buildings between green regions and these buildings are classified using KNN, Figure 8. b, and RF, Figure 8.d, but not when using LDA, Figure 8. c. The blue box shows the water class in RGB image, Figure 8. a, and water was classified by RF only, Figure 8.d.

4 APPLICATION AREA (NEW ADMINISTRATIVE CAPITAL)

Over the past few years, the New Administrative Capital has received a great deal of interest and rapid development in the field of construction and road networks in a way that has completely changed the features of the region. Therefore, the study of this region will give clear and important results to decision-makers. Therefore, the RF algorithm was relied upon in the classification of the application area because RF is a robust ML technique in classification compared with the trained model (KNN and LDA). It produced promising results with a minimal number of training datasets used for six classes as shown in Figure 9.

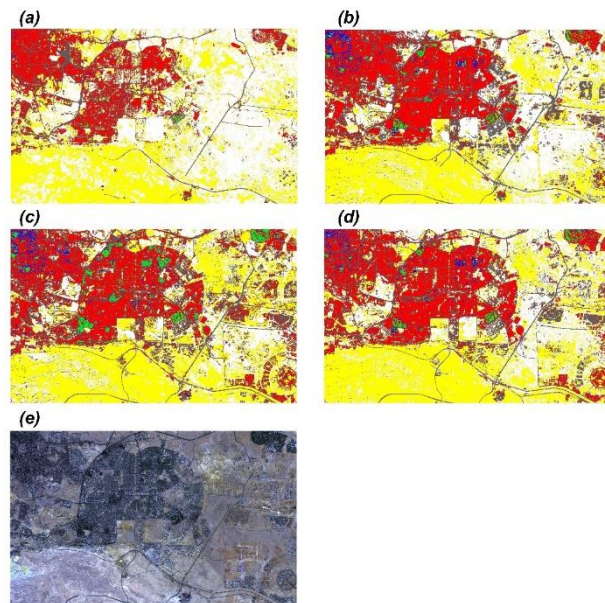


Figure 9: Applying RF model in the application area, (a) 2018, (b) 2019, (c) 2020, (d) 2021, (e) RGB image

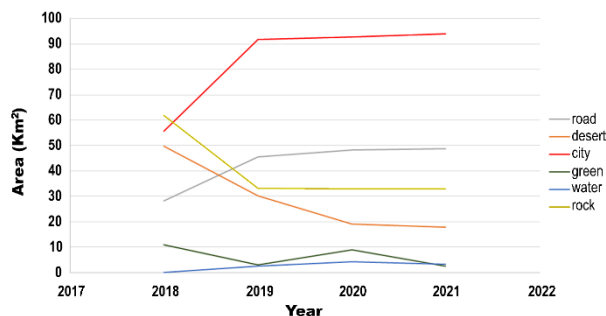


Figure 10: The progress on all six classes, This scatter chart shows the progress in all classes at New Administrative Capital

The New Administrative Capital was monitored, as shown in Figure 10, in a span of four years from 2018 to 2021. There was about a double increase in the estimated transport networks from 28.14 km² to 48.84 km². Besides, there was also a large surge until 2021 in the urban areas from 55.7 km² to 94 km². The improvement in these classes matches with the government advancement in the application area that included the UTNM for providing services to society in the future. There was a decrease in the area of the desert and the area of rocks, as well as a distinct shift in the green area, as a result of the development and pace of progress in building and transportation networks, and thus there have been changes throughout the years during the application area.

5 THE EFFECT OF URBAN DEVELOPMENT ON THE ENVIRONMENT

5.1 Carbon monoxide (CO)

Generally, the development of urbanization can affect the environment positively or negatively. Therefore, the impact of the CO in the application area was monitored for a period of 5 years from 2018 to 2022. As could be seen from the graph, the CO levels increased from 2018 to 2019 when it was 0.03 m^{oll}/m², and then they became 0.032 m^{oll}/m². In 2020, because of the Covid19 lockdown, all the construction projects in the application area were terminated; thus, the environment started to regain its strength. As displayed in **Figure 11**, chart f, the amount of CO dropped to 0.0315 m^{oll}/m². By 2021, the negative impact of the virus started to decline; hence, construction projects commenced to increase. Therefore, there was a gradual increase in the gas till it scored 0.0335 m^{oll}/m² in 2022.

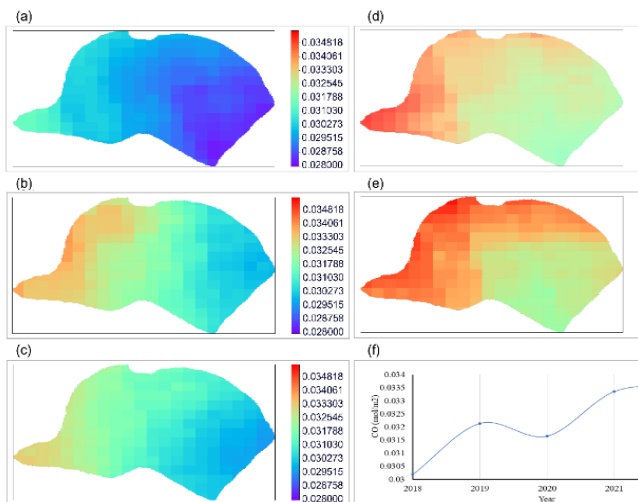


Figure 11: (a) Mapping of CO levels in the New Administrative Capital in 2018, (b) 2019, (c) 2020, (d) 2021, (e) 2022, (f) Accumulative change of CO concentration from 2018 to 2022

5.2 Nitrogen Dioxide (NO₂)

Commonly, the advancement of urbanization can affect the environment positively or negatively; therefore, the impact of the NO₂ in the application area was monitored for a period of 5 years from 2018 to 2022. As could be seen from the graph, the NO₂ witnessed a significant increase from 2018 to 2019 when it was 0.00016 m^{oll}/m² and then became 0.00018 m^{oll}/m². In 2020, because of the epidemic of Coronavirus, all the construction projects in the application area were terminated; thus, the environment started to regain its power. As displayed in **Figure 12**, chart F, the amount of NO₂ dropped to 0.0014 m^{oll}/m². By 2021, the negative impact of the virus started to decline; hence, construction

projects commenced to increase. Therefore, there was a sudden rise in gas till it scored 0.0003 m^{oll}/m² in 2022.

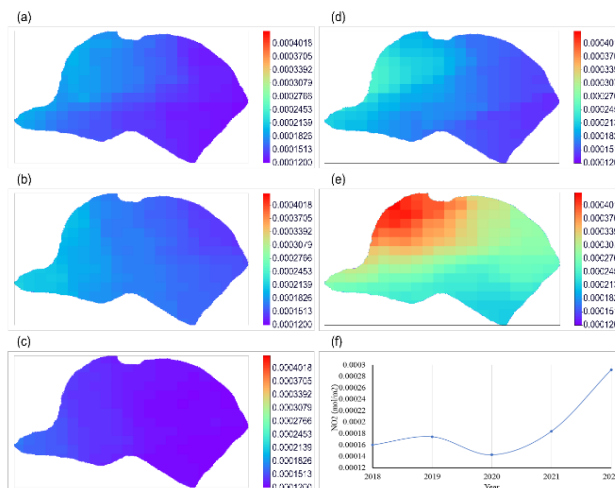


Figure 12: (a) Mapping of NO₂ levels in the New Administrative Capital in 2018, (b) 2019, (c) 2020, (d) 2021, (e) 2022, (F) Accumulative change of NO₂ concentration from 2018 to 2022

6 RESULTS AND DISCUSSION

By studying the classification accuracy of several models covering simple histogram thresholding, spectral indices, and machine learning techniques, it was found that Random Forest produces the best performance for land use classification and urban segmentation in our study regions. The RF model utilizes Sentinel-2 images and produces a classification accuracy of 98.80%. This result demonstrates the effectiveness of the RF model in land use classification using the medium resolution Sentinel-2 images, which offers the opportunity for continuous monitoring of urbanized regions to assess the development rate with minimal cost and high temporal resolution.

The RF model was used to monitor the New Administrative Capital in Egypt from the year 2018 to 2021 which shows a rapid and significant change in the land use classification, especially in the urban and road classes which reflects the high development rate of the region. The monitoring results show that the area of urban regions and roads nearly doubled from the year 2018 to the year 2019, after that the area of these classes shows a small increase from the year 2019 to the year 2021. This can be attributed to two main reasons: first, the temporary suspension of construction projects because of the Covid19 lockdown. Second: the first construction stage of the New Administrative Capital project was nearly finished by the end of the year 2019 and did not start the second stage until the end of the year 2021. During the same time frame, the desert and rock classes areas were reduced by nearly the same amount of the increase in the urban and road classes. These results demonstrate the high-level urbanization

and development rate of the New Administrative Capital during the past few years.

For the environmental assessment of the New Administrative Capital, we utilized Sentinel 5P to monitor the concentrations of CO and NO₂ over the study region from the year 2018 to the year 2022. The monitoring results show a significant increase in these pollutant levels which is attributed to the ongoing construction operation in the region. Also, it was found a temporary drop in the CO and NO₂ concentration levels in the year 2020 which is attributed to the Covid19 lockdown and suspension of construction projects.

7 CONCLUSIONS

In this paper, several classification techniques were tested utilizing medium-resolution Sentinel-2 images. It was found that the RF technique produces the best performance with a classification accuracy of 98.80%, which demonstrates the capability of utilizing freely available satellite data in continuous monitoring of urbanized regions with high temporal frequency. By using the produced RF model to monitor the urbanization rate of the New Administrative Capital in Egypt from the year 2018 to the year 2021, it was found that the areas of the urban region have increased from 55.7 km² to 94 km² and transportation networks areas have increased from 28.14 km² to 48.84 km². This demonstrates the high urbanization and development rate in this region which was found to be associated with a high increase in air pollutants. This study shows a general image of the CO and NO₂ pollution dynamics across the New Administrative Capital, based on Copernicus Sentinel-5P data. From the year 2018 to the year 2022 the CO concentration levels increased from 0.0315 mol/m² to 0.0335 mol/m², while the NO₂ concentration levels increased from 0.00016 mol/m² to 0.0003 mol/m². The monitoring results show a drop in the air pollutants level in the year 2020 during the Covid19 lockdown which proves the association between the construction process and the increase in air pollution levels.

ACKNOWLEDGMENTS

The authors acknowledge the Faculty of Engineering, Suez Canal University for providing the necessary help and research facility to conduct the current study. The authors also would like to express gratitude to Dr. Ahmed Islam ElManawy, Agricultural Engineering Department, Suez Canal University for his help in analyzing the data.

RESEARCH AND PUBLICATION ETHICS

Author Contributions

Each author is expected to have made considerable contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the

creation of new software used in the work; or have drafted the work or substantively revised it; AND has approved the submitted version (and subsequent versions edited by journal staff; AND agrees to be personally accountable for the author's contributions and for ensuring that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved.

REFERENCES

- [1] T. ElGharbawi and F. Zarzoura, "Damage detection using SAR coherence statistical analysis, application to Beirut, Lebanon," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, no. September 2020, pp. 1–9, 2021, doi: 10.1016/j.isprsjprs.2021.01.001.
- [2] D. Yacouba, H. Guangdao, and W. Xingping, "Applications of Remote Sensing in Land Use/Land Cover Change Detection in Puer and Simao Counties, Yunnan Province," *Rep. Opin.*, vol. 2, no. 9, pp. 7–16, 2010.
- [3] S. Ishaya and O. O. Ifatimehin, "Application of remote sensing and GIS techniques in mapping fadama farming areas in a part of Abuja, Nigeria," *Am. J. Sustain. Agric.*, vol. 3, no. 1, pp. 37–44, 2009.
- [4] K. Minho, J. Minyoung, and K. Yongil, "Histogram Matching of Sentinel-2 Spectral Information to Enhance PlanetScope Imagery for Effective Wildfire Damage Assessment," *Korean J. Remote Sens.*, vol. 35, no. 4, pp. 517–534, 2019.
- [5] V. Fernández-García, M. Santamarta, A. Fernández-Manso, C. Quintano, E. Marcos, and L. Calvo, "Burn severity metrics in fire-prone pine ecosystems along a climatic gradient using Landsat imagery," *Remote Sens. Environ.*, vol. 206, no. February 2017, pp. 205–217, 2018, doi: 10.1016/j.rse.2017.12.029.
- [6] M. E. D. Chaves, M. C. A. Picoli, and I. D. Sanches, "Recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping: A systematic review," *Remote Sens.*, vol. 12, no. 18, 2020, doi: 10.3390/rs12183062.
- [7] P. Addabbo, M. Focareta, S. Marcuccio, C. Votto, and S. L. Ullo, "Contribution of Sentinel-2 data for applications in vegetation monitoring," *Acta IMEKO*, vol. 5, no. 2, pp. 44–54, 2016, doi: 10.21014/acta_imeko.v5i2.352.
- [8] R. Meng *et al.*, "Using high spatial resolution satellite imagery to map forest burn severity across spatial scales in a Pine Barrens ecosystem," *Remote Sens. Environ.*, vol. 191, no. 631, pp. 95–109, 2017, doi: 10.1016/j.rse.2017.01.016.
- [9] E. Dragozi, I. Z. Gitas, S. Bajocco, and D. G. Stavrakoudis, "Exploring the relationship

- between burn severity field data and very high resolution GeoEye images: The Case of the 2011 Evros wildfire in Greece,” *Remote Sens.*, vol. 8, no. 7, 2016, doi: 10.3390/rs8070566.
- [10] A. Hamedianfar and H. Z. M. Shafri, “Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images,” *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3380–3396, 2015, doi: 10.1080/01431161.2015.1060645.
- [11] M. Drusch *et al.*, “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services,” *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012, doi: 10.1016/j.rse.2011.11.026.
- [12] A. Saber, I. El-Sayed, M. Rabah, and M. Selim, “Evaluating change detection techniques using remote sensing data: Case study New Administrative Capital Egypt,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 24, no. 3, pp. 635–648, 2021, doi: 10.1016/j.ejrs.2021.03.001.
- [13] A. El-Zeiny, H. Effat, K. Mansour, A. Shahin, and K. Elwan, “Geo-environmental monitoring of coastal and land resources of Port Said Governorate, Egypt,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 25, no. 1, pp. 157–172, 2022, doi: 10.1016/j.ejrs.2022.01.009.
- [14] S. Gakhar and K. Chandra Tiwari, “Spectral – spatial urban target detection for hyperspectral remote sensing data using artificial neural network,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 24, no. 2, pp. 173–180, 2021, doi: 10.1016/j.ejrs.2021.01.002.
- [15] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, “Semi automatic road extraction from digital images,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 20, no. 1, pp. 117–123, 2017, doi: 10.1016/j.ejrs.2017.03.001.
- [16] K. Shahi, H. Z. M. Shafri, E. Taherzadeh, S. Mansor, and R. Muniandy, “A novel spectral index to automatically extract road networks from WorldView-2 satellite imagery,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 18, no. 1, pp. 27–33, 2015, doi: 10.1016/j.ejrs.2014.12.003.
- [17] 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, May 2018, doi: 10.1080/01431161.2018.1433343/SUPPL_FILE/TRES_A_1433343_SM5998.ZIP.
- [18] “The New Administrative Capital <https://www.nytimes.com/2022/10/08/world/middleeast/egypt-new-administrative>
- [19] “Home - 10th Ramadan.” http://www.newcities.gov.eg/know_cities/Tenth_Ramadan/default.aspx
- [20] L. Shikwambana, P. Mhangara, and N. Mbatha, “Trend analysis and first time observations of sulphur dioxide and nitrogen dioxide in South Africa using TROPOMI/Sentinel-5 P data,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 91, no. December 2019, p. 102130, 2020, doi: 10.1016/j.jag.2020.102130.
- [21] M. Hereher, R. Eissa, A. Alqasemi, and A. M. El Kenawy, “Assessment of air pollution at Greater Cairo in relation to the spatial variability of surface urban heat island,” *Environ. Sci. Pollut. Res.*, vol. 29, no. 15, pp. 21412–21425, 2022, doi: 10.1007/s11356-021-17383-9.
- [22] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” *Remote Sens. Environ.*, vol. 202, no. 2016, pp. 18–27, 2017, doi: 10.1016/j.rse.2017.06.031.
- [23] S. Van Der Walt *et al.*, “Scikit-image: Image processing in python,” *PeerJ*, vol. 2014, no. 1, pp. 1–18, 2014, doi: 10.7717/peerj.453.
- [24] C. M. Lewandowski, N. Co-investigator, and C. M. Lewandowski, *Learning OpenCV 3 Computer Vision with Python*, vol. 1. 2015.
- [25] E. Duchesnay *et al.*, “Statistics and Machine Learning in Python To cite this version : HAL Id : hal-03038776 Statistics and Machine Learning in Python,” 2020.
- [26] D. Lu, P. Mausel, E. Brondízio, and E. Moran, “Change detection techniques,” *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004, doi: 10.1080/0143116031000139863.
- [27] O. J. Tobias and R. Seara, “Image segmentation by histogram thresholding using fuzzy sets,” *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1457–1465, 2002, doi: 10.1109/TIP.2002.806231.
- [28] S. V. Stehman, “Selecting and interpreting measures of thematic classification accuracy,” *Remote Sens. Environ.*, vol. 62, no. 1, pp. 77–89, 1997, doi: 10.1016/S0034-4257(97)00083-7.
- [29] D. M. W. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” pp. 37–63, 2020.
- [30] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN model-based approach in classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, no. June 2015, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.
- [31] F. Ye, Z. Shi, and Z. Shi, “A comparative study of PCA, LDA and kernel LDA for image classification,” *Proc. - 2009 Int. Symp. Ubiquitous Virtual Reality, ISUVR 2009*, pp. 51–54, 2009, doi: 10.1109/ISUVR.2009.26.