


Adoption of Machine Learning Technique in Nile River Islands Classification

Noha Kamal ^{a,*}, Ahmad A. Aziz El-Banna^b and Nagwa El-Ashmaw^c

^a Nile Research Institute, National Water Research Center, Cairo, Egypt

^b Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt

^c Drainage Research Institute, National Water Research Center, Cairo, Egypt

*Corresponding author. E-mail: noha_kamal2002@hotmail.com, noha_kamal@nwrc.gov.eg

 NK, 0000-0001-5209-5505

ABSTRACT

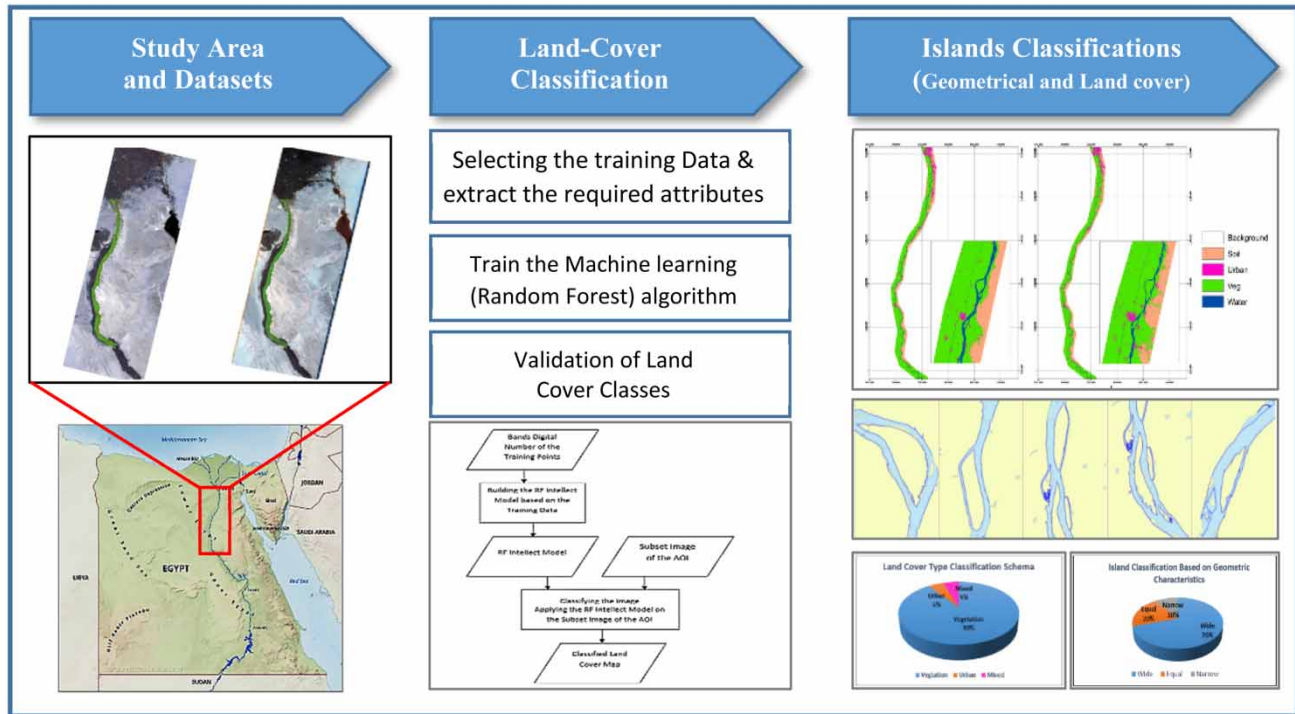
Wider adoption of machine learning methods in water resources has the potential to greatly accelerate the efficiency and quality of analysis. The Nile River is one of the major fluvial hydro-systems in the world. Fluvial islands are present in nearly all natural and regulated rivers. The Nile River is characterized by numerous natural phenomena and human interventions represented in multiple islands characteristics. This paper investigates the formation and development of the Nile River islands in the fourth reach, which extends between Assuit and Delta barrages. A machine learning (ML) technique, with the Random Forest (RF) algorithm, has been introduced as a potential technique to replace the traditional ones, to extract and classify the land cover types and the geometrical characteristics of the Nile River islands. The assessment of the results of extracting the Nile River islands and the land cover types are included. The accuracy of the extracted boundaries of the islands is assessed using field surveying data. The classification of the islands based on the islands' geometric characteristics represented that 70% of the extracted islands are classified as Wide Island, 20% are classified as Equal Island, and 10% as Narrow Island. The islands' classification, based on the land cover, results show that there is only 5% of the islands that are urban areas, 5% of the islands are mixed class (both vegetation and urban), and the rest of the islands 90% have a vegetation land cover type. The accuracy assessment was performed using the error matrix, the results show that the overall accuracy of the land cover classification is greater than 84%. The proposed islands' classification scheme can become an important tool that provides the decision-makers with more detailed information to improve the planning of the Nile River islands development projects. Furthermore, this schema can be expanded to other climatic and topographic regions.

Key words: fluvial islands, machine Learning (ML), Nile River, Random Forest (RF), satellite images classification

HIGHLIGHTS

- The study investigates the applicability of using the Machine Learning technique in Nile island boundaries determination out of the satellite images.
- The study utilizes the Random Forest algorithm to extract the land cover types for the Nile islands and the surrounding area.
- The study introduces new schemes for Nile island classification based on geometric and land cover aspects.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The Nile River is characterized by numerous natural phenomena and human interventions represented in multiple qualities of the islands' characteristics and the spread of many activities along the stream. Damming changes hydrological regimes quickly and alters aquatic environments leading to sediment transport and islands formation in regulated rivers (Mahmoud *et al.* 2021). Since, in Egypt, the Nile water flows into the High Aswan Dam (HAD), there is a necessity to conduct a hydrological assessment and island classification for better river management strategies and development plans. Fluvial islands are important in both hydrologic and biotic capacities and can therefore be indicators of the general health and energy of the system (Breiman 2001). Islands associated with the river can provide a detailed account of that river's past and present activities. Because islands separate the total river flow into at least two individual channels, they create varying hydraulic conditions due to different widths, depths, slopes, etc. (Thanh Noi & Kappas 2017). Because islands and rivers are so fundamentally linked, any river restoration and management strategies must incorporate islands as well. The islands' classification will improve the understanding of the range of possible river responses. The Nile River has many islands that have been formed after the HAD construction. Researchers are interested to study the development and evolution of these islands to reflect the esthetic aspects and improvement of the environment surrounding the islands since they have an economic value, and they can be used in agriculture and recreation.

The construction of the HAD has caused major changes in hydrological characteristics downstream the dam, consequently affecting the river morphology including island formation and types (Sadek *et al.* 1999). Sadek *et al.* (NRI 1989) studied and evaluated the effect of the HAD on the morphology of Rosetta Branch. Based on the analysis of aerial photos, three types of islands along the branch have been reported. The first was permanent islands, which have permanent vegetation and distinct from sand bars (NRI 1989), and that stayed unchanged without being attached to the banks. The second type of islands was the attachment islands which joined the main banks and worked as a part of the flood plain. This process has increased the land areas neighboring Rosetta Branch. The under-forming islands were considered the third type that has begun to form due to the small amount of water flowing and the heavy sedimentation resulted from the weak water current (Sadek 2012).

Other types for island classification are required, for instance, the classification of islands based on the relation between the dimensions of the islands to the dimensions of the river is required for studying the hydraulic condition of the river, and the

applicability of using the river for navigation; this type of classification has not been considered for the Nile River islands yet. Additionally, the classification of the islands based on land cover types is important in determining the water consumption for each area, as well as, this information will support the decision-makers in assessing the value of the land if required for development studies.

Remote sensing with the advanced classification techniques is a powerful tool for extracting the water bodies (and the islands), as well as in the land cover classification. The machine learning (ML) classification technique has gained much interest recently in various fields of engineering sciences including the remote sensing studies. This extensive application of various ML techniques has taken place due to their outstanding capabilities in finding proper solutions for plenty of problems with lower implementation complexities compared to complex mathematical models. In addition, ML techniques can enhance the performance by finding solutions faster than the heuristic models (Nguyen *et al.* 2018).

Classification can be defined as the process of identifying and grouping objects into predetermined categories, and by using pre-categorized training datasets, ML classification techniques/algorithms can classify future datasets into similar categories to the trained ones. Classification problems can be solved using plenty of ML algorithms, such as Logistic Regression, k-Nearest Neighbors, Decision Trees (DTs), Random Forest (RF), Support Vector Machine, Naive Bayes, and even Neural Networks. Each algorithm has its pros and cons; however, the basic principle of the RF algorithm is the DTs algorithm. DT techniques are multilateral ML algorithms that can perform classification, regression, as well as multi-output tasks. DTs are powerful and fast algorithms, capable of fitting complex datasets (Mutanga *et al.* 2012). DT models employ two phases, i.e., the training and prediction phases (the same case as all supervised techniques). During the training phase, features and labels samples are utilized to compose a tree model. The tree classification model is a decision support tool that is constructed from decision nodes, branches, and leaf nodes (Rodriguez-Galiano *et al.* 2012). Each decision node is considered a test on a certain attribute, while each branch acts as the outcome of the test, and each leaf node represents a class label that indicates the decision taken after computing all attributes. DT classifiers are simple in visualization and interpreting and require small data preprocessing. However, the prediction time of DT is cost-effective and depends on the data attributes. DTs are also the fundamental components of the RF, which are among the most powerful ML algorithms available today (Rodriguez-Galiano *et al.* 2012).

RF is an ensemble of DT, generally trained via the bagging method, i.e., bootstrap aggregation (Mutanga *et al.* 2012). Typically, ensemble methods combine some similar or different algorithms to classify objects based on the likelihood that a combination of learning models could increase the overall selected result. In other words, the RF algorithm builds multiple DTs with training data, then fits the new data within one of the trees as a 'random forest'. The main steps for implementing the RF algorithm can be summarized as (Rodriguez-Galiano *et al.* 2012): (1) select random K data samples (subsets) from the training set; (2) build the DT for each selected subset; (3) compute the prediction result from every DT; (4) perform voting for every predicted result; and (5) nominate the most voted prediction result as the final prediction result. A simple description of the RF technique is shown in Figure 1.

The selection of the classifier significantly influences the results of the satellite image classification. Using ML technique in satellite image classification has intensely increased in the last decade. Several ML algorithms have proved their superiority over the traditional classification techniques (Zhang *et al.* 2021). The ML algorithms are robust and more accurate than traditional classifiers, especially when dealing with large amounts of data and data noise (Jamali 2019). One of the most popular ML algorithms is RF, which has been widely applied in the remote sensing field as a classification algorithm (Beechie *et al.* 2006). RF classifier was suggested by Breiman in 2001, this classifier is a supervised ensemble classification algorithm based on the DTs (Thorp 1992). This ensemble learning method is developed to improve the classification and regression trees (CART) method by combining a large set of the DT. Additionally, the RF overcomes the problem of overfitting to the training set in the CART method. The RF is a non-parametric statistical technique, which has the capability to produce classification functions based on continuous or discrete datasets and can deal with complex relationships between predictors due to the large amounts of data (Beechie *et al.* 2006). The RF is operated by constructing a multitude of DT at training time, based on the non-geometry attribute data of the training sample to train the algorithm. In contradiction to the single tree in the CART, the RF generates multiple trees with a different selection of training samples. Each tree in the forest is grown from training pixels which are randomly selected to train RF classification and give a vote, the pixel, then, is assigned to the class that gets the most votes. Some of the advantages of using the RF in remote sensing are that it runs efficiently on large data sets, it can handle thousands of input variables without variable deletion, it gives estimates of what variables are important in the classification, and it is computationally lighter than other tree ensemble methods (Thorp 1992;

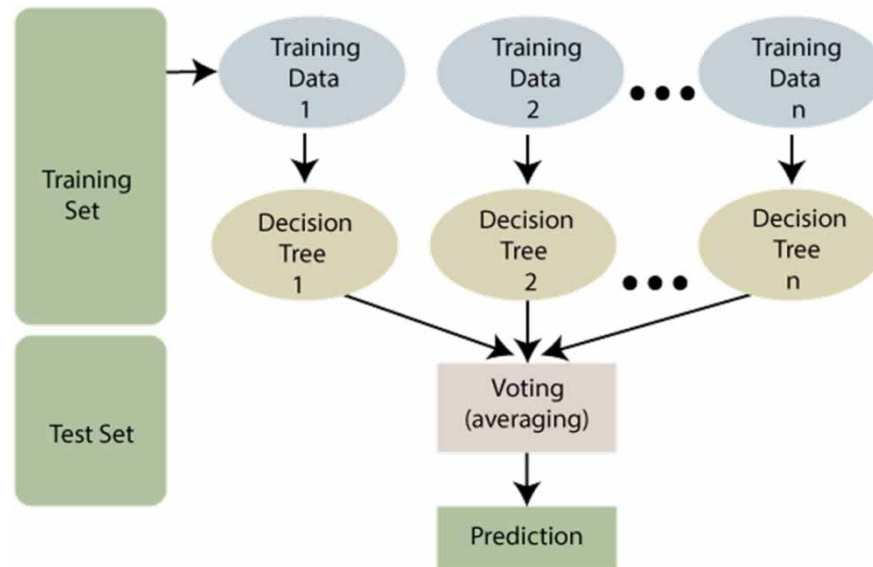


Figure 1 | Methodology of the RF algorithm (Mutanga *et al.* 2012).

Wyrick & Klingeman 2011). Furthermore, the results of the land cover classification using the RF algorithm are more accurate compared to other classifiers (Talukdar *et al.* 2020).

According to the lack of information about the Nile island's classes, the main purpose of this study is to introduce classification schemes to classify the Nile islands that have not been conducted to the Nile islands in Egypt. The two introduces classification schemes for the Nile islands are based on (1) the geometry of the islands and (2) the land cover types of the islands. Investigating the suitability of using the ML technique as a modern technique for extracting the islands, and for land cover classification, with the RF algorithm, is another aim of this research work.

2. MATERIALS AND METHODS

2.1. Study area and data sets

The selected study area is the large fluvial islands within the fourth reach of the Nile River in Egypt. The fourth reach is located between Assiut Barrage (31°11'14.76"E, 27°12'20.35"N) and Delta Barrage (31°8'20.70"E, 30°10'25.78"N). The study area is around 300 km along the Nile River. Figure 2 shows the location of the selected study area as appearing on Google Earth.

To achieve the purpose of this paper, two sets of Landsat images were downloaded from the United States Geology Survey (USGS) site for the study areas on two different dates, July 2003 and April 2021. There are available data of the island outlines that were collected in the field by a team of the Nile Research Institute (NRI) in 2003. Therefore, the Year 2003 has been selected as a date of satellite images capturing to use the available data for assessing the obtained results of the classification technique used in island extraction. However, the Year 2021 has been selected to represent the current situation. The downloaded data are illustrated in Table 1.

2.2. Research methodology

The ML technique with the RF algorithm has been used, in this research, for extracting the fluvial islands and for land cover classification of these islands. The work procedure starts by downloading the required images for the study area, mosaic the images, and then cropping the area of interest (AOI). The distinguished land cover classes should be identified before conducting the classification process. To apply the ML algorithm, training sample data should be selected first and used to produce the RF model. After producing the intellect model, this model will be applied to the entire image to assign the pixels to the appropriate class.

The proposed island classification scheme based on geometry classification will be applied. Where, the geometric characteristics of islands represent features that can be objectively determined, including the relative width (R) of the island which is the ratio of the island's maximum width to the combined width of the flanking flow channels as shown in Figure 3. Islands

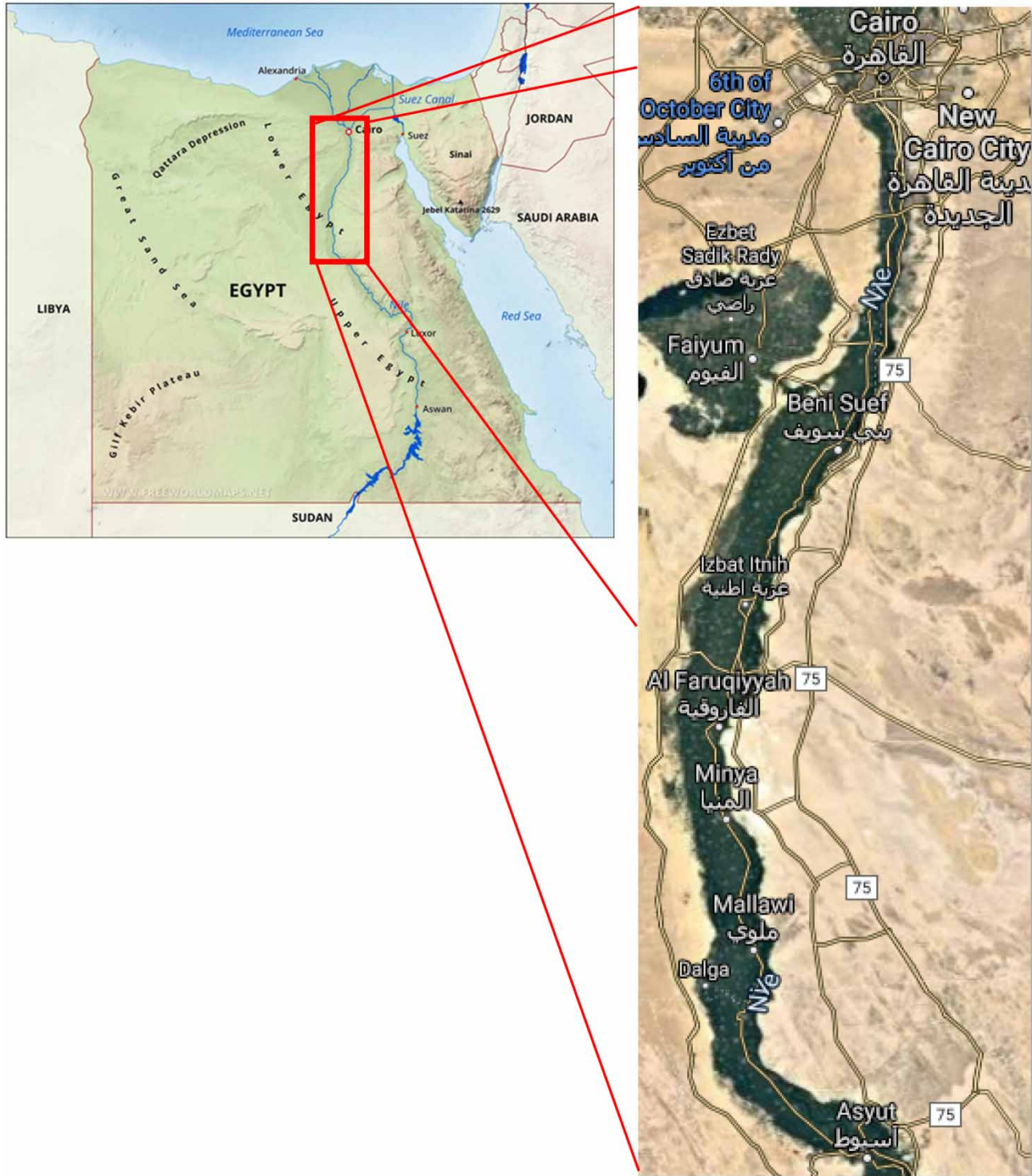


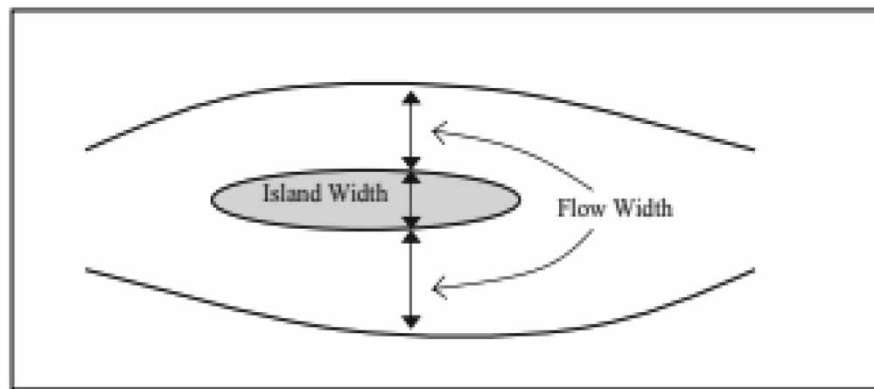
Figure 2 | Selected study area (fourth reach of the Nile River).

can be distinguished as (a) wide, having a ratio greater than 1.5, (b) equal, having a ratio between 0.5 and 1.5, and (c) narrow, having a ratio less than 0.5 (Wyrick & Klingeman 2011).

The second schema is based on the classification of land cover types of islands based on the percentage of the land cover types. The proposed scheme for the classification is that if the land cover type percent is larger than 66%, the island will be

Table 1 | Downloaded Landsat images for the study area

Sensor	Date	Path	Row
Landsat 5 TM	07/07/2003	176	039
Landsat 5 TM	07/07/2003	176	040
Landsat 5 TM	07/07/2003	176	041
Landsat 8 OLI	03/04/2021	176	039
Landsat 8 OLI	03/04/2021	176	040
Landsat 8 OLI	03/04/2021	176	041

**Figure 3** | Ratio of island width to flow width to determine the relative width of island for use in the classification of the island type.

considered covered with this land cover type, if there are two types whose percent is larger than 33%, it will be considered as mixed (of these two types).

The results of the RF classifier should be evaluated before continuing the process of island classification. The geometric characteristics, the *R* ratio, will be calculated for each island. Moreover, the area of each land cover type within each island will be determined and used for calculating the percentage of each land cover type within the islands. The final step is using the calculated geometric characteristics and the percentages of the land cover types to conduct the island classification based on the geometric characteristics and the land cover types. Figure 4 illustrates the proposed flow of the work in this research paper.

3. RESEARCH RESULTS

By following the described work procedure, the obtained results are as follows:

- The required satellite images should be uploaded to the computer, if they are not available, they can be downloaded from the internet. In this work, the Landsat images (as shown in Table 1) have been downloaded from the USGS website, as shown in Figure 5. Since the study area is covered by three different Landsat satellite images, the images are mosaiced and the AOI is cropped. The AOI is the area surrounding the fourth reach of the Nile River stream in Egypt. Cropping the AOI is done to avoid spending much time in the classification of unnecessary areas, Figure 6 shows the mosaic images and the cropped AOI in both dates.
- Four distinguished classes have been identified in the study area, namely soil (for the plain soil areas and undeveloped lands), urban (for the built-up areas), vegetation (for the cultivated areas), and water (for the Nile and any other water bodies within the study area). Training data have been formed as groups of points within each of the distinguished classes. The whole mosaiced image has been used for collecting the training data. Several locations for each class have been

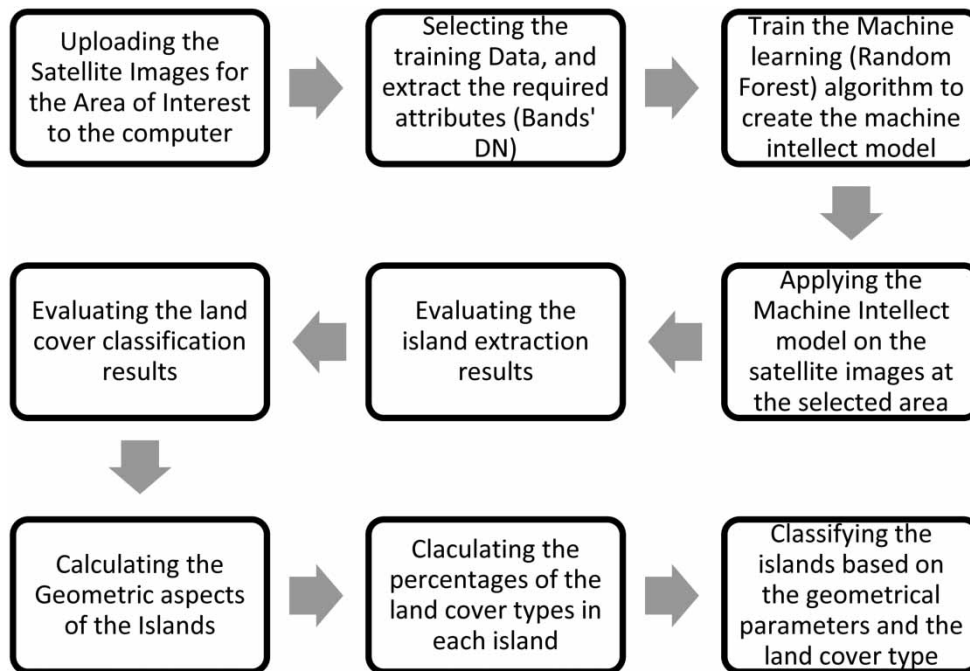


Figure 4 | Work procedure of fluvial island classification.

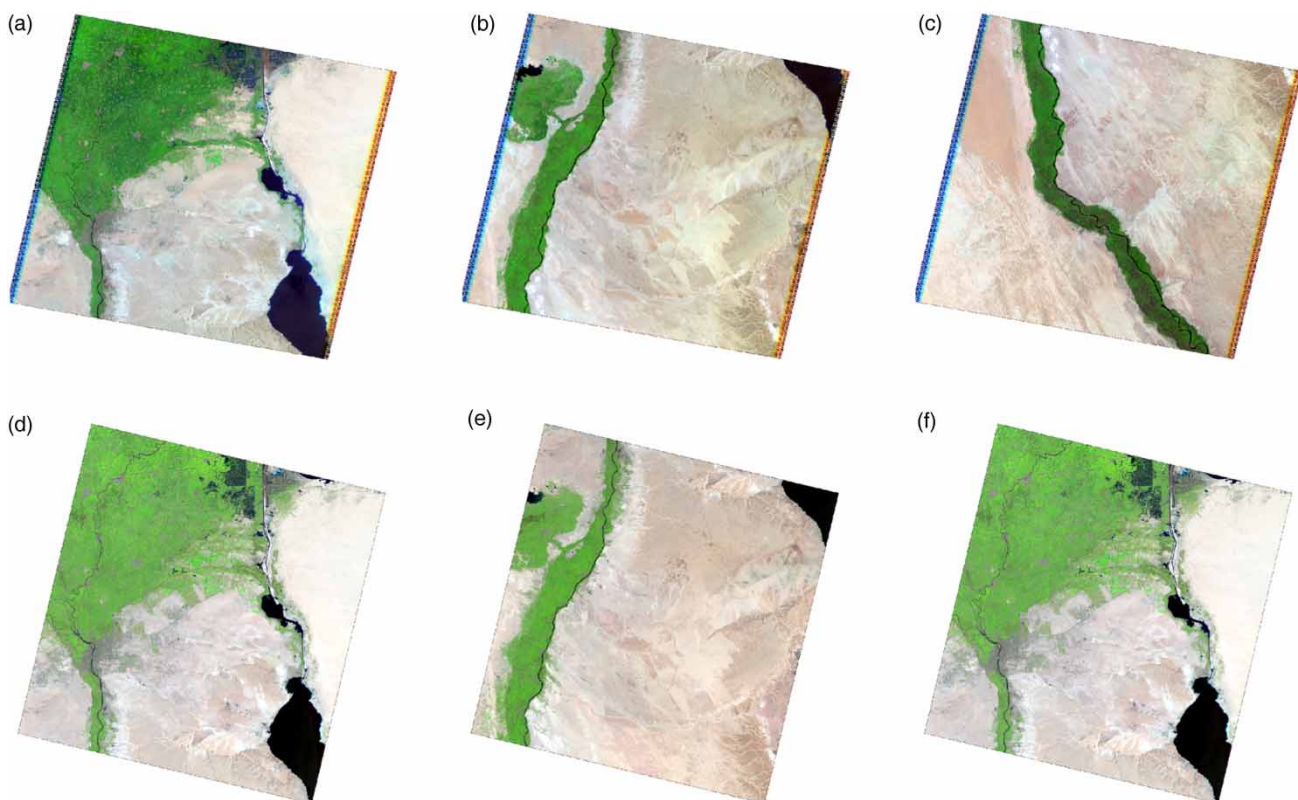


Figure 5 | Downloaded Landsat images: (a–c) Landsat 5 TM for 2003 and (d–f) Landsat 8 OLI for 2021.

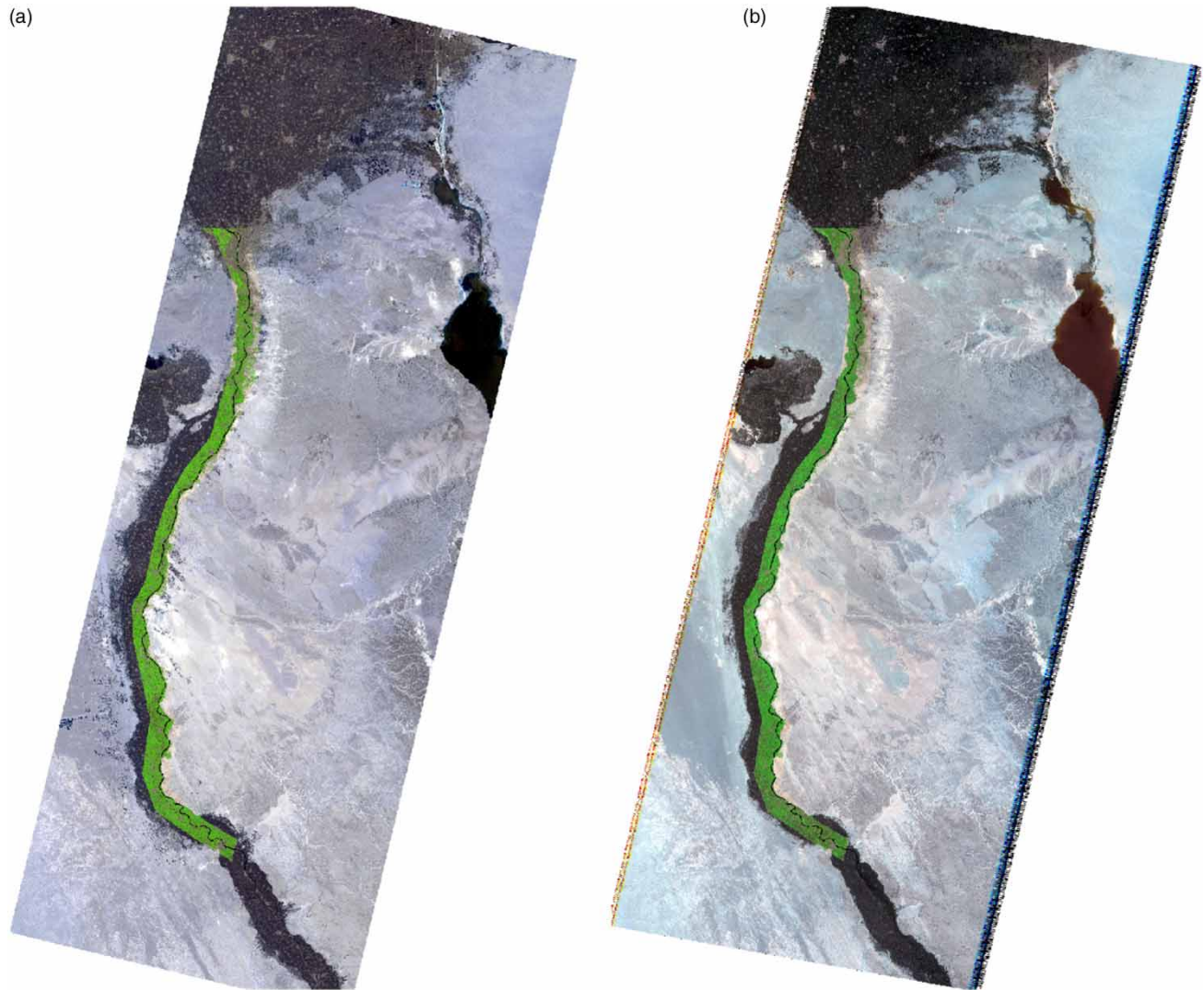


Figure 6 | Mosaiced images and the AOI: (a) Landsat 5 TM of 2003 and (b) for Landsat 8 OLI of 2021.

selected randomly to collect the training point data. The total number of training points is 30,222 points forming 116 areas. The locations of the training data are illustrated in Figure 7. For each point of the training samples, the attribute data (the Digital Number (DN) of each band) have been extracted for each of the Year 2003 and Year 2021 images.

- The attribute values of the training points have been, then, used to train the RF algorithm, and an intellect model has been produced for each year based on the attribute data determined in the previous step. Figure 8 illustrates the flow chart of the RF classifier.
- Based on the output information of the training procedure, the intellect model has been applied to the AOI, as shown in Figure 8. The output of the classification process is land cover classified maps for the study area with the following four classes: soil, urban, vegetation, and water, in the Years 2003 and 2021, as shown in Figure 9.
- The non-water classes (i.e., soil, urban, and vegetation) are gathered as Land class, then the Land class has been converted to polygons. The boundaries of these polygons within the Nile stream are extracted for further processes. Before going on in the proposed workflow, the extracted islands' boundaries have been compared to the field survey data collected in 2003. Figure 10 illustrates a sample of the extracted islands.
- For evaluating the correctness of the islands' boundaries, the 2003 extracted boundaries by the ML technique were compared to field collected data (field survey data conducted in 2003). Transects with 250 m distance apart were drawn

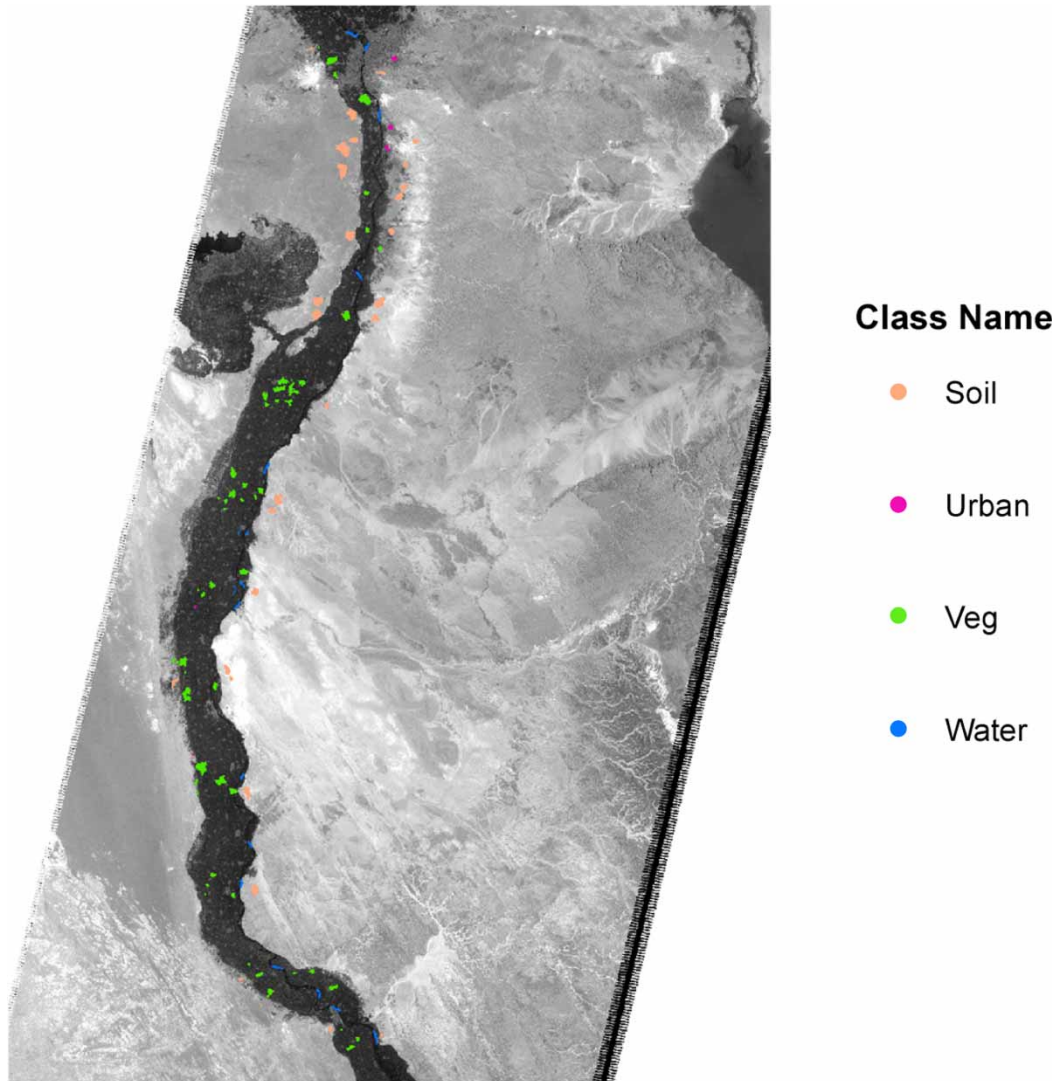


Figure 7 | Location and class of the training data.

along the island's boundary and perpendicular to the boundary line. The difference between the two data sets of the boundaries (extracted by ML and the actual boundaries by field survey) were measured. By analyzing the differences between the two boundaries datasets, it was found that the differences between them (errors in the ML classification results) have an average of 11.7 m. This value is less than half pixel size of the Landsat images (30 m spatial resolution).

- The land cover classification results have been assessed using the confusion matrix method. Around 550 groups of points have been selected randomly over the entire classified image, each group covers an area within the same class, where the total number of check points is 196,422 points. These points are used for checking the accuracy of the land cover classification produced by the ML technique with the RF algorithm. The ground truth (reference data) has been collected based on visual interpretation from the satellite image. The confusion matrix and the accuracy calculations are illustrated in [Figure 11](#). The producer, user, and overall accuracies and Kappa statistics are calculated for the accuracy assessment.
- After classifying the images, the large islands, which have areas larger than 50 fed., within the fourth reach have been extracted to be classified. Twenty islands have been extracted out of the Year 2021 classified image to represent the current situation. The widths of the islands are measured, as well as the width of the streams to the right and left of the islands. [Table 2](#) shows the measured data of the islands and the calculation of the *R* ratio of the islands.

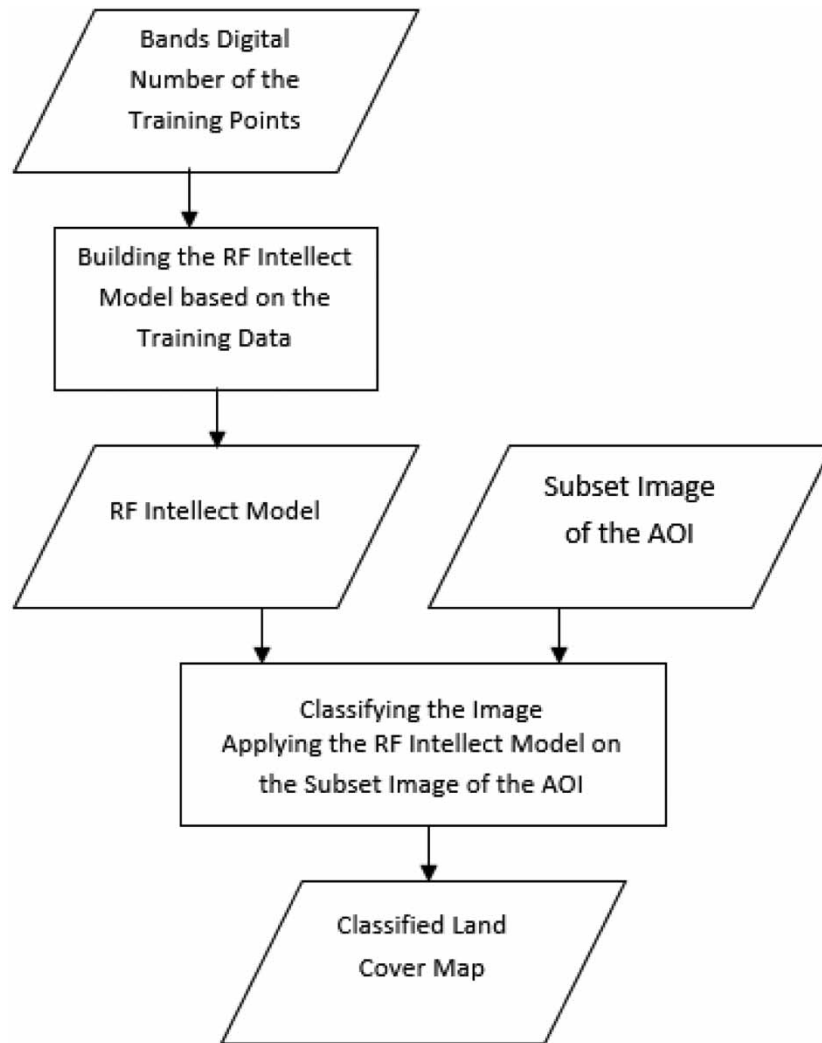


Figure 8 | Steps of the RF classifier.

- For the land cover classification, the land cover classes of each island have been extracted and the area of each class in each island has been measured. After that, the percentages of the urban and vegetation classes in each island have been calculated. Table 3 illustrates the percentage of the land cover classes in each island.
- The last step of the proposed work procedure is the classification of the islands based on the calculated relative width ratio and the land cover types. Based on the classification criteria mentioned in the Introduction section, the islands have been classified as wide, equal, or narrow islands, and urban, vegetation, or mixed islands.

4. DISCUSSION

From the illustrated results of the proposed island classification workflow, and the results of using ML for extracting the islands and for land cover classification of the islands, it can be noticed that.

4.1. Satellite image classification and results evaluation

There are two different types of images for the selected study area, each on a different date, namely Landsat TM images for 2003 and Landsat OLI for 2021 images. Therefore, the training data are not the same for the two sets of images, that is because the Landsat TM images have seven bands (six of them have been used) and the Landsat LOI images have 11

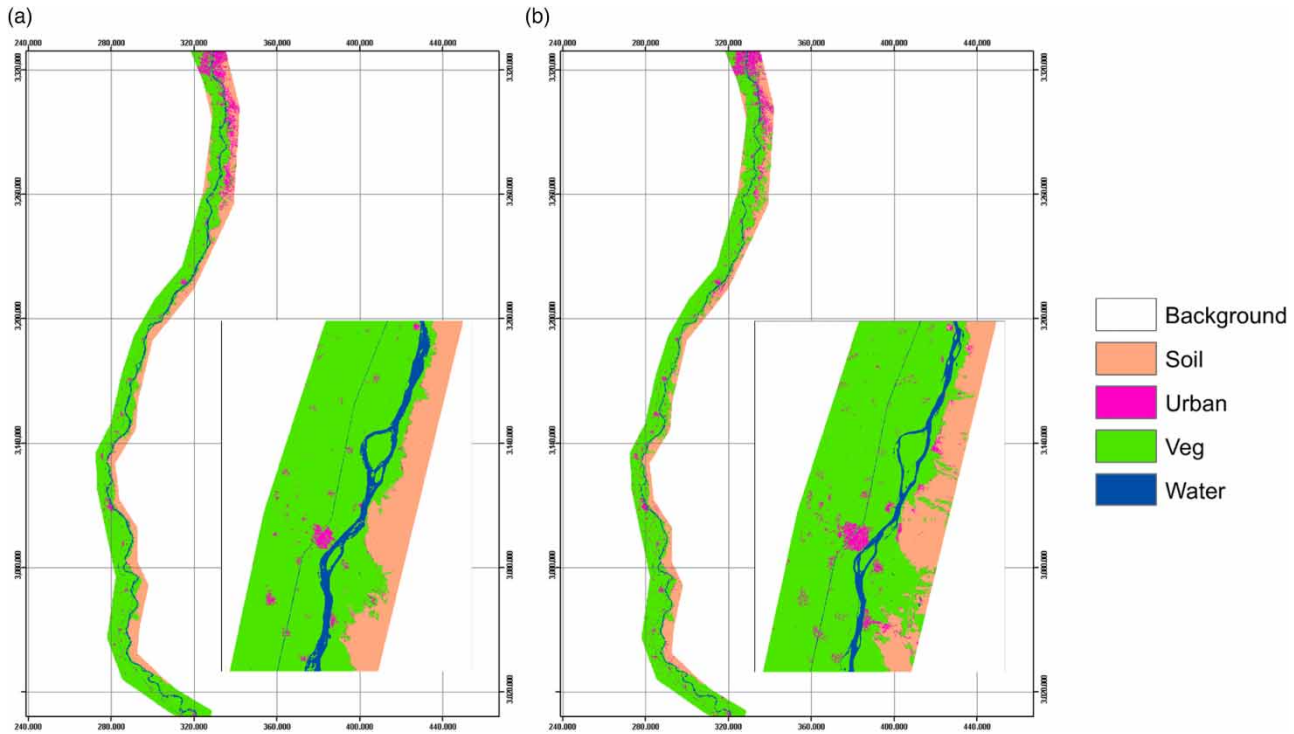


Figure 9 | Land cover maps for the study area: (a) for Year 2003 and (b) for Year 2021.

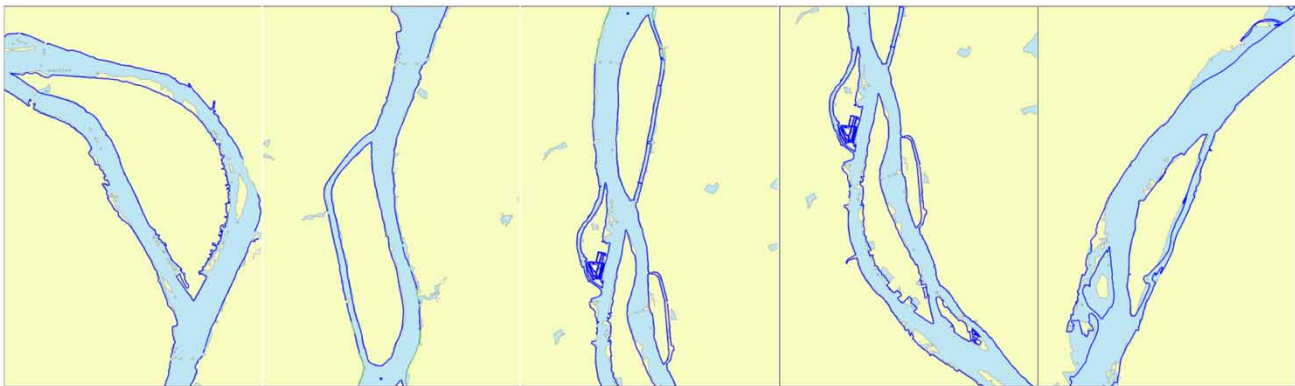


Figure 10 | Sample of the extracted islands in the fourth reach of the Nile River in Egypt in 2003.

bands (eight of them have been used). Hence, the number of attribute fields is different in the two datasets. Additionally, the changes in the conversion of a land cover type to another between 2003 and 2021 affect the DN of bands at the randomly selected training points. This conversion is due to urbanization, land reclamation, and changes in the water level between the two dates, as shown in Figure 9. In addition, by observing the results shown in Figure 9 there are some changes in the boundaries of the Nile stream and the fluvial islands between the 2003 and 2021 images.

For the evaluation process, it can be noticed, from Figure 10, that there is misclassification of the land cover types (water and land), that is because the exact date of collecting the field data was not known, and the water level may change from month to month within the year, which influences the classification results if the images that used for classification are not in the same date of field surveying data collection.

Year 2003	Soil	Urban	Veg	Water	Kappa	Producer Accuracy	User Accuracy
Soil	24755	2833	1459	2	0.83	89%	85%
Urban	2802	20123	1924	124	0.76	52%	81%
Veg	309	15532	108825	1639	0.67	96%	86%
Water	0	97	1549	14449	0.89	89%	90%

Overall Accuracy = **86%**

Overall Kappa = **0.75**

Year 2021	Soil	Urban	Veg	Water	Kappa	Producer Accuracy	User Accuracy
Soil	22151	2978	1818	9	0.80	86%	82%
Urban	1667	22607	2003	169	0.81	51%	85%
Veg	1794	18600	107372	1776	0.60	96%	83%
Water	2	477	1071	11924	0.88	86%	89%

Overall Accuracy = **84%**

Overall Kappa = **0.71**

Figure 11 | Assessment of the accuracy of the land cover classification using the confusion matrix method.

For validating the accuracy of the islands' boundaries, transects with 250 m distance apart were used. The difference between the extracted boundaries by ML and the actual boundaries extracted by the field survey were measured. By analyzing the differences between the ML class result and the ground truth, it was found that the error in the ML classification results is within 11.7 m, which is less than 0.5-pixel size, which can be considered as an acceptable boundaries extraction result for further analysis.

Additionally, for the land cover classification assessment, the confusion matrix method with the large number of points is adequate for the evaluation. The obtained overall accuracy and overall Kappa statistics show a trust in the obtained land cover types for both years, where the overall accuracies are 86 and 84% for 2003 and 2021, and the overall Kappa statistics are 0.75 and 0.71 for 2003 and 2021, respectively. These Kappa statistics values can be considered substantial results according to [McHugh \(2012\)](#).

4.2. Island classification

The large islands in the fourth reach of the Nile River in Egypt have been extracted from the images of the year 2021. These large islands are the islands that have areas greater than 50 fed. For the island classification, two aspects are considered: the geometric characteristics of the islands represented in the relative width of the islands and the type of the land cover of the islands.

4.2.1. Geometric classification schema

According to the described geometric classification scheme in the Methodology section, the geometric characteristics of islands represented by the relative width (R) of the island are calculated. As illustrated in [Table 2](#), by following the criteria

Table 2 | Geometric measurements of the islands

Island #	Area (m ²)	Island width (m)	Island length (m)	Left stream width (m)	Right stream width (m)	R ratio
1	5,492,824.25	1635.4	5,805.81	318.5	282	2.72
2	2,707,465.53	838.4	3,844.1	141.8	156	2.82
3	1,592,887.43	649.6	3,207	372.2	51	1.53
4	494,505.14	518.6	1,687.5	63.5	249	1.66
5	1,730,291.10	648	4,064.2	213.4	313	1.23
6	226,976.68	270.4	1,389.5	534.2	89	0.43
7	551,073.50	445.5	1,778.8	448.8	65	0.87
8	476,519.73	381.9	1,725.25	338.5	146	0.79
9	2,420,487.29	995.7	4,097.93	465.2	80	1.83
10	345,517.02	205.5	1,744.6	50	483	0.39
11	3,115,568.36	1296.8	4,139	205.6	304	2.54
12	1,916,386.12	860.6	3,704.6	311.9	56	2.34
13	5,176,880.12	1363.8	5,641.5	50.9	490	2.52
14	3,614,221.39	1182.4	4,695.6	453.5	93	2.16
15	5,175,426.49	1871.5	5,017.5	110.6	237	5.39
16	1,122,166.31	727.9	2,312.4	204.4	179	1.90
17	1,445,527.52	992.7	2,381	242.4	67	3.21
18	2,248,042.08	841.9	3,459.5	51	374	1.98
19	574,114.33	529	1,769.95	289.6	122	1.29
20	2,181,603.55	1204.3	3,280.3	148.7	349	2.42

shown in [Table 4](#), and by analyzing the calculated *R* for the 20 islands, it can be noted that there are 70% of the extracted islands (14 islands) classified as Wide Island, 20% (four islands) are classified as Equal Island, and 10% (two islands) as Narrow Island, [Figure 12](#).

4.2.2. Land cover type classification schema

In the land cover type classification, the islands are classified according to the percentage of the land cover type to the total area of the island. Based on the calculated percentage of the land cover type data in [Table 5](#), and based on the mentioned criteria in the methodology, if two-thirds of the island is covered by one land cover class, the island will be assigned to this class. However, if the island is covered by multiple classes, it will be assigned to a mixed class, where the classes cover more than one-third of the islands. By analyzing the data illustrated in [Table 3](#), there is only one island that has an urban area greater than 66%, one island has both the vegetation and the urban are greater than 33%, and the rest of the islands (18 islands) have a vegetation land cover type which is greater than 66% of the area. The results of the island classification based on the land cover are illustrated in [Table 5](#) and are shown in [Figure 13](#). The classification of the islands based on land cover types is important in determining the water consumption for each area, as well as, this information will support the decision-makers in assessing the value of the land if is required for developing studies. Furthermore, vegetation is generally a good indicator of stability ([Ward et al. 1999](#)). The large number of islands classified as vegetation islands can be seen in the results of land cover classification, which is a good indicator of the stability of these islands and, as a result, a good indicator for development plans.

5. CONCLUSIONS

This research proposed a workflow for classifying the fluvial islands in the fourth reach of the Nile River in Egypt. This workflow included the process of the satellite images to extract the fluvial islands and for classifying the land cover of the islands. The process of the images included the use of the ML technique in the classification based on the RF algorithm. After extracting the islands with their land cover types, the results are evaluated and used for the classification of the islands. This study

Table 3 | Percentage of each land cover classes within the islands

Island #	Island area (m ²)	Class percentage		
		Soil (%)	Urban (%)	Vegetation (%)
1	5,492,824.25	0	11	89
2	2,707,465.53	2	40	58
3	1,592,887.43	1	74	25
4	494,505.14	0	1	99
5	1,730,291.10	0	1	99
6	226,976.68	1	3	96
7	551,073.50	1	4	94
8	476,519.73	0	2	98
9	2,420,487.29	0	0	100
10	345,517.02	0	0	100
11	3,115,568.36	1	4	95
12	1,916,386.12	0	0	100
13	5,176,880.12	0	2	98
14	3,614,221.39	0	0	100
15	5,175,426.49	0	0	100
16	1,122,166.31	0	0	100
17	1,445,527.52	0	0	100
18	2,248,042.08	1	4	95
19	574,114.33	1	0	99
20	2,181,603.55	0	0	100

Table 4 | Island classification based on geometric characteristics

Relative width (<i>R</i>)	No. of islands	Class
$R > 1.5$	14	Wide island
$0.5 < R < 1.5$	4	Equal island
$R < 0.5$	2	Narrow island

introduced a scheme for classifying the Nile River islands based on the geometric characteristics of the islands with respect to the dimension of the river (the relation between the islands' width and the width of the river). Another classification scheme has been introduced in this work based on land cover of the islands. These classification types for the islands have not been conducted before. For the assessment process, the error matrix was used to evaluate the land cover classification. For the accuracy assessment of the islands' extraction, a comparison between the boundaries of the islands extracted by the ML technique and surveyed field data has been done, using visual assessment. By analyzing the applied work procedure, the following conclusions can be derived:

- The ML technique with the RF algorithm can be used to extract the fluvial islands out of the Landsat images.
- The RF classifier can be used to produce land cover maps with overall accuracy greater than 84%, with Kappa statistics greater than 0.7.
- There are changes in the boundaries of the Nile stream and the fluvial islands within the fourth reach of the Nile River in Egypt between the Years 2003 and 2021.

Island Classification Based on Geometric Characteristics

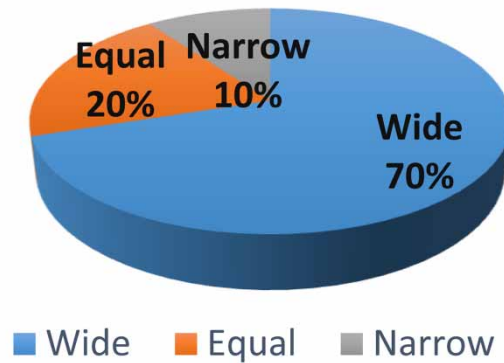


Figure 12 | Island classification based on geometric characteristics.

Table 5 | Island classification based on land cover types

No. of islands	Class
18	Vegetation island
1	Urban island
1	Mixed island

Land Cover Type Classification Schema

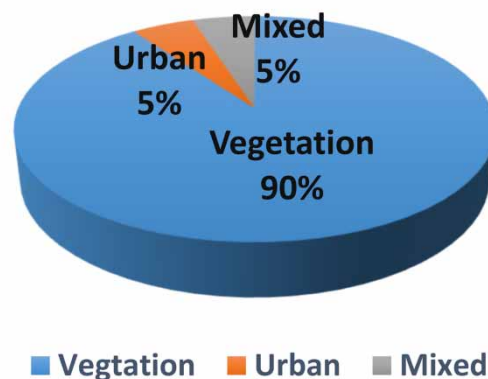


Figure 13 | Island classification based on land cover types.

- Most of the large islands within the fourth reach are wide islands, which have a relative width greater than 1.5, where 14 of 20 islands are wide islands.
- Most of the islands in the fourth reach are covered by vegetation, which more than two-thirds of the island is covered by vegetation (18 islands of 20). However, there is only one island that has an urban land cover type and only one island that has a mixed land cover type (vegetation and urban).

6. RECOMMENDATIONS AND FURTHER RESEARCH WORK

From the experimental work and the results obtained in this research, the following ideas are recommended:

- Applying such work procedure on the entire Nile River within Egypt to support the decision-makers with general information that can be used in the development projects is recommended.
- Using satellite images that have been captured on the same date, or at least in the same month, of the collected field data to avoid the effect of the changes in the water level on the extracted borders of the islands is recommended.
- Using quantitative measurements to evaluate the quality and accuracy of the extracted islands, such as the area and perimeters, or the Digital Shoreline Analysis System to measure the difference between the extracted borders and the real borders in the field will be considered in the future work.
- Depending on other aspects in the island classification will be considered in future work. Such as the stability of the islands, i.e., the changes in the shape, size, and existence of the islands. Using other geometric characteristics, such as aspect ratio (ratio between the length and the width of the islands).
- In this study, island classification can be expanded in similar and different environments (other climatic and topographic regions), but it should be taken into consideration, the stability of islands especially in unregulated rivers. Additionally, applying the ML technique with the RF algorithm is applicable for extracting the fluvial islands and the land cover types. This approach can be applied in any other water body.

ACKNOWLEDGEMENT

The authors thank the United States Geological Survey (USGS) for providing the data for free.

DISCLOSURE STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Beechie, T. J., Liermann, M., Pollock, M. M., Baker, S. & Davies, J. 2006 [Channel pattern and river-floodplain dynamics in forested mountain river systems](#). *Geomorphology* **78**, 124–141.
- Breiman, L. 2001 [Random forests](#). *Machine Learning* **45** (1), 5–32.
- Jamali, A. 2019 [Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran](#). *SN Applied Sciences* **1** (11), 1–11.
- Mahmoud, O., George, Z., Mohammed, H., Ahmad, W. & Jos, L. 2021 [Optimizing regional climate model output for hydro-climate applications in the Eastern Nile Basin](#). *Earth Systems and Environment* **5**, 185–200. <https://doi.org/10.1007/s41748-021-00222-9>.
- McHugh, M. L. 2012 [Interrater reliability: the kappa statistic](#). *Biochemia Medica* **22** (3), 276–282.
- Mutanga, O., Adam, E. & Cho, M. A. 2012 [High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm](#). *International Journal of Applied Earth Observation and Geoinformation* **18**, 399–406.
- Nguyen, H. T. T., Doan, T. M. & Radeloff, V. 2018 [Applying random forest classification to map land use/land cover using Landsat 8 OLI](#). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLII-3/W4**, 363–367.
- NRI 1989 *Nile Research Institute Database*. El-Qanater.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J. P. 2012 [An assessment of the effectiveness of a random forest classifier for land-cover classification](#). *ISPRS Journal of Photogrammetry and Remote Sensing* **67**, 93–104.
- Sadek, N. 2012 [Island development impacts on the Nile River morphology](#). *Ain Shams Engineering Journal*. <http://dx.doi.org/10.1016/j.asej.2012.06.006>.
- Sadek, N., Soliman, M. M. & Abdelbary, M. R. 1999 Nile River morphological changes; case Study Rosetta Branch. In: *ASCE Middle East Regional Conference and International symposium's environmental*, May 16–18, Egypt, Cairo, p. 587.
- Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A. & Rahman, A. 2020 [Land-use land-cover classification by machine learning classifiers for satellite observations – a review](#). *Remote Sensing* **12**, 1135. <https://doi.org/10.3390/rs12071135>.
- Thanh Noi, P. & Kappas, M. 2017 [Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery](#). *Sensors* **18** (2), 18. doi:10.3390/s18010018.

- Thorp, J. H. 1992 Linkage between islands and benthos in Ohio River, with implications for riverine management. *Canadian Journal of Fisheries and Aquatic Sciences* **49** (9), 1873–1882.
- Ward, J. V., Tockner, K., Edwards, P. J., Kollman, J., Bretschko, G., Gurnell, A. M., Petts, G. E. & Rosarro, B. 1999 A reference river system for the Alps: the Fiume Tagliamento. *Regulated Rivers* **15**, 63–75.
- Wyrick, J. R. & Klingeman, P. C. 2011 Proposed fluvial island classification scheme and its use for river restoration. *River Research And Applications* **27**, 814–825. John Wiley & Sons, Ltd. doi:10.1002/rra.1395.
- Zhang, T., Su, J., Xu, Z., Luo, Y. & Li, J. 2021 Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Applied Sciences* **11** (2), 543. doi:10.3390/app11020543.

First received 13 November 2021; accepted in revised form 25 February 2022. Available online 9 March 2022