

Automation of image processing through ML algorithms of GRASS GIS using embedded Scikit-Learn library of Python

Polina Lemenkova 

Department of Biological, Geological and Environmental Sciences (BiGeA), Alma Mater Studiorum – Università di Bologna, Via Irnerio

42, Bologna, IT-40126, Emilia-Romagna, Italy

Faculty of Agricultural, Environmental and Food Sciences, Libera Università di Bolzano, Piazza Università 5, Bolzano, IT-39100, Trentino-Alto Adige, Italy

ARTICLE INFO

Communicated by F. Bertrand

MSC:

86A30

86-08

86A99

86A04

Keywords:

Geoinformation

Machine learning

Cartography

Informatics

Remote sensing

Satellite image

Geospatial analysis

ABSTRACT

Image processing using Machine Learning (ML) and Artificial Neural Network (ANN) methods was investigated by employing the algorithms of Geographic Resources Analysis Support System (GRASS) Geographic Information System GIS with embedded Scikit-Learn library of Python language. The data are obtained from the United States Geological Survey (USGS) and include the Landsat 8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) multispectral satellite images. The images were collected on 2013 and 2023 to evaluate land cover categories in each of the year. The study area covers the region of Nile Delta and the Faiyum Oasis, Egypt. A series of modules for raster image processing was applied using scripting language of GRASS GIS to process the remote sensing data. The satellite images were classified into raster maps presenting the land cover types. These include 'i.cluster' and 'i.maxlik' for non-supervised classification used as training dataset of random pixel seeds, 'r.random', 'r.learn.train', 'r.learn.predict' and 'r.category' for ML part of image processing. The consequences of various ML parameters on the cartographic outputs are analysed, such as speed and accuracy, randomness of nodes, analytical determination of the output weights, and dependence distribution of pixels for each algorithm. Supervised learning models of GRASS GIS were tested and compared including the Gaussian Naive Bayes (GaussianNB), Multi-layer Perceptron classifier (MLPClassifier), Support Vector Machines (SVM) Classifier, and Random Forest Classifier (RF). Though each algorithm was developed to serve different objectives of ML applications in RS data processing, their technical implementation and practical purposes present valuable approaches to cartographic data processing and image analysis. The results shown that the most time-consuming algorithms was noted as SVM classification, while the fastest results were achieved by the GaussianNB approach to image processing and the best results are achieved by RF Classifier.

1. Introduction

1.1. Background

Remote Sensing (RS) data are the key source of spatial information that enable to monitor the physical characteristics of the Earth. Satellite images provide information on the variety and dynamics of landscapes visible from space. Comparison of images as time series enables to recognise land cover types and detect changes remotely, which presents a cost-effective approach to environmental mapping. Numerous existing RS-based environmental analysis presented numerous case studies that aimed at revealing changes in land cover types [1–4]. Such and similar works reported the use of Geographic Information System (GIS) as a tool for RS-based mapping.

The question now arises regarding how to process RS data effectively and accurately. Image processing and classification are considered one of the most significant issues in RS. Accuracy and speed of

image processing strongly depend on selection of methods and approaches which requires advanced tools and algorithms. The advanced methods of RS data handling enable to retrieve the multi-faceted information which highlights diverse aspects of the Earth's landscapes. They can be further utilised for diverse subjects, including geological mapping [5,6], climate change [7–9], environmental monitoring [10, 11] and hazard risk assessment in the coastal areas [12–17].

The geoinformation obtained from the satellite images contributes to deepening our knowledge in Earth sciences and thus increase our understanding of the complex processes interrelated between the environmental processes reflected in ecosystems, climate change and anthropogenic activities and their responses by the natural landscapes. The multidisciplinary of the RS data and geoinformation on the one hand and the applicability of the advanced technical methods (programming, Artificial Intelligence (AI) and Machine Learning

E-mail addresses: polina.lemonkova2@unibo.it, polina.lemonkova@unibz.it.

<https://doi.org/10.1016/j.exco.2025.100180>

Received 7 January 2024; Received in revised form 2 October 2024; Accepted 25 January 2025

Available online 3 February 2025

2666-657X/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(ML)) on the other hand, provide effective tools for environmental analysis.

The Earth observation data are provided in large quantities which requires advanced methods of their rapid and accurate processing [18–20]. Diverse approaches provide different technical solutions to process satellite images effectively. These include either programming libraries [21–24], or embedded modules in the existing GIS software [25–27] or developed plugins for processing data from airborne or spaceborne sensors [28–31]. However, technical parameters of images and approaches of algorithms vary between programs, which results in the problem of selecting optimal methods of image processing.

ML presents a powerful and robust method for RS data processing. Though not all GIS include the ML options in their toolsets for diverse reasons (e.g., lack of fixed tools, awareness of options), the choice methods for processing RS data depends on the availability of algorithms. To meet the needs of image processing, these may vary from either scripting libraries or combining GIS with programming in a diversified workflow [32–37]. Considering the diversity of the ML approaches and their existent nuances that result in quality of image processing, the algorithms should be tested and compared to evaluate their performance. Rather, the choice of methods of RS data processing enables options that allow for algorithms to be used at either background tools or tuning steps for better image analysis and interpretation [38–40].

The expanded opportunity to exercise choice in ML methods of RS data processing is argued to promote advancements in satellite image analysis by not only expanding the choices of methods but also stimulating higher quality of mapping and cartographic output [41–43]. Hence, ML facilitates mapping workflow through moving the priorities from technical issues of image processing to a deeper analysis. Such benefits are meant to operate through the selection of optimal ML methods creating opportunities and choice amongst the available tools of RS data processing [44].

1.2. State-of-the-art

In the early 1970s, when the GIS has started its development, the cartographic possibilities of satellite image processing were limited to the existing restricted functionality of GIS. In contrast, modern frameworks use programming libraries to enhance technical capabilities of GIS through the AI and ML. To achieve the main objective of RS data processing, that is, to accurately and rapidly extract information from satellite images, the data should be processed using advanced computer-based techniques. In this case, the algorithms of RS data processing leverage diverse classification methods [45,46].

Nowadays, processing remote sensing data has become more flexible and accurate due to the use of the Artificial Neural Networks (ANN). The ANN is a model in ML whose architecture and operation are simulated after biological neural networks. Layers of artificial nodes, which are networked units or hidden nodes that approximate connected brain neurons, make up the ANN [47,48]. In RS approach, ANN-based image classification uses this technique to improve classification efficacy. Using ANN algorithms that optimise hidden neurons, land cover types and ground objects on the satellite images are recognised and discriminated more correctly, which enhances classification quality, as reported in existing studies [49–53]. The advantage of this approach is that it employs different environmental data sources, GIS, and statistical datasets to train ANN model [54]. Such data-driven approach enables to better recognise complex landscape patterns for environmental mapping using spatial-temporal data and ANN algorithms [55] and is useful in environmental forecasting [56,57].

The traditional approaches of satellite image processing are implemented by many algorithms, including supervised and non-supervised classification, clustering, segmentation, semantic identification, partition, and object based image analysis (OBIA). However, ML algorithms of image analysis are not accessible to all GIS due to development

constraints and their complexity. Based on this, the choice of the ML data-driven methods is a practice that is not always available to the cartographers and that is mainly utilised in computer science. In contrast to the conventional approaches, the principal aspect of the ML consists in the application of the AI paradigm.

This novel ML-based methods of satellite image processing use the mathematical and statistical models which enable to automatically learn from the input dataset. The generated output based on this learning is performed in a real time regime, during image processing [58,59]. Such approach specifically uses input information as a seed to learn complex patterns and relationships within the dataset. The use of ML choice has been flexibly adjusted to diverse applications to find patterns and links in datasets. For instance, ML can be applied for precise forecasting [60], predictive analysis in hazard risk assessment [61,62], geological mapping [63,64], hydrological modelling [65,66], environmental assessment [67], marine research [68] and evaluation of urban sprawl [69]. In this work, we use the techniques of Geographic Resources Analysis Support System (GRASS) GIS scripting software [70] for processing RS data using AI and ML.

1.3. Objectives and goals

The objective of this study is to investigate the performance of ML and ANN methods for image processing by employing the algorithms of GRASS GIS with embedded Scikit-Learn library of Python. This investigation implies the comparison of methods and models' performance for each case: Gaussian Naive Bayes (GaussianNB), Multi-layer Perceptron classifier (MLPClassifier), Support Vector Machines (SVM), and Random Forest (RF). The comparison aims to determine which algorithm provides the best performance for RS data processing with an example of multispectral satellite images.

The technical goal of this study is to validate the ML tools applied to support the advanced visualisation of the RS data for geospatial modelling and mapping of the coasts of northern Egypt. Unlike the traditional methods of mapping that are usually applied in RS data processing, the integrated approach which combines programming for satellite image processing and cartographic methods for visualisation paves the way to the more advanced ways of handling Earth observation data. Specifically, it combines programming and GIS which is useful for mapping highly heterogeneous landscapes in the arid regions where the mosaic vegetation pattern contrasts from the desert areas, hinterland with agriculture and Faiyum Oasis, northern Egypt (Fig. 1).

This research is focused on the environmental mapping of selected region of Egypt using RS data processed by AI and ML methods using GRASS GIS scripts. The GIS that have included ML modules to select from perform better for image processing compared to simplified programs. Amongst the types of such available tools, GRASS GIS provides the options of ML for RS data processing that are based on the Python's library Scikit-Learn [71], such as, for instance, RF, Nearest Neighbours algorithms (KNeighborsClassifier), Stochastic Gradient Descent (SGD) learning classifier, linear classifiers that include SVM, logistic regression and GradientBoostingClassifier.

The ML solutions use a range of mathematical techniques that can be adjusted to diverse goals of image processing. Besides, the robust Python toolset enhances the image processing through optimised workflow. In this way, the libraries of Python embedded in GRASS GIS ensure the quality and accuracy of image classification [72]. While GRASS GIS also has a wide variety of traditional modules for processing vector and raster types of geospatial data, such as tools for geomorphological computations, hydrological modelling or analytical data visualisations, these options differ in that they predominantly adjusted to cartographic data using internal modules rather than Python algorithms.

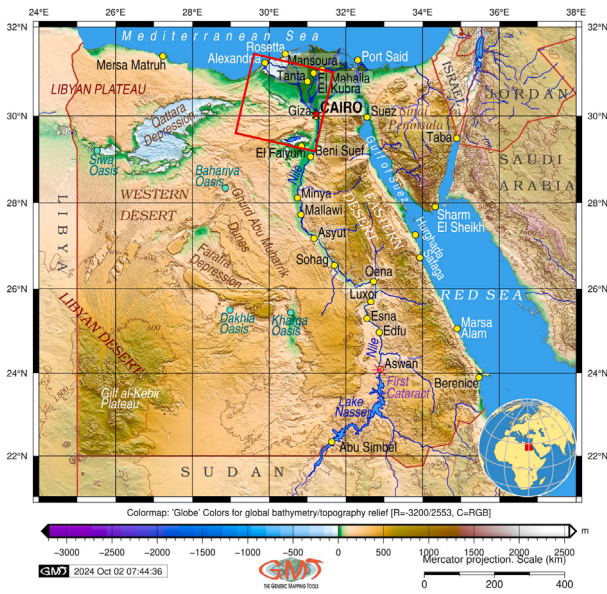


Fig. 1. Location of the study area (rotated red square) on the topographic map of Egypt: western Nile Delta and the Faiyum Oasis.

Source: Author.

2. Image processing

2.1. Data

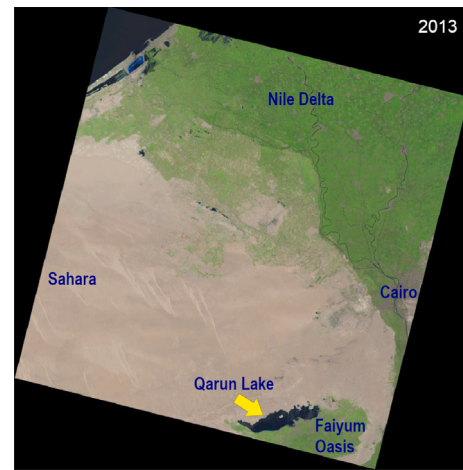
The data are presented by the Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) satellite images covering western segment of the Nile Delta and the Faiyum Oasis, northern Egypt in 2013 and 2023 Fig. 1. The satellite images Fig. 2 were selected based on their quality, coverage and suitability, and downloaded from the EarthExplorer website: <https://earthexplorer.usgs.gov/>.

The individual technical characteristics of the two images are summarised in Table 1. The data are provided by the USGS as open source repository from the Collection Number 2, Category T1. The row/path of the Landsat swath is 39/177 for both the scenes, collected in date time in nadir. Both images were projected in the Universal Transverse Mercator (UTM) Zone 36 for Egypt in World Geodetic System 1984 (WGS84) Datum. The images on 2013 and 2023 were selected to evaluate the decade dynamics of the landscapes around the Faiyum Oasis, northern Egypt within the 10-year time span.

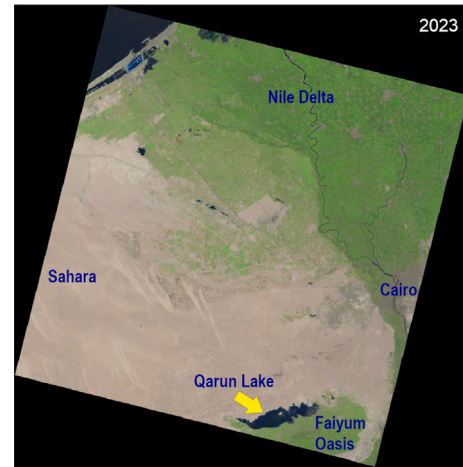
2.2. Workflow

The workflow of GRASS GIS tested on the RS data includes image preprocessing, classification, ML algorithms and accuracy analysis. The images were imported into GRASS GIS using module 'r.import'. The preprocessing and atmospheric corrections were applied to normalise the moderate-resolution images and eliminate the effects of the atmosphere using 'i.landsat.toar:' module. The atmospheric correction of the images was carried out in GRASS GIS using 'i.landsat.toar' which computed the top-of-atmosphere reflectance and temperature for Landsat images. This adjustment transforms the calibrated Digital Number (DN) of the Landsat scenes to top-of-atmosphere reflectance thus improving the quality of the image. Here the atmospheric correction model was applied based on the dark object subtraction (DOS) correction approach.

At the next step, 10 groups of land cover categories were defined using information received on previous step in order to recognise these classes on actual study area. Following that, the images were processed using four different ML algorithms of GRASS GIS by the combination



(a)



(b)

Fig. 2. Data sources: Landsat 8 OLI/TIRS multispectral satellite images covering western Nile Delta and the Faiyum Oasis, Egypt: (a): July 2013; (b): July 2023. Major geographic features are annotated on the images.

of modules "r.random" (which generates training pixels from the land cover classification), "r.learn.train" (which trains models using various ML algorithms) and "r.learn.predict", which performs the prediction of the pixels' assignment into target categories based on the values of their spectral signatures.

The accuracy of the assigned classes was estimated using Kappa statistics which evaluates chance agreement of the correctly classified pixels. This method is based on the estimation of the observed versus expected values of pixels in each land cover class of the output raster matrix. Practically, it has been calculated module "r.kappa" of GRASS GIS to validate the results of image classification. The training dataset was based on the raster data from the Global Land Cover Characteristics (GLCC) achieved from the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) Pathfinder satellite images.

2.3. Maximum likelihood

While maximum likelihood (MaxLike) estimation (Fig. 3) mostly pursue statistical and evaluation goals, the advent of this method which aims at the estimating the probability distribution in dataset have created opportunities to the pursuit of satellite image processing. Thus, the MaxLike method enables a straightforward approach to partition of pixels into groups using clusters of cells with similar spectral reflectance

Table 1

Technical parameters of the Landsat-8 OLI/TIRS images used as dataset.

Source: EarthExplorer by the United States Geological Survey (USGS).

Data set attribute	Attribute value	Attribute value
Landsat Product Identifier L2	LC08_L2SP_177039_20230721_20230803_02_T1	LC08_L2SP_177039_20130725_20200912_02_T1
Landsat Product Identifier L1	LC08_L1TP_177039_20230721_20230803_02_T1	LC08_L1TP_177039_20130725_20200912_02_T1
Landsat Scene Identifier	LC81770392023202LGN00	LC81770392013206LGN01
Date Acquired	2023/07/21	2013/07/25
Roll Angle	-0.001	0.000
Date Product Generated L2	2023/08/03	2020/09/12
Date Product Generated L1	2023/08/03	2020/09/12
Start Time	2023-07-21 08:29:28	2013-07-25 08:31:45.103183
Stop Time	2023-07-21 08:30:00	2013-07-25 08:32:16.873179
Land Cloud Cover	0.00	0.00
Scene Cloud Cover L1	0.00	0.00
Ground Control Points Model	662	821
Geometric RMSE Model	7.456	6.176
Geometric RMSE Model X	4.877	3.954
Geometric RMSE Model Y	5.640	4.744
Processing Software Version	LPGS.16.3.0	LPGS.15.3.1c
Sun Elevation LORA	66.45746837	66.44030819
Sun Azimuth LORA	109.05483881	111.66370294
TIRS SSM Model	FINAL	ACTUAL
Data Type L2	OLI_TIRS_L2SP	OLI_TIRS_L2SP
Scene Centre Latitude	30.30634	30.30620
Scene Centre Longitude	30.40632	30.40229
Corner Upper Left Latitude	31.34283	31.34003
Corner Upper Left Longitude	29.15158	29.14854
Corner Upper Right Latitude	31.39299	31.39022
Corner Upper Right Longitude	31.60323	31.59696
Corner Lower Left Latitude	29.20342	29.20333
Corner Lower Left Longitude	29.23428	29.23120
Corner Lower Right Latitude	29.24948	29.24941
Corner Lower Right Longitude	31.63332	31.62715

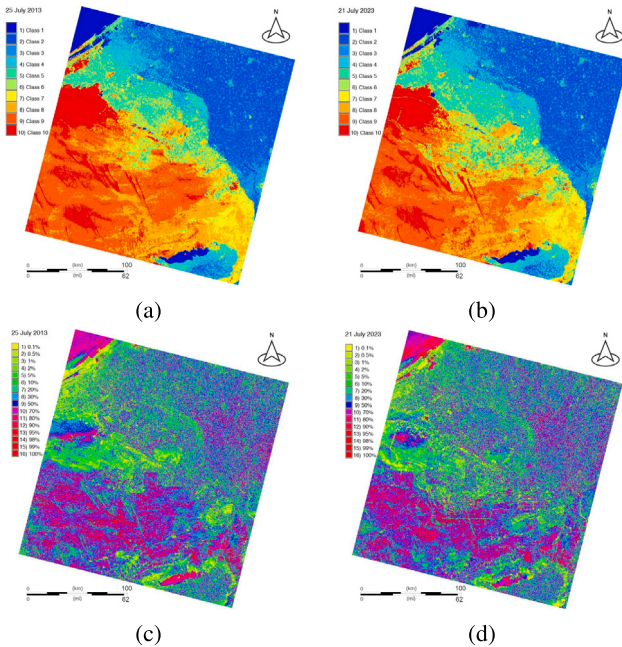


Fig. 3. Unsupervised automatic classification using clustering and maximal likelihood approaches, and computed rejection probability classes by GRASS GIS: (a): Classification 2013; (b): Classification 2023; (c): Accuracy assessment: 2013; (d): Accuracy assessment: 2023.

within the observed dataset [73]. Here, only multispectral bands (Bands 1 to 7) of the Landsat –8 images were selected, without panchromatic (Band 8), Cirrus (Band 9) and Thermal Infrared (Bands 10 and 11). To this end, the multispectral bands were grouped using command *subgroup = res_30m* that enables to select only the target bands. Then, the number of classes was defined to 10 and the classification was

performed using 'i.cluster' and 'i.maxlik' modules of GRASS GIS, as demonstrated in script below:

```
i.cluster group=L8_2023_Faiyum subgroup=res_30m \
signaturefile=cluster_L8_2023_Faiyum \
classes=10 reportfile=rep_clust_L8_2023_Faiyum.txt
i.maxlik group=L8_2023_Faiyum subgroup=res_30m \
signaturefile=cluster_L8_2023_Faiyum \
output=L8_2023_cluster_classes \
reject=L8_2023_cluster_reject
```

The numerical results of the classification are stored using 'report-file' function which contains statistical parameters of cluster classes: class size, separation, percent convergence, number of iterations, means and standard deviations for Landsat bands and class separability matrix. Sample pixel sizes were selected for every images (sample size: 6676 points for year 2013 and 6678 for 2023); the standard and mean deviation of the partition were calculated. Training cluster maps are computer for 10 classes of land cover types over the study area for each scene: (1) urban areas, (2) agricultural fields, (3) water areas in the Nile river and Mediterranean Sea, (4) coastal shelf zones, (5) consolidated base lands and soils; (6) sandy areas in the desert; (7) natural vegetation in the non-urban areas; (8) croplands; (9) salt; (10) sabkha. The polygons of these categories were used as training data in the ML approach of image classification. The outcomes are validated with computed reject threshold results using chi square test to determine levels of confidence of correctly classified pixels on raster scenes. The output of the images processed using MaxLike and computed rejection probability classes are shown in Fig. 3.

2.4. GaussianNB Classifier

The GaussianNB Classifier is originally incepted in 18th century by Reverend Thomas Bayes and described in the theorem. Now it is widely used in ML and computer science applications [74]. In the GRASS GIS, it is derived from the Scikit-Learn Library of Python. The GaussianNB algorithm is generally based on the supervised learning method which

implies the probabilistic approach and normal Gaussian distribution of the data [75]. Here the likelihood of the pixels assigned to each land cover class during image partition is assumed to be Gaussian according to Eq. (1):

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (1)$$

where the parameters σ_y and μ_y are computed using the approach of maximum likelihood. Thus, the algorithm of GaussianNB Classifier computes the likelihood of the features using the Bayes' theorem which assumes the independence between every pair of pixels according to Gaussian distribution and relies on probabilities of samples [76]. Specifically, the technique of GaussianNB Classifier examines feature types as variables by considering the value of each cells assigned to target classes of the images [77]. In the scripting environment of GRASS GIS, the Gaussian NB Classifier was implemented using the following script:

```
r.random input=L8_Faiyum_2013_cluster_classes seed=100 \
  npoints=1000 raster=training_pixels
r.learn.train group=L8_2023_Faiyum
  training_map=training_pixels \
  model_name=GaussianNB n_estimators=500 \
  save_model=GNB_model.gz
r.learn.predict group=L8_2013_Faiyum \
  load_model=GNB_model.gz output=GNB_classification
r.category GNB_classification
```

The parameters were selected as follows: number of seed for rand() function is selected as 100 with number of evaluated cells within the raster image is 1000. The selected value is chosen to achieve the better accuracy, since cells in the resulting raster map overlap with those of the input image, which based on the current computational region according to the coordinate extent of the images. The points in the resulting map based on image classification are located in cell centres of these cells, hence the cells are associated with the random point locations in the original Landsat scene.

2.5. MLPClassifier

MLPClassifier is a Multi-layer Perceptron classifier derived from the Scikit-Learn Library of Python and is based on fundamentals of predictive learning [78]. The performance of the MLPC differs from the GaussianNB approach since it uses the principle of feedforward Artificial Neural Network (ANN). For data analysis, ANN uses three layers that are used as structures of network topology in a flow of information for data partition. The input, hidden and output layer consist of nodes (i.e., pixels of raster layer) which train the model using supervised learning [79,80].

The principal approach of this process consists in the connection between each node in one layer to those in the following layer through a certain weight (w_{ij}). MLPClassifier iteratively evaluates the training data using these connections in weights of the evaluated pixels. The algorithm changes weights repetitively using estimated error until the output image is approached to the expected result and the error is minimised. This sequential operation is formulated in Eq. (2):

$$\epsilon(n) = \frac{1}{2} \sum_{output\ node\ j} e_j^2(n) \quad (2)$$

where $e_j(n)$ is the degree of error in an output node j in the n th pixel of the raster dataset and $\epsilon(n)$ is the node weights which are iteratively adjusted using the minimisation of the errors in the classified raster image for the n th pixel of the raster matrix. Using optimisation, each weight w_{ij} is estimated iteratively and changed accordingly using Eq. (3):

$$\Delta w_{ji}(n) = -\eta \frac{\partial \epsilon(n)}{\partial w_{ji}(n)} y_i(n) \quad (3)$$

where $y_i(n)$ is the result of the previous step of classification, and η is the tuning parameter of optimisation, which aims at quick convergence of the weights of pixels during the iterative process of image classification.

Hence, the MLPC algorithm is sensitive to feature scaling which is related to the resolution of the original raster image. The essential approach of this algorithms consists in random selection of the hidden nodes and analytical determination of the output weights of neural networks [81]. As a result, the MLPClassifier algorithm ensures higher generalisation output at a faster learning speed.

2.6. SVC

The SV Classifier is a branch of the SVM which are a set of supervised learning methods [82]. Based on the statistical learning frameworks developed by [83], SVM is now among the most applicable models in image processing due to the efficiency of its algorithm [84]. The main approach of the SVM is classifying the new data into one or two given classes using the decision process which tracks the largest separation, i.e., margin, between the given classes. Hence, the best separation is obtain using the discrimination of the location of the samples within the finite-dimensional space of dataset, and detected largest distance to the nearest training sample of one of these classes [85,86].

The mathematical foundation of the SVM algorithm consists in the following. When SVC uses training vectors, that is, the sample pixels that are located within the margin of classes, $x_i \in R^p$ where $i = 1 \dots n$ as two classes and a vector $y \in \{1, -1\}^n$, it aims to find the $w \in R^p$ and $b \in R$ so that the assignment of pixels to correct land cover class is true for most samples. The formulated problem is expressed in Eq. (4):

$$\min_{\omega, b, \zeta} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \zeta_i \quad (4)$$

which depends on the conditions in the definitions of $y_i(\omega^T \varphi(x_i) + b) \geq 1 - \zeta_i$ which should ideally be greater than one for the optimal prediction of the correctly classified pixels on a raster scene, and $\zeta_i \geq 0, i = 1, \dots, n$. The SVC solves this problem through the maximisation of margins of the sample classes by analysing the Digital Numbers (DN) of pixels on the raster image and inverse regularisation parameter. After the optimisation of the raster dataset using iterative analysis, the estimated decision function in a classified matrix consisting of cells for a sample of x pixels corresponds to Eq. (5):

$$\sum_{i \in SV} y_i \alpha_i K(x_i, x) + b \quad (5)$$

The resulting output of the classification assigns the predicted class to its sign and the support vectors are summarised using attributes of the classification. Such effective approach has a memory-efficient performance since it utilises a set of training points as decision function, i.e., support vectors which optimises the use of computational facilities of the machine. Moreover, the flexibility of this algorithms is caused by the possibility to adjust the model using different Kernel functions defined which includes both common and custom kernels.

2.7. RandomForestClassifier

The RF Classifier initially developed by Tin Kam Ho in 1995 [87] became a well established method of data complexity analysis among ML algorithms [88]. The essential of the RF ML ensemble learning approach of image classification is that it generates a randomised multitude of decision trees during the process of training [89]. It uses the training set of classified pixels and builds each tree in the ensemble from this sample with a replacement by introducing randomness in the classification workflow. In image processing, the training algorithm of RF uses the approach of bootstrap aggregating to analyses of pixels [90–92].

The application of RF in image processing consists in selecting a random pixel from the raster matrix, replacing it in the training set and fitting the trees to the samples. This is done through the training set of pixels on the raster image $X = x_1, \dots, x_n$ with responses $Y = y_1, \dots, y_n$. Hence, for $b = 1, \dots, B$, the algorithm uses n training sample pixels and replaces them using data from training map. Afterwards, the classification tree f_b is being trained using sample training of pixels. Finally, the predictions for new sample pixels x' are made using average predictions from all the decision trees using the rule formulated in Eq. (6):

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (6)$$

2.8. Accuracy assessment

The correct identification of the land cover classes is important because the results of mapping can be used for predicting agricultural productivity, land use planning and management. Therefore, it is essential to evaluate the accuracy of image classification using ML algorithms. In GRASS GIS, the Estimated kappa (κ) value in “r.kappa” is the value that was computed for land cover classes, that is, each of the 10 categories. Hence, the observed values of the classified images were evaluated against the training polygons for land cover categories into the target class. The GLCC data were used as reference to validate the results of the classified images.

A sample of points as a ground truth data from the GLCC was compared to the classification results to determine the accuracy of the evaluation. Methodologically, the estimation of kappa (κ) presents the observed agreement between the results of the assigned pixels for those cells that have been classified by each classifier (RF, DT, MLPC and SVM) into the target class of land cover categories. In GRASS GIS, kappa statistics was calculated using “r.kappa” following the formula in Eq. (7):

$$kpp_i = (pii_i - p_{i_i} * p_{j_i}) / (p_{i_i} - p_{i_i} * p_{j_i}); \quad (7)$$

where the pii_i is the number of pixels for which there is agreement divided by overall number of the classified pixels which expresses the probability of agreement; P_{i_i} is the probability of the correct classification i where the target pixel is assigned to the correct category of land cover class i ; P_{j_i} is the probability of classification j to accurately classify pixel i . Practically, the validation was performed using GRASS GIS syntax where input data is given with an example for 2023 and reference taken from the GLCC file. Afterwards, the results on κ statistics were exported as csv file. In GRASS GIS, the accuracy assessment was computed using the following code:

```
g.region raster=L8_2023_Faiyum -p
r.kappa -w classification=L8_2023_Faiyum \
reference=GLCC_Egypt output=kappa.csv -m -h
# saving Kappa matrix to tabular format
r.kappa classification=L8_2023_Faiyum \
reference=GLCC_Egypt output=k_Faiyum.csv -m -h
```

3. Results and discussion

The results of the unsupervised automatic classification of the images using clustering and maximal likelihood approaches are presented in Fig. 4. It also shows the computed rejection probability classes by GRASS GIS visualisation of the processed images. The computed class separability matrix for 10 classes of land cover types identified on the

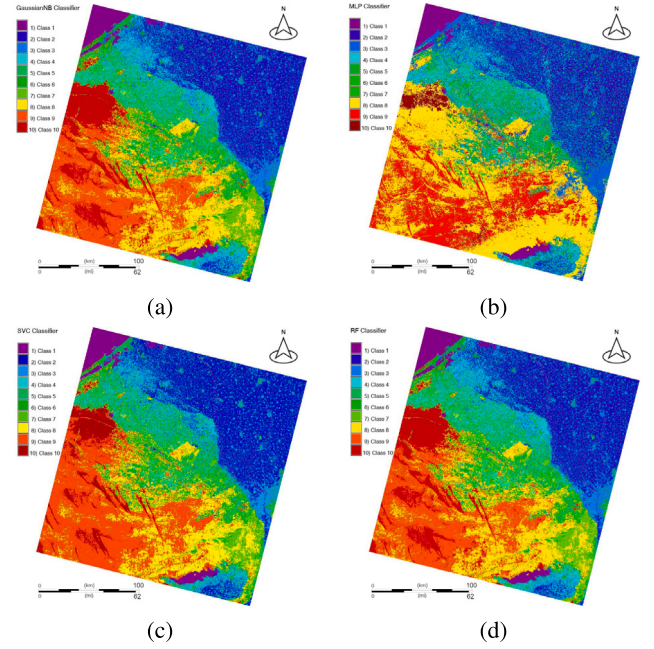


Fig. 4. Cartographic visualisation of the supervised ML-based classification using four different models of GRASS GIS embedded in r.learn.train' module: (a): GaussianNB Classifier; (b): MLP Classifier; (c): SVC Classifier; (d): Random Forest Classifier.

image on 2013 is given in matrix block below.

	1	2	3	4	5	6	7	8	9	10
1	0									
2	1.8	0								
3	2.6	1.7	0							
4	2.4	1.0	1.0	0						
5	3.7	2.8	1.0	1.7	0					
6	5.0	4.3	2.2	2.9	1.2	0				
7	6.9	6.3	3.6	4.4	2.5	1.3	0			
8	8.9	8.3	4.9	5.7	3.6	2.3	1.0	0		
9	10.0	9.4	5.6	6.4	4.4	3.0	1.8	1.0	0	
10	8.6	8.1	5.4	6.0	4.4	3.3	2.4	1.9	1.2	0

The results of the MaxLike unsupervised classification were employed as seed training dataset for ML algorithms. The classification for 2013 reported 98.26% points stable in 10 classes; the classification for 2013 reported 98.01% points stable. The cartographic visualisation of the images processed with four methods is presented in Fig. 4. An essential point when classifying satellite image consists in the accurate selection of the land cover categories which should be identified, recognised on the images and discriminated on the resulting maps. The computed class separability matrix for 10 classes of the land cover types recognised in image on 2023 is given in matrix block below:

	1	2	3	4	5	6	7	8	9	10
1	0									
2	1.8	0								
3	2.8	1.8	0							
4	2.6	1.0	1.1	0						
5	3.9	2.7	0.9	1.6	0					
6	4.9	4.0	1.9	2.8	1.1	0				
7	6.8	5.9	3.2	4.2	2.3	1.1	0			
8	9.0	8.1	4.5	5.8	3.6	2.3	1.1	0		
9	10.0	9.1	5.3	6.5	4.4	3.0	2.0	1.0	0	
10	8.3	7.5	4.9	5.8	4.2	3.2	2.4	1.8	1.1	0

The class separability matrix indicates the values of the class separation for each of the 10 land cover class. Hence, the values means the

distance between the individual classes defined automatically during the iteration process. Below this distance, the categories of land cover types are grouped into one class, which is repeated during the next iterations until the results are optimised. Larger values (over 1.5) indicate the specified distance between the land cover classes, while lower values (0.5 to 1.5) are merged and the process is repeated. Hence, the percentage of accurately classified images with correctly assigned land cover categories is calculated and summarised in the two matrices for each image showing class separability for the years 2013 and 2023.

The land cover types in the northern Egypt were identified using reports from previous existing studies and existing data. These were modified and generalised according to the current study area. For this purpose, land cover type were proposed for the study area with ten categories adopted from the existing classification scheme. As a result, the landscape types around the Nile Delta and Faiyum Oasis were classified and grouped in the ten classes. The identification of these land cover types was then performed automatically using computer vision and ML algorithms embedded in the GRASS GIS. Among the compared outputs, the MLP and SVC Classifiers differ from the RF and NB approaches, due to the different approaches (ANN and SVM methods). The map generated using Naive Bayes Classified (Fig. 4a) is comparable with the map based on the RF Classifier (Fig. 4d), while MLP Classifier presented a more generalised map (Fig. 4b) using deep learning network approach.

Being one of the most efficient inductive learning algorithms of ML in data science, Naive Bayes has advantages in the classification performances among the ML algorithms due to the effects of dependence distribution of nodes within classes and integrated dependence of all nodes [93]. Hence, the Naive Bayes is optimal if the dependences distribute evenly in classes which is excellently suits to the homogeneous types of landscapes with diversified land cover classes. However, this approach is not optimised for images with high landscape heterogeneity.

The advantages of SVM (Fig. 4c) can be explained by its effectivity in high dimensional datasets which is valuable for satellite images that consists thousands of pixels as cells. In our case, 61,405,182 cells were classified automatically for each band of multispectral Landsat 8 OLI/TIRS images with 30-m resolution where each raster image contained 7911 rows and 7762 columns of cells. More specifically, the SVM presents high search accuracy compared to other algorithms after a few iterations. Besides, using SVM is supportive for time series analysis where several satellite images are compared since this method is effective when number of dimensions is larger than samples. This is particular valuable for detecting complex land cover patterns in the regions with high heterogeneity of landscapes, such as mixed areas with urban areas and natural spaces.

The RF classifier presented the best results compared to the other approaches in the presented maps (Fig. 4d). Its superiority is explained by the tuned optimisation of the image classification. Using the parameter tuning in GRASS GIS, each node is being split during the construction of a tree with the best split found either from all input pixels of the image or a randomised subset depending on the resolution and extent of the training dataset of the image which depends on the study area and heterogeneity of the visualised landscapes. This randomness aims at decreasing the variance of the RF estimator which allows to overcome overfitting of the individual decision trees with high variance in case of heterogeneous landscapes. The artificially increased randomness of the RF algorithm enables either to decrease the prediction errors of the decision trees or cancel them out, thus making image classification more accurate.

The assessment of the statistical analysis for the year of 2013 shows the final results report convergence=98.3% and 98.26% points stable. For the year 2023, the classification results report 98.01% points stable and convergence=98.0%. Since the computed accuracy is higher than 80%, the classification of the multispectral image is performed robustly. Therefore, the application of the ML, as demonstrated in this study, can

be advised for similar future studies where image processing is applied for environmental monitoring and mapping. With a kappa value (κ) of 0.74, the study's total classification accuracy was 82.4% for image on year 2013 and kappa value (κ) of 0.73 and classification accuracy was 84.2% for image on 2023. Since the κ coefficient is considered significant for both images (on 2013 and 2023), it points at the applicability of the ML methods of GRASS GIS and their effectiveness for investigations on image processing in future similar research.

4. Conclusion

Nowadays, the important role of programming and ML applications in satellite image analysis is more and more recognised due to the powerful techniques which enable to process images rapidly yet accurately. As a result, modern geoinformatics and RS domain benefit from methods derived from the programming domains, data science, and statistical packages. The use of these tools significantly enriches the cartographic capacities of RS data processing through the advanced methods of modelling and visualisation. ML algorithms lead the environmental mapping to a principally novel stage that complements and enhances commonly used traditional GIS methods for RS data processing. With this regards, this paper demonstrated the case of such applications where ML algorithms were applied and compared to address the issue of the effective RS data handling.

This work demonstrated the examples of the supervised training in ML approaches to geospatial applications. The research evaluated ML algorithms of image classification for improved workflow of image processing and tested their applicability in cartographic works. Four algorithms of ML supervised training were compared using GRASS GIS modules that are based Python's library Scikit-Learn. Based on the maps generated from the processed satellite images four distinct ML approaches of GRASS GIS were compared both for mathematical foundation and graphical output.

The insights presented in this paper have shed light on the potential future directions in RS data processing using ML tools integrated with geoinformatics software GRASS GIS. Further advances in cartographic RS data processing in arid areas can be extended to the common in environmental mapping tasks: droughts, erosion, land suitability for agriculture, etc. Such process affects both natural and social systems, cause damage in infrastructure and disrupt landscapes. Satellite images processed using advanced tools of ML can significantly support geospatial analysis and help to mitigate or prevent risks of natural hazards. Using demonstrated approach, the Research and Development (R&D) groups integrated into the coastal management can perform the RS-based data analysis using programming and scripting methods.

Novel framework of environmental monitoring has an multi- and interdisciplinary profile that combines diverse scientific branches: advanced RS datasets, computer graphics, data science, computing algorithms integrated with GIS methodology or visualisation and spatial analysis. Such integrative approaches pave the ways to further development of marine research and open new perspectives of the environmental mapping. In this paper, we demonstrated such integrated study with a case of Faiyum Oasis, Egypt, and discussed practices of RS data processing in geoinformatics for environmental mapping of arid zones. The study was supported by an overview of modern AI and ML methods of RS data processing in modern geoinformation. Hence, the perspectives of the GIS-based integrated use of programming, ML, DL techniques and satellite image analysis were discussed and highlighted. We also provided the most essential descriptions of the novel characteristics of the AI and ML track specifically for RS data analysis.

Satellite imagery, geo-information resources and maps are a primary requirement in visualising data in environmental monitoring of arid areas. As demonstrated in this study, the use of the open spatial data, such as multispectral Landsat imagery, is crucial for environmental mapping since they provide reliable information with sufficient spatial coverage. The information obtained from such data can be reused for

updated vegetation mapping and operative monitoring to analyse land cover changes. The decisions made based on these datasets include the cost-benefit computations, social-economic analysis, or hazard risk assessment in mixed arid and agricultural areas. In turn, processing multi-temporal time series of satellite images ensures environmental management and assist to mitigate implications for affected ecosystems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] R.C. Carvalho, R. Reef, Quantification of coastal change and preliminary sediment budget calculation using SfM photogrammetry and archival aerial imagery, *Geosciences* 12 (10) (2022) <http://dx.doi.org/10.3390/geosciences12100357>.
- [2] A. Marsden, R. Smith, R. Craig, B. Heath, N. Raisbeck-Brown, M. Steber, J. Adams, K. Church, Continental mapping of fire risk across Australia using the NOAA-AVHRR, in: *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium* (Cat. No.01CH37217), Vol. 3, 2001, pp. 991–993, <http://dx.doi.org/10.1109/IGARSS.2001.976724>.
- [3] P. Lemenkova, Mapping ghana by GMT and r scripting: advanced cartographic approaches to visualize correlations between the topography, climate and environmental setting, *Adv. Geod. Geoinf.* 71 (2022) 1–20, <http://dx.doi.org/10.24425/gac.2022.141169>.
- [4] T.J. Malthus, E.L. Hestir, A. Dekker, J. Anstee, H. Botha, N. Cherukuru, V. Brando, L. Clementsen, R. Oliver, Z. Lorenz, Inland water quality monitoring in Australia, in: *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 2013*, pp. 2872–2875, <http://dx.doi.org/10.1109/IGARSS.2013.6723424>.
- [5] M.M. Abd-AlHai, S.A.S. Araffa, M.M. Mekki, A.A.-M. ElGalladi, A reconnaissance study for tracing and ordering new mineralisation zones using integrated remote sensing, GIS, and aeromagnetic techniques, west allagi-heiani-suture, Egypt, *NRIAG J. Astron. Geophys.* 12 (1) (2023) 19–44, <http://dx.doi.org/10.1080/20909977.2022.2156068>.
- [6] H.M. Hussien, M. Yousif, A. El Sheikh, Investigation of groundwater occurrences in structurally controlled terrain, based on geological studies and remote sensing data: Wadi el morra, south sinai, Egypt, *NRIAG J. Astron. Geophys.* 9 (1) (2020) 512–531, <http://dx.doi.org/10.1080/20909977.2020.1788270>.
- [7] G. Ramat, E.S. and Simonetta Paloscia, and Giacomo Fontanelli, S. Pettinato, L. Santurri, N. Souissi, E.D. Ponte, M.M.A. Wahab, A.A. Khalil, Y.H. Essa, M. Ouessar, H. Dhaou, A. Sghaier, A. Hachani, Z. Kassouk, Z.L. Chabaane, Remote sensing techniques for water management and climate change monitoring in drought areas: case studies in Egypt and Tunisia, *Eur. J. Remote. Sens.* 56 (1) (2023) 2157335, <http://dx.doi.org/10.1080/22797254.2022.2157335>.
- [8] P. Lemenkova, Deep learning methods of satellite image processing for monitoring of flood dynamics in the Ganges Delta, Bangladesh, *Water* 16 (8) (2024) <http://dx.doi.org/10.3390/w16081141>.
- [9] E. Seif, H. Shokr, A.M. Youssef, M.S. Eddeen, Y. Diab, Y.E. Imam, Climate change effect on water temperature and evaporation in lake qarun, *Water Sci.* 37 (1) (2023) 1–10, <http://dx.doi.org/10.1080/23570008.2023.2189220>.
- [10] T.D. Tanushri Govender, C. Shoko, Remote sensing of land use-land cover change and climate variability on hydrological processes in sub-saharan africa: key scientific strides and challenges, *Geocarto Int.* 37 (25) (2022) 10925–10949, <http://dx.doi.org/10.1080/10106049.2022.2043451>.
- [11] P. Lemenkova, Landscape fragmentation and deforestation in sierra leone, west africa, analysed using satellite images, *Transylv. Rev. Syst. Ecol. Res.* 26 (1) (2024) 13–26, <http://dx.doi.org/10.2478/trser-2024-0002>.
- [12] R. Elshazly, A. Armanuos, M. Elshemy, B.A. Zeidan, Chapter 11 - modeling the bathymetry of coastal lakes using remote sensing and gis. Case study of Lake Manzala, Egypt, in: B.A. Zeidan, A.I.M. Aly (Eds.), *The Nile River System, Africa*, in: *Ecologyhydrology from Catchment to Coast*, vol. 2, Elsevier, 2024, pp. 261–273, <http://dx.doi.org/10.1016/B978-0-323-90122-2.00006-1>.
- [13] W. Sun, C. Chen, W. Liu, G. Yang, X. Meng, L. Wang, K. Ren, Coastline extraction using remote sensing: a review, *GIScience & Remote. Sens.* 60 (1) (2023) 2243671, <http://dx.doi.org/10.1080/15481603.2023.2243671>.
- [14] S. Kamh, M. Ashmawy, A. Kilias, B. Christaras, Evaluating urban land cover change in the hurghada area, Egypt, by using GIS and remote sensing, *Int. J. Remote Sens.* 33 (1) (2012) 41–68, <http://dx.doi.org/10.1080/01431161.2010.550331>.
- [15] M. El-Raey, A.F. Abdel-Kader, S.M. Nasr, H.I. El-Gamily, Remote sensing and GIS for an oil spill contingency plan, ras-mohammed, Egypt, *Int. J. Remote Sens.* 17 (11) (1996) 2013–2026, <http://dx.doi.org/10.1080/01431169608948756>.
- [16] P. Lemenkova, Exploitation d'images satellitaires landsat de la région du cap (afrique du sud) pour le calcul et la cartographie d'indices de végétation à l'aide du logiciel GRASS GIS, *Physio- Géo* 20 (2024) 113–129, <http://dx.doi.org/10.4000/11pyj>.
- [17] M. Braham, A. Boufekane, H. Bourenane, B.N. Amara, R. Bensalem, E.H. Oubaiche, Y. Bouhadad, Identification of groundwater potential zones using remote sensing, GIS, machine learning and electrical resistivity tomography techniques in guelma basin, northeastern Algeria, *Geocarto Int.* 37 (26) (2022) 12042–12072, <http://dx.doi.org/10.1080/10106049.2022.2063408>.
- [18] Y. Li, X. Li, Y. Zhang, D. Peng, L. Bruzzone, Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing big data, *Int. J. Appl. Earth Obs. Geoinf.* 120 (2023) 103345, <http://dx.doi.org/10.1016/j.jag.2023.103345>.
- [19] J. Yin, J. Dong, N.A. Hamm, Z. Li, J. Wang, H. Xing, P. Fu, Integrating remote sensing and geospatial big data for urban land use mapping: A review, *Int. J. Appl. Earth Obs. Geoinf.* 103 (2021) 102514, <http://dx.doi.org/10.1016/j.jag.2021.102514>.
- [20] Y. Huang, Z. xin Chen, T. YU, X. zhi Huang, X. fa Gu, Agricultural remote sensing big data: Management and applications, *J. Integr. Agric.* 17 (9) (2018) 1915–1931, [http://dx.doi.org/10.1016/S2095-3119\(17\)61859-8](http://dx.doi.org/10.1016/S2095-3119(17)61859-8).
- [21] A. Cuartero, M.E. Paoletti, P. García-Rodríguez, J.M. Haut, PyCircularStats: A python-based tool for remote sensing circular statistics and graphical analysis, in: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 2876–2879, <http://dx.doi.org/10.1109/IGARSS46834.2022.9884758>.
- [22] P. Lemenkova, O. Debeir, Recognizing the wadi fluvial structure and stream network in the qena bend of the Nile river, Egypt, on landsat 8-9 OLI images, *Information* 14 (2023) 249, <http://dx.doi.org/10.3390/info14040249>.
- [23] S. Barma, S. Damarla, S.K. Tiwari, Semi-automated technique for vegetation analysis in sentinel-2 multi-spectral remote sensing images using python, in: *2020 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA, 2020*, pp. 946–953, <http://dx.doi.org/10.1109/ICECA49313.2020.9297369>.
- [24] M.S. İşik, V. Özbey, S. Erol, E. Tari, GNSSpy: Python Toolkit for GNSS Data, in: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021*, pp. 8550–8553, <http://dx.doi.org/10.1109/IGARSS47720.2021.9553880>.
- [25] A.M.E. Ngolo, T. Watanabe, Integrating geographical information systems, remote sensing, and machine learning techniques to monitor urban expansion: an application to luanda, angola, *Geo- Spat. Inf. Sci.* 26 (3) (2023) 446–464, <http://dx.doi.org/10.1080/10095020.2022.2066574>.
- [26] P. Lemenkova, O. Debeir, Computing vegetation indices from the satellite images using GRASS gis scripts for monitoring mangrove forests in the coastal landscapes of Niger Delta, Nigeria, *J. Mar. Sci. Eng.* 11 (2023) 871, <http://dx.doi.org/10.3390/jmse11040871>.
- [27] P. Lukas, A.M. Melesse, T.T. Kenea, Prediction of future land use/land cover changes using a coupled CA-ANN model in the upper omo-gibe river Basin, Ethiopia, *Remote. Sens.* 15 (4) (2023) <http://dx.doi.org/10.3390/rs15041148>.
- [28] L. Duarte, A. Teodoro, O. Moutinho, J. Gonçalves, Open-source GIS application for UAV photogrammetry based on MicMac, *Int. J. Remote Sens.* 38 (8–10) (2017) 3181–3202, <http://dx.doi.org/10.1080/01431161.2016.1259685>.
- [29] R.B. Lyngdoh, A.S. Sahadevan, T. Ahmad, P.S. Rathore, M. Mishra, P.K. Gupta, A. Misra, AVHYAS: A Free and Open Source QGIS Plugin for Advanced Hyperspectral Image Analysis, in: *2021 International Conference on Emerging Techniques in Computational Intelligence, ICETCI, 2021*, pp. 71–76, <http://dx.doi.org/10.1109/ICETCI51973.2021.9574057>.
- [30] A. Cuartero, M.E. Paoletti, P. García-Rodríguez, J.M. Haut, QCircularStats: A QGIS-plugin for evaluation bidimensional data by circular statistics, *IEEE Access* 11 (2023) 47552–47562, <http://dx.doi.org/10.1109/ACCESS.2023.3272637>.
- [31] Y. Liu, M.C. Wimberly, J. Hu, On the construction of eastweb framework — A plug-in framework for processing earth observation data streams, in: *IEEE International Conference on Electro/Information Technology, 2014*, pp. 463–467, <http://dx.doi.org/10.1109/EIT.2014.6871809>.
- [32] X. Song, J. Liu, Scheduling geo-processing workflow applications with QoS, in: *2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing*, Vol. 1, 2008, pp. 460–463, <http://dx.doi.org/10.1109/ETTandGRS.2008.98>.
- [33] P. Lemenkova, O. Debeir, Satellite image processing by python and r using landsat 9 OLI/TIRS and SRTM DEM data on côte d'ivoire, west africa, *J. Imaging* 8 (2022) 317, <http://dx.doi.org/10.3390/jimaging8120317>.
- [34] Q. Chen, L. Wang, Z. Shang, MRGIS: A MapReduce-enabled high performance workflow system for GIS, in: *2008 IEEE Fourth International Conference on EScience, 2008*, pp. 646–651, <http://dx.doi.org/10.1109/eScience.2008.169>.

- [35] J. Zhao, Y. Wang, H. Zhang, Automated batch processing of mass remote sensing and geospatial data to meet the needs of end users, in: 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011, pp. 3464–3467, <http://dx.doi.org/10.1109/IGARSS.2011.6049966>.
- [36] P. Lemenkova, O. Debeir, Quantitative morphometric 3D terrain analysis of Japan using scripts of GMT and R, *Land* 12 (2023) 261, <http://dx.doi.org/10.3390/land12010261>.
- [37] M. Zhang, P. Yue, X. Guo, GIScript: Towards an interoperable geospatial scripting language for GIS programming, in: 2014 the Third International Conference on Agro-Geoinformatics, 2014, pp. 1–5, <http://dx.doi.org/10.1109/Agro-Geoinformatics.2014.6910592>.
- [38] M. Crivellaro, A. Vitti, G. Zolezzi, W. Bertoldi, Characterization of active riverbed spatiotemporal dynamics through the definition of a framework for remote sensing procedures, *Remote. Sens.* 16 (1) (2024) <http://dx.doi.org/10.3390/rs16010184>.
- [39] P. Lemenkova, O. Debeir, Multispectral satellite image analysis for computing vegetation indices by R in the Khartoum Region of Sudan, Northeast Africa, *J. Imaging* 9 (2023) 98, <http://dx.doi.org/10.3390/jimaging9050098>.
- [40] E. Abdali, M.J. Valadan Zoej, A. Taheri Dehkordi, E. Ghaderpour, A parallel-cascaded ensemble of machine learning models for crop type classification in google earth engine using multi-temporal sentinel-1/2 and landsat-8/9 remote sensing data, *Remote. Sens.* 16 (1) (2024) <http://dx.doi.org/10.3390/rs16010127>.
- [41] I. Tselka, S.E. Detsikas, G.P. Petropoulos, I.I. Demertzi, Chapter 7 - google earth engine and machine learning classifiers for obtaining burnt area cartography: a case study from a mediterranean setting, in: N. Stathopoulos, A. Tsatsaris, K. Kalogeropoulos (Eds.), *Geoinformatics for Geosciences*, in: Earth Observation, Elsevier, 2023, pp. 131–148, <http://dx.doi.org/10.1016/B978-0-323-98983-1.00008-9>.
- [42] P. Lemenkova, Sentinel-2 for high resolution mapping of slope-based vegetation indices using machine learning by SAGA GIS, *Transylv. Rev. Syst. Ecol. Res.* 22 (2020) 17–34, <http://dx.doi.org/10.2478/trser-2020-0015>.
- [43] M. Tarek, T. Sadek, G. Hayet, Flood-Prone Urban Area mapping using machine learning, a case study of M'sila city (Algeria), in: 2023 International Conference on Earth Observation and Geo-Spatial Information, ICEOGI, 2023, pp. 1–4, <http://dx.doi.org/10.1109/ICEOGI57454.2023.10292983>.
- [44] N. Wu, L.G.T. Crusiol, G. Liu, D. Wuyun, G. Han, Comparing the performance of machine learning algorithms for estimating aboveground biomass in typical steppe of northern China using sentinel imageries, *Ecol. Indic.* 154 (2023) 110723, <http://dx.doi.org/10.1016/j.ecolind.2023.110723>.
- [45] G. Delogu, E. Caputi, M. Perretta, M.N. Ripa, L. Boccia, Using PRISMA hyperspectral data for land cover classification with artificial intelligence support, *Sustainability* 15 (18) (2023) <http://dx.doi.org/10.3390/su151813786>, URL <https://www.mdpi.com/2071-1050/15/18/13786>.
- [46] P. Lemenkova, O. Debeir, R libraries for remote sensing data classification by k-means clustering and NDVI computation in Congo River Basin, DRC, *Appl. Sci.* 12 (2022) 12554, <http://dx.doi.org/10.3390/app122412554>.
- [47] L. Lv, W. Fang, L. Yuan, Y. Tian, Optimal ANN-snn conversion with group neurons, in: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2024, pp. 6475–6479, <http://dx.doi.org/10.1109/ICASSP48485.2024.10448202>.
- [48] Q.R. Adebawale, N. Faruk, K.S. Adewole, A. Abdulkarim, A.A. Oloyede, H. Chiroma, O.A. Sowande, A.D. Usman, I.-F.Y. Olayinka, A.Y. Kassim, L.S. Taura, Effect of training algorithms and network architecture on the performance of multi-band ANN-based path loss prediction model, in: 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON, 2022, pp. 1–5, <http://dx.doi.org/10.1109/NIGERCON54645.2022.9803057>.
- [49] K. Zun-You, A. Ru, L. Xiang-Juan, ANN based high spatial resolution remote sensing wetland classification, in: 2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science, DCABES, 2015, pp. 180–183, <http://dx.doi.org/10.1109/DCABES.2015.52>.
- [50] P. Lemenkova, Artificial neural networks for Mapping Coastal lagoon of chilika lake, India, using earth observation data, *J. Mar. Sci. Eng.* 12 (5) (2024) <http://dx.doi.org/10.3390/jmse12050709>.
- [51] R. Hedjam, A. Abdesselam, F. Melgani, Change detection from unlabeled remote sensing images using siamese ANN, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 1530–1533, <http://dx.doi.org/10.1109/IGARSS.2019.8898672>.
- [52] P. Lemenkova, Artificial intelligence for computational remote sensing: Quantifying patterns of land cover types around cheetham wetlands, port phillip bay, Australia, *J. Mar. Sci. Eng.* 12 (8) (2024) <http://dx.doi.org/10.3390/jmse12081279>.
- [53] I. Kotaridis, M. Lazaridou, Cnns in land cover mapping with remote sensing imagery: a review and meta-analysis, *Int. J. Remote Sens.* 44 (19) (2023) 5896–5935, <http://dx.doi.org/10.1080/01431161.2023.2255354>.
- [54] S.V. Razavi-Termeh, A. Sadeghi-Niaraki, S.-M. Choi, Gully erosion susceptibility mapping using artificial intelligence and statistical models, *Geomatics, Nat. Hazards Risk* 11 (1) (2020) 821–844, <http://dx.doi.org/10.1080/19475705.2020.1753824>.
- [55] H. Smida, M. Tarki, N. Gammoudi, L. Dassi, GIS-based multicriteria and artificial neural network (ANN) investigation for the assessment of groundwater vulnerability and pollution hazard in the braga shallow aquifer (central Tunisia): A critical review of generic and modified DRASTIC models, *J. Contam. Hydrol.* 259 (2023) 104245, <http://dx.doi.org/10.1016/j.jconhyd.2023.104245>.
- [56] M. Shao, Z. Han, J. Sun, H. Gao, S. Zhang, Y. Zhao, A novel framework for wave power plant site selection and wave forecasting based on GIS, MCDM, and ANN methods: A case study in hainan island, southern China, *Energy Convers. Manage.* 299 (2024) 117816, <http://dx.doi.org/10.1016/j.enconman.2023.117816>.
- [57] H.L. Vu, D. Bolingbroke, K.T.W. Ng, B. Fallah, Assessment of waste characteristics and their impact on GIS vehicle collection route optimization using ANN waste forecasts, *Waste Manage.* 88 (2019) 118–130, <http://dx.doi.org/10.1016/j.wasman.2019.03.037>.
- [58] E.P. dos Santos, D.D. da Silva, C.H. do Amaral, E.I. Fernandes-Filho, R.L.S. Dias, A machine learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from sentinel-1 and landsat 8, *Comput. Electron. Agric.* 194 (2022) 106753, <http://dx.doi.org/10.1016/j.compag.2022.106753>.
- [59] E.M. Rocha, J.L. Drewry, R.M. Willett, B.D. Luck, Assessing kernel processing score of harvested corn silage in real-time using image analysis and machine learning, *Comput. Electron. Agric.* 203 (2022) 107415, <http://dx.doi.org/10.1016/j.compag.2022.107415>.
- [60] S.D. Latif, N. Alyaa Binti Hazrin, C. Hoon Koo, J. Lin Ng, B. Chaplot, Y. Feng Huang, A. El-Shafie, A. Najah Ahmed, Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches, *Alex. Eng. J.* 82 (2023) 16–25, <http://dx.doi.org/10.1016/j.aej.2023.09.060>.
- [61] P. Xiong, L. Tong, K. Zhang, X. Shen, R. Battiston, D. Ouzounov, R. Iuppa, D. Crookes, C. Long, H. Zhou, Towards advancing the earthquake forecasting by machine learning of satellite data, *Sci. Total Environ.* 771 (2021) 145256, <http://dx.doi.org/10.1016/j.scitotenv.2021.145256>.
- [62] L. Wu, R. Liu, N. Ju, A. Zhang, J. Gou, G. He, Y. Lei, Landslide mapping based on a hybrid CNN-transformer network and deep transfer learning using remote sensing images with topographic and spectral features, *Int. J. Appl. Earth Obs. Geoinf.* 126 (2024) 103612, <http://dx.doi.org/10.1016/j.jag.2023.103612>.
- [63] W. Han, X. Zhang, Y. Wang, L. Wang, X. Huang, J. Li, S. Wang, W. Chen, X. Li, R. Feng, R. Fan, X. Zhang, Y. Wang, A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities, *ISPRS J. Photogramm. Remote Sens.* 202 (2023) 87–113, <http://dx.doi.org/10.1016/j.isprsjprs.2023.05.032>.
- [64] H.M. El-Sayed, A.R. Elgendy, Geospatial and geophysical insights for groundwater potential zones mapping and aquifer evaluation at wadi abu marzouk in el-nagila, Egypt, *J. Aquat. Res.* 50 (1) (2024) 23–35, <http://dx.doi.org/10.1016/j.ejar.2023.12.008>.
- [65] A. Rafik, Y. Ait Brahimi, A. Amazirh, M. Ouarani, B. Bargam, H. Ouatiki, Y. Bouslih, L. Bouchaou, A. Chehbouni, Groundwater level forecasting in a data-scarce region through remote sensing data downscaling, hydrological modeling, and machine learning: A case study from Morocco, *J. Hydrol. Reg. Stud.* 50 (2023) 101569, <http://dx.doi.org/10.1016/j.ejrh.2023.101569>.
- [66] M.H. Aouragh, S. Jilil, N. Essahlaoui, A. Essahlaoui, A. El Hmaidi, A. El Ouali, A. Mridekh, Remote sensing and GIS-based machine learning models for spatial gully erosion prediction: A case study of rdat watershed in sebou basin, Morocco, *Remote. Sens. Applications: Soc. Environ.* 30 (2023) 100939, <http://dx.doi.org/10.1016/j.rsase.2023.100939>.
- [67] S. Amani, H. Shafizadeh-Moghadam, A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data, *Agricult. Water. Manag.* 284 (2023) 108324, <http://dx.doi.org/10.1016/j.agwat.2023.108324>.
- [68] F. Erdem, B. Bayram, T. Bakirman, O.C. Bayrak, B. Akpinar, An ensemble deep learning based shoreline segmentation approach (WaterNet) from landsat 8 OLI images, *Adv. Space Res.* 67 (3) (2021) 964–974, <http://dx.doi.org/10.1016/j.asr.2020.10.043>.
- [69] F. Li, T. Yigitcanlar, M. Nepal, K. Nguyen, F. Dur, Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework, *Sustain. Cities Soc.* 96 (2023) 104653, <http://dx.doi.org/10.1016/j.scs.2023.104653>.
- [70] M. Neteler, M.H. Bowman, M. Landa, M. Metz, GRASS GIS: A multi-purpose open source GIS, *Environ. Model. Softw.* 31 (2012) 124–130, <http://dx.doi.org/10.1016/j.envsoft.2011.11.014>.
- [71] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [72] D. Rocchini, V. Petras, A. Petrasova, Y. Chemin, C. Ricotta, A. Frigeri, M. Landa, M. Marcantonio, L. Bastin, M. Metz, L. Delucchi, M. Neteler, Spatio-ecological complexity measures in GRASS GIS, *Comput. Geosci.* 104 (2017) 166–176, <http://dx.doi.org/10.1016/j.cageo.2016.05.006>.
- [73] K.C. Tan, H.S. Lim, M.Z.M. Jaffri, Comparison of neural network and maximum likelihood classifiers for land cover classification using landsat multispectral data, in: 2011 IEEE Conference on Open Systems, 2011, pp. 241–244, <http://dx.doi.org/10.1109/ICOS.2011.6079300>.

- [74] T.F. Chan, G.H. Golub, R.J. LeVeque, Updating formulae and a pairwise algorithm for computing sample variances, in: H. Caussinus, P. Ettinger, R. Tomassone (Eds.), *COMPSTAT 1982 5th Symposium Held At Toulouse 1982*, Physica-Verlag HD, Heidelberg, 1982, pp. 30–41.
- [75] D. Berrar, Bayes' theorem and naive Bayes classifier, in: S. Ranganathan, M. Gribskov, K. Nakai, C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology*, Academic Press, Oxford, 2019, pp. 403–412, <http://dx.doi.org/10.1016/B978-0-12-809633-8.20473-1>.
- [76] O. Peretz, M. Koren, O. Koren, Naive Bayes classifier – an ensemble procedure for recall and precision enrichment, *Eng. Appl. Artif. Intell.* 136 (2024) 108972, <http://dx.doi.org/10.1016/j.engappai.2024.108972>.
- [77] Y. Shi, Y. Yang, An algorithm for incremental tree-augmented naive Bayesian classifier learning, in: 2010 International Conference on Artificial Intelligence and Computational Intelligence, 1, 2010, pp. 524–527, <http://dx.doi.org/10.1109/AICI.2010.115>.
- [78] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, *Psychol Rev* 65 (1958) 386–408, <http://dx.doi.org/10.1037/h0042519>.
- [79] A.E. Al-Dousari, A. Mishra, S. Singh, Land use land cover change detection and urban sprawl prediction for Kuwait metropolitan region, using multi-layer perceptron neural networks (MLPNN), *Egypt. J. Remote. Sens. Space Sci.* 26 (2) (2023) 381–392, <http://dx.doi.org/10.1016/j.ejrs.2023.05.003>.
- [80] R. Sugiarto, A.H. Saputro, W. Handayani, Study on Venation Visualization System Using Hyperspectral Images and Multi-Layer Perceptron Classifier, in: 2019 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI, 2019, pp. 478–482, <http://dx.doi.org/10.1109/ISRITI48646.2019.9034557>.
- [81] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing* 70 (1) (2006) 489–501, <http://dx.doi.org/10.1016/j.neucom.2005.12.126>, Neural Networks.
- [82] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297, <http://dx.doi.org/10.1007/BF00994018>.
- [83] V.N. Vapnik, The support vector method, in: W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (Eds.), *Artificial Neural Networks — ICANN'97*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1997, pp. 261–271.
- [84] M.A. Chandra, S. Bedi, Survey on SVM and their application in image classification, *Int. J. Inf. Technol.* 13 (5) (2021) 1–11, <http://dx.doi.org/10.1007/s41870-017-0080-1>.
- [85] S. Manthira Moorthi, I. Misra, R. Kaur, N.P. Darji, R. Ramakrishnan, Kernel based learning approach for satellite image classification using support vector machine, in: 2011 IEEE Recent Advances in Intelligent Computational Systems, 2011, pp. 107–110, <http://dx.doi.org/10.1109/RAICS.2011.6069282>.
- [86] P. Lemenkova, Support vector machine algorithm for mapping land cover dynamics in Senegal, west africa, using earth observation data, *Earth* 5 (3) (2024) 420–462, <http://dx.doi.org/10.3390/earth5030024>.
- [87] T.K. Ho, Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, 1995, pp. 278–282, <http://dx.doi.org/10.1109/ICDAR.1995.598994>.
- [88] T.K. Ho, M. Basu, Complexity measures of supervised classification problems, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3) (2002) 289–300, <http://dx.doi.org/10.1109/34.990132>.
- [89] L. Dong, H. Du, F. Mao, N. Han, X. Li, G. Zhou, D. Zhu, J. Zheng, M. Zhang, L. Xing, T. Liu, Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical area for example, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 13 (2020) 113–128, <http://dx.doi.org/10.1109/JSTARS.2019.2953234>.
- [90] T. Bai, K. Sun, S. Deng, D. Li, W. Li, Y. Chen, Multi-scale hierarchical sampling change detection using random forest for high-resolution satellite imagery, *Int. J. Remote Sens.* 39 (21) (2018) 7523–7546, <http://dx.doi.org/10.1080/01431161.2018.1471542>.
- [91] P. Lemenkova, Random forest classifier algorithm of geographic resources analysis support system geographic information system for satellite image processing: Case study of bight of sofala, mozambique, *Coasts* 4 (1) (2024) 127–149, <http://dx.doi.org/10.3390/coasts4010008>.
- [92] M. Bessinger, M. Lück-Vogel, A. Skowno, F. Conrad, Landsat-8 based coastal ecosystem mapping in South Africa using random forest classification in google earth engine, *S. Afr. J. Bot.* 150 (2022) 928–939, <http://dx.doi.org/10.1016/j.sajb.2022.08.014>.
- [93] H. Zhang, The optimality of naive Bayes, in: The Florida AI Research Society, 2004, pp. 1–6, URL <https://api.semanticscholar.org/CorpusID:8891634>.