

Article

Not peer-reviewed version

Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal

[Orlando Bhungeni](#)^{*}, Ashadevi Ramjatan, [Michael Gebreslasie](#)

Posted Date: 10 April 2024

doi: 10.20944/preprints202404.0713.v1

Keywords: uMngeni River Catchment; Machine learning; LULC; Landsat 8; Remote sensing



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal

Orlando Bhungeni *, Ashadevi Ramjatan and Michael Gebreslasie

School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban, 4000, South Africa

* Correspondence: e-mail: 223146923@stu.ukzn.ac.za

Abstract: Analysis of land use/land cover (LULC) in the catchment areas is the first action toward safeguarding the freshwater resources. The LULC information in the watershed has gained popularity in the natural science field as it helps water resource managers and environmental health specialists develop natural resource conservation strategies based on available quantitative information. Thus, remote sensing is the cornerstone in addressing environmental-related issues at the catchment level. In this study, the performance of four machine learning algorithms (MLAs), such as Random Forests (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Naïve Bayes (NB) was investigated to classify the catchment into nine relevant classes of the undulating watershed landscape using Landsat 8 Operational Land Imager (L8-OLI) imagery. The assessment of the MLAs were based on the visual inspection of the analyst and the commonly used assessment metrics, such as user's accuracy (UA), producers' accuracy (PA), overall accuracy (OA), and kappa coefficient. The MLAs produced good results, where RF (OA= 97.02%, Kappa= 0.96), SVM (OA= 89.74 %, Kappa= 0.88), ANN (OA= 87%, Kappa= 0.86), and NB (OA= 68.64 Kappa= 0.58). The results show the outstanding performance of the RF model over SVM and ANN with a small margin. While NB yielded satisfactory results, which could be primarily influenced by its sensitivity to limited training samples. In contrast, the robust performance of RF could be due to an ability to classify high-dimensional data with limited training data.

Keywords: uMngeni River Catchment; Machine learning; LULC; Landsat 8; Remote sensing

1. Introduction

The catchment areas are important ecological components that provide a variety of ecosystem services that benefit society and biodiversity [1,2]. These include the availability of freshwater for consumption, support the societal needs such as food production, sport, and recreation, and providing a habitat for both [3–5]. Such essential services can be sustained when the biophysical environment of the catchment areas is sustainably managed. However, natural land cover and freshwater resources in the water catchment areas are threatened by Land use/Land cover (LULC) changes in response to societal demands [6]. It is essential to map and monitor the water catchment landscape dynamics to provide valuable information on the distribution of the land use activities in the catchment, which serve as the tool for initiatives whose mandates aim to strike a balance between developments and the natural environment in the catchment areas. In addition, this information will help water resource managers and other environmental health practitioners establish environmental management plans and policies based on available information [7,8]. Moreover, the availability of such information is important as it can be used to troubleshoot water quality issues and prompt the identification of land resource degradation [9].

Mapping the water catchment land use activities can be achieved by either traditional field observation or remote sensing methods. Remote sensing emerged as the convenient and cost-effective way to provide the spatiotemporal data required to analyze heterogeneous landscapes at a large scale [10]. In addition, the remote sensing data can be systematically stored, maintained, and openly shared to end users. A variety of remote sensing sensors (optical and active sensors) have been utilized to map the LULC in the catchment areas. [11] assessed the influence of the contextual data features and various dimensionality filters, including Discrete Wavelet Transform (DWT), in

mapping LULC using Headwall Hyperspec data in Bergama, Turkey. The random forest (RF) and support vector machine (SVM) were employed, and RF and DWT data obtained higher overall accuracy and kappa scores of 88.13% and 0.88, respectively. In comparison, SVM recorded lower accuracy than RF. [12] have used Environmental Mapping and Analysis Program (EnMAP) imagery data to map the LULC using pixel-based machine learning algorithms (MLAs), such as RF and SVM image classification. Their results showed that the SVM image classifier returned an overall accuracy of 92.6%, while the RF model obtained an overall accuracy of 88.1%. [13] used PRISMA hyperspectral data with machine learning methods—such as RF, artificial neural network (ANN), and convolutional neural network (CNN), to classify the land use activities in Naples, Italy. The CNN achieved the best results at 0.973% and 0.968 overall accuracy and Kappa, respectively. As the second-best model, ANN recorded 0.963% and 0.956 and RF showed the lowest classification accuracy at 0.887% and 0.867, respectively.

Other studies have demonstrated the potential of very high spatial resolution (VHR) multispectral data in providing the same service as hyperspectral data in different terrestrial settings and achieving higher classification accuracies. For instance, [14], utilized Pleiades-1 imagery to map the LULC in Côte d’Emeraude, France. Maximum likelihood (ML) and SVM algorithms were used to implement the classification. ML algorithm yielded a higher overall accuracy of 84.64% compared to SVM, which only produced a map with an overall accuracy of 76.13%. In Fayoum City, Egypt, [15] evaluated a variety of classifiers in mapping the city land cover classes using PlanetScope imagery. They found the ANN model more accurate with a 0.97 kappa compared to Naïve Bayes (NB), SVM, RF, and decision trees (DT), which recorded 0.93, 0.90, 0.86, and 0.87, respectively.

Some studies explored the capabilities of moderate to low-resolution sensors' capabilities in classifying spatial land use classes. For example, in a comparative study, [7] employed Landsat-8 (L8-OLI) data and SVM and RF algorithms to map Nkandla Forest cover classes in South Africa. L8-OLI was found efficient in mapping the forest cover, and SVM was found more accurate with overall accuracy and kappa records of 95.83 % and 0.94, compared to 95.24 0.93 achieved by RF, respectively. On the west coast of India, a study by [16] assessed the efficiency of the MLAs in mapping LULC using Sentinel-2 (S2) imagery. NB, Classification And Regression Trees (CART), gradient tree boost (GTB), and RF MLAs were assessed, and the models achieved an overall accuracy of 81.34%, 90.18%, 91.06%, and 92.14%, respectively. A study by [10] also used L8-OLI data to evaluate the capabilities of MLAs in land use analysis in Casablanca, Morocco. The RF, SVM, CART, and GTB algorithms were among the tested methods, and RF showed robust performance with an overall accuracy of 95.42%, compared to other classifiers with an overall accuracy of 83%, 91.50%, and 93.46%, respectively. A study by [17] compared SVM and RF algorithms using S2 datasets without and with spectral indices to analyze the LULC hilly terrain of Gopeshwar, India. The RF models provided more accurate results in both datasets, with overall accuracy and kappa of 86% and 0.83 for a dataset with only spectral bands and 88% and 0.85 for a dataset with incorporated spectral indices, respectively. While SVM reported the least accuracy, with overall accuracy and kappa of 82% and 0.79 for spectral bands and 87% and 0.83 for incorporated indices, respectively. In the Munneru River Basin, India, [18] explored the performance of the MLAs viz. SVM, RF, and CART are used to map the LULC based on L8-OLI and S2 data. They found S2 and RF to dominate their counterparts. The average overall accuracy for RF, SVM, and CART in L8-OLI was 94.85%, 90.88%, and 82.88%. In comparison, S2 average overall accuracy for RF, SVM, and CART was 95.84%, 93.65%, and 86.48%, respectively. [19] evaluated the MLAs in classifying the land use dynamics using L8-OLI data in the Lake Haramaya catchment, Ethiopia. Object-based image analysis (OBIA), SVM, RF, and ANN were tested, and the SVM dominated other classifiers with an overall accuracy of 94%, while RF and ANN achieved 92% and 89%, respectively. While OBIA reported the least overall accuracy at 75%. [20], compared the parametric and non-parametric classifiers in mapping LULC based on L8-OLI data in the Big Sunflower River watershed, United States of America (USA). RF, SVM, and the ML were factored in, and their results showed SVM showed a robust performance with an overall accuracy (kappa) of 93.5% (0.88), 88.8% (0.82), and RF with the least performance at 84.6%, (0.72).

The success of land cover mapping is dependent on the classification algorithm utilized, not only the remote sensing data. As discussed above, various machine-learning algorithms have been used in the remote sensing field to explain the distribution of spatial features in catchment areas, with varying degrees of accuracy reported. Their performance varies with the biophysical features of

the landscapes [16]. Hence, there is no generic approach for LULC classification for all terrestrial landscapes [19]. Therefore, it is vital to compare the MLA models to determine their suitability and sensitivity to different settings [19]. In view of that, if any, little has been done to assess the suitability of the MLAs in the analysis of LULC in the case of South African river catchment landscapes. The current study employed L8-OLI data to map the thematic land use classes of the uMngeni river catchment in KwaZulu Natal, South Africa. Further, the performance of four MLAs, such as NB, RF, SVM, and ANN, was assessed in the uMngeni River Catchment terrestrial landscape. The current study aims to (i) map the LULC that defines the uMngeni River catchment area and (ii) compare the performance of four non-parametric pixel-based MLAs in a supervised classification based on L8-OLI data. The study is expected to improve the understanding of LULC configuration over the uMngeni River catchment, a key element in building a LULC change and projection module within the catchment, to generate information for water resource management strategies.

2. Materials and Methods

The analysis of the LULC was carried out in a series of steps, as illustrated in Figure 1. The acquired data was preprocessed before implementing the machine learning algorithms, and the LULC was computed. The post-classification analysis was carried out, including accuracy assessment and computing variable importance.

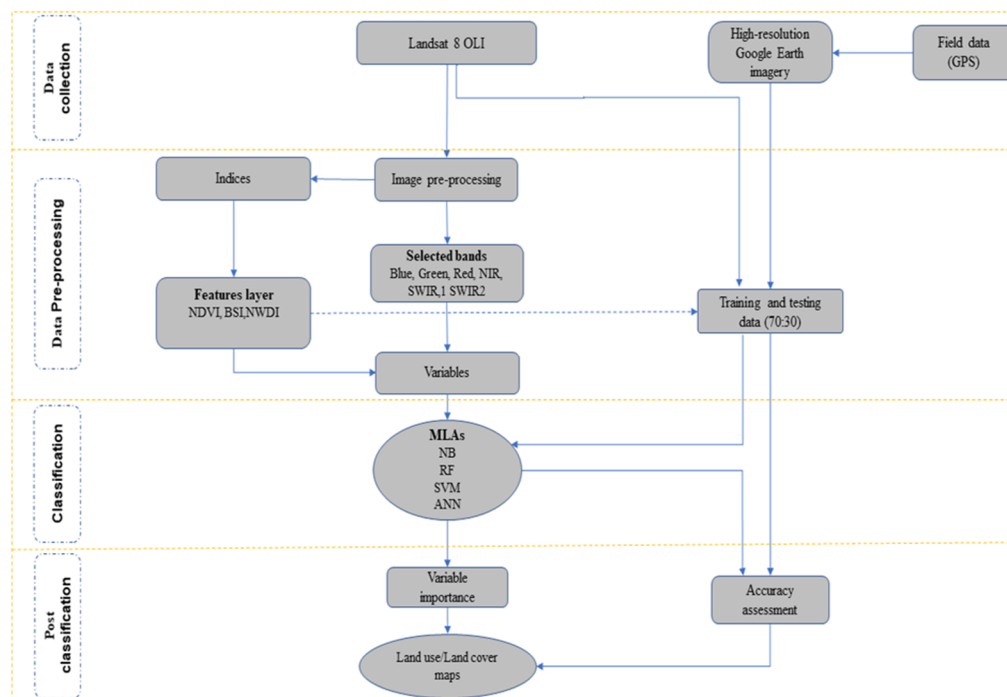


Figure 1. The conceptual framework of the study.

2.1. Study Area

The study area is located in the uMngeni River catchment, KwaZulu-Natal, South Africa (29°29'05.0"S 29°46'52.8"E and 29°48'41.8"S 31°02'27.0"E). The area is situated in the central-western part of KwaZulu-Natal province (Figure 2). It covers 4444 km², which is subdivided into three main regions (lower, middle, and upper) [21]. Its biophysical environment is characterized by a complex river network with an undulating landscape in the upper reaches and gentle slopes with hills towards the lower parts of the catchment, shown in green in Figure 2.

The river network forms the uMngeni river, which stretches from Drakensberg hills in the inner lands of KZN and flows down through Pietermaritzburg and final deposits in the Indian Ocean via the Durban north estuary. The uMngeni River has four main dams (Midmar, Albert Falls, Nagle, and Inanda), which serve as a water source for 45% of the KZN province's total population [9], including industrial activities, recreation, and agricultural purposes in the cities along the river path, such as

Pietermaritzburg and Durban [21]. Such services are largely influenced by the moderate climate conditions and slope gradient of the catchment.

The catchment has an elevation of 1913 m above sea level [22]. Rainfall occurs during the summer (October to March), ranging from 700 to 550 ml per annum. It has moderate temperatures with an annual average of 12 to 20 °C. Its climate and geographic position play a critical role in the province's economic development. Due to these biophysical characteristics, natural features, and economic activities, the catchment is characterized by a range of land use types.

From our field observation, the land cover types in the catchment include the following but are not limited to build-up: green patches (cropland, plantation, and forest), barren land, and water bodies. The Natural vegetation and cropland dominate the uMngeni catchment [9]. The plantation is one of the important land use classes in the catchment, which dominates the upper parts of the catchment. The middle part is characterized by heterogeneous patches of different land use types (plantation and built-up areas) [9]. While the lower part is mainly dominated by built-up patches with fragments of green vegetated land cover. It is worth unpacking the information in such biodiversity-rich areas to provide a piece of updated scientific information.

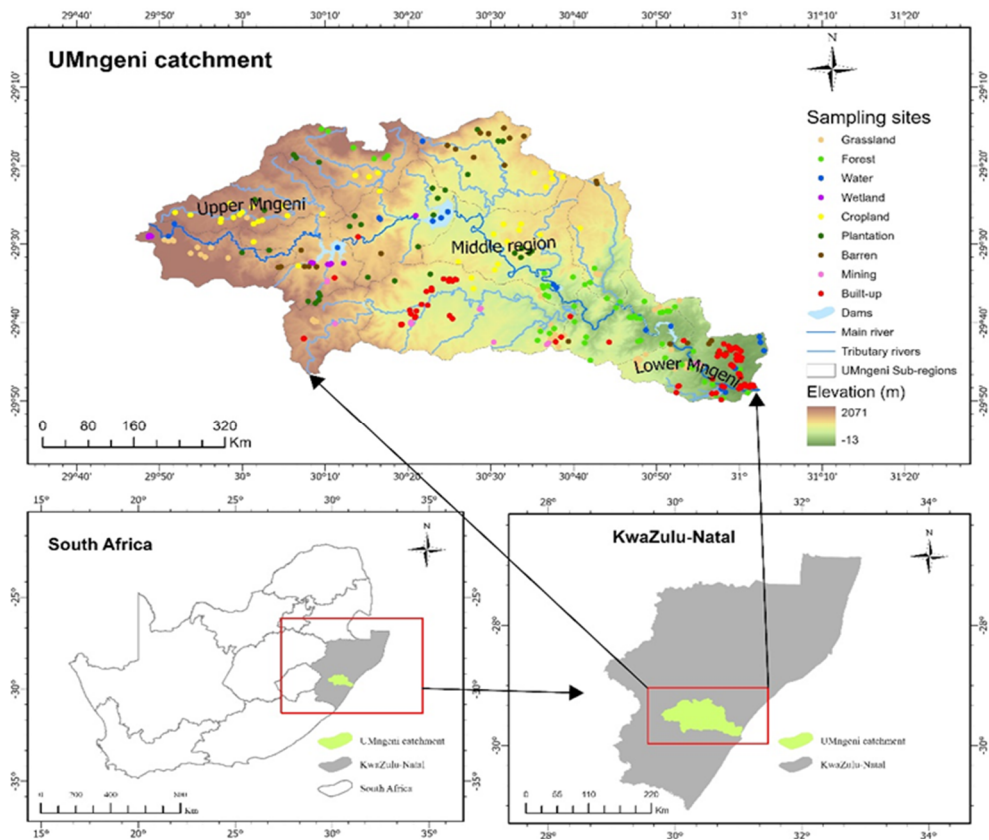


Figure 2. The uMngeni River catchment layout with elevation gradients, river network, and dams.

2.2. Definition of LULC Classes Umngeni River Catchment

To assess the LULC in the catchment, a total of nine thematic land use classes were identified, which include built-up, cropland, water, plantations, mining, forest, barren, grassland, and wetlands, as described in Table 1.

Table 1. Shows the description of nine thematic LULC classes used in the current study.

Class ID	LULC classes	Description
1	Grassland	Includes grass and shrubs open areas, as well as golf courses, and sport fields grounds
2	Forest	Natural forests or transition foarest areas ranging from low to highly dense canopy cover, dominated with trees of approximately 5 m high.
3	Water	This class considers natural or artificial surface fresh-waterbodies within the study area, which includes main rivers and tributaries, dams, lakes, and ponds
4	Wetland	This class is mainly made up of hydrphytes and herbaceous species which are solely based on the moisture of the soil to survive. The wetlands can be permanent or temporary depending on the local climate or its size and water holding capacity.
5	Cropland	Composed of both subsistance and ecommercial farming, which might be annual crops, or seasonal. During the post harvest, cultivated lands are charecterised as bare land in the case of seasonal crops.
6	Plantation	Range of tree species cultivated for commercial purposes, this class includes greener patches of the mature trees, sub adult, young, jivinales, and tree stumps.
7	Barren	Partial vegetated areas, bare land due to either erosion, natural degradation or human factors, post harvest crop fields, dry river banks, and stripped rock areas.
8	Mining	Small patches of mining scars, which comprise of open cast pits, sand mining, and the open raw material processing sites.
9	Built-up	Built-up structures, which includes all forms residential areas, economic corridors, industrial, commercial, educational, religious and health infrastructures.

2.3. Remote Sensing Data Acquisition and Preprocessing

In this study, a 30 m resolution image from L8-OLI image was used for LULC analysis (Table 2). The multispectral image was obtained on 15 May 2023 from the open Google Earth Engine (GEE) data catalog, in surface reflectance, level 2 category (<https://developers.google.com/earth-engine/datasets/catalog/landsat>). The surface reflectance category provides images that are ready for analysis with a set of preprocessing requirements that have already been performed to improve the quality of the image [23]. According to [24], surface reflectance images do not require fully blown image preprocessing. Hence the *ee.filter* functions were used to set the criteria to select data that meet the needs of the study. It must also be noted that scenes of different dates but the same season have been collected using *ee.Filter.date()*. As the images of the same season will not have an impact on the analysis, since the atmospheric variation is minimal. A total of four scenes (each with a total area of 185 km by 180 km) from two rows and paths were collected to cover our study area. Then, the *filterMetadata* function was set to return only scenes with a cloud cover of less than 10% [25].

Table 2. L8-OLI spectral bands details and their respective spatial resolutions.

Bands	Wavelength (λ)	Resolution (m)
Blue	0.45–0.51μm	30 m
Green	0.53–0.59 μm	30 m
Red	0.64–0.67 μm	30 m
Near Infrared (NIR)	0.85–0.87 μm	30 m
Shortwave Infrared 1 (SWIR 1)	1.56–1.65 μm	30 m
Shortwave Infrared 2 (SWIR 2)	2.10–2.29 μm	30 m

However, the filtering processing alone was infeasible to prepare the image for the analysis. Therefore, image pre-processing was performed as the prerequisite for any satellite image in the GEE interface. The median composite was formed from a series of scenes from the winter season, 01 May to 30 August 20 2022.

To stretch our specifications, the boundary shape file was used to create a subset of our study area from the composite image using the *Clip()* function. Then, the subset was subject to further processing, as the clouds are unavoidable in the satellite image [26]; the function of the mask from C code, commonly known as CFmasking, was applied to remove cloud-infested pixels in the image, as used by [27]. The approach involves the quality assurance (QA-PIXEL) band, which assigns zero values to the cloud-contaminated pixels and cloud shadows, producing a cloud-free composite [28].

2.3.1. Calculated Spectral Indices

Spectral indices are derived pixel-wise ratios of two or more bands of the multispectral image, which serve as both logistics and added explanatory variables for image classification [29]. Many studies have achieved meaningful LULC classification results by supplementing multispectral bands with information from spectral indices [18,30]. The indices layers facilitate the discrimination of specific features and assign the pixels to predicted classes. Our study area possesses diverse spatial attributes, making distinguishing the class members from others difficult. Hence, the current study employed various mathematical equations (Table 3) to compute the Normalized difference indices (NDI) for vegetation (NDVI) to improve the separation of vegetation classes [31], water indices (NDWI) to distinguish waterbodies from hill shadows [32], and bare soil indices (BSI) to identify bare surfaces [33]. These indices were also used as additional explanatory inputs to L8-OLI visible, NIR, and SWIR bands.

Table 3. shows the supplementary indices used in the land use classification.

Index	Equation	Rerferences
NDVI	$NDVI = \frac{NIR - Red}{Red + NIR}$	[34]
NDWI	$NDWI = \frac{Green - NIR}{Green + NIR}$	[32]
BSI	$BSI = \frac{(Red + SWIR) - (NIR + Blue)}{(Red + SWIR) + (NIR + Blue)}$	[35]

2.4. Reference Data Collection

Field survey approach and Google Earth Pro 7.3.6 high resolution images were used in collecting sampling data for building and measuring the mapping accuracy of the classifiers. The field survey was carried out on the 16th of August 2023, where coordinates of predefined thematic land use classes (Table 1) were extracted using a portable global positioning system (GPS) through a random sampling method. The same method was also use in Google Earth Pro-based sample collection. This approach widely places the sampling points across the study area (Figure 2) and ensures all the land cover classes are adequately represented in the reference data [36,37]. The GPS coordinates were imported into the Google Earth Pro 7.3.6 platform and overlaid on high-resolution image to locate the land cover classes [38]. Around each coordinate, the irregular-sized polygons were then digitized to represent the corresponding land cover class (Table 4). The digitized reference polygons were imported into GEE and overlayed to L8-OLI image. Then the spectral properties of cover classes were extracted from the L8-OLI image per pixels, as shown in Table 4 and randomly divided into two datasets, a training dataset (70%) and a validation data set (30%) [8]. These two datasets were used to train and assess the accuracy of the MLA models—NB, RF, SVM, and ANN.

Table 4. Training and test samples used in the study.

LULC Class	No. of polygons	Training pixel count	Test pixel count
Grassland	63	6944	1685
Forest	51	5823	1483
Water	38	1865	435
Wetland	11	357	86
Cropland	38	8965	2163
Plantation	32	7820	1957
Barren	30	4492	1075
Built-up	86	2504	598
Mining	13	667	159

2.5. Image Classification

Regions with diverse land use classes can make it difficult for machine learning models to discover the relationship between explanatory features [39]. Therefore, before commencing the

classification stage, it is important to scrutinize the variables based on their spectral similarities in the given class. To address that, the Pearson correlation coefficient was used to perform variable correlation analysis. It is a commonly used pair-wise technique that helps to identify the collinearity between two variables and eliminate noise [40]. It improves the accuracy of the MLAs by reducing redundancy and improving the computation processing rate. The stepwise correlation analysis was performed in L8-OLI explanatory variables; (1) spectral reflectance information for each variable was extracted from sampling sites, (2) then imported into Python to compute variable correlation using *corr()* algorithm (3) Seaborn's package was employed to plot the correlation array as a heat map matrix.

2.5.1. Selected Classifiers and Parameter Tuning Naïve Bayes

The NB classification approach evolved from the Bayesian and Naive theorem. Bayesian technique assumes the possibility of a recurrence of the event based on learning from the previous event [41]. While Naive is a probabilistic classifier with a strong independent assumption of the feature from the others [30]. In other words, dimensions that define a certain variable are independent of each other. Naïve Bayesian theory can be expressed mathematically as the probability of a pixel being in class k given X as training sample, which is equivalent to the density of the sample size given the class, multiplied by the probability of the pixel being in class k , divided by the density of sample size:

$$P(y = k|X) = \frac{p(X|y=k)(P(y=k))}{p(X)} \quad (1)$$

Where y is the unassigned variable, k is the class, and X is the training sample. Therefore $P(y = k|X)$ is the probability of the y and p is the density. NB models can be trained in supervised classification methods either in the form of parametric or non-parametric models by incorporating Gaussian functions.

The NB model is easy to tune, in GEE lambda (λ) is the only parameter that can be adjusted to avoid assigning zero probability to classes that could not be identified in the training stages [39]. The model computes the statistics (mean and the variance of the variables for each class) required for classification from the provided training data [16]. In LULC analysis, it has been reported as a promising classifier with satisfactory results [30]. Its performance is attributed to the independence of the feature dimensions from the others. For the purpose of our study, the NB model from “*smileNaiveBayes*” library in GEE was used.

Random Forests

The RF classifier is an ensemble classifier that uses regression decision trees to classify the dataset [42]. It gained popularity in the remote sensing community due to its reduced computational processing time, unique classification approach, and robust performance [29]. The RF proposed by [42], involves bagging CART split criterion. Bagging facilitates the building of the bootstrap samples from the original training sample data, the process also known as bootstrapping, predictor trees are formed from each bootstrap sample. The class membership of the unknown pixel is determined by the votes from all decision trees [8]. The decision to assign the pixels to the class is based on the aggregate (average of the votes) results from multiple RF decision trees [43,44]. The advantage of the RF models is the ability to handle high-dimensional datasets [16] and it is insensitive to outliers and overfitting [45]. The RF models left out some samples for each newly generated tree, commonly known as the out-of-bag samples (OOB), which report on the important features in the classification and misclassification errors of each tree [46]. For the model to meet analysts' needs, requirements must be predefined in the parameter tuning stages.

Two important parameters to be set when building the model are the number of decision trees to be created ($Ntree$) and the number of variables to be selected per split ($Mtry$). The higher value of $Ntree$ does not improve the accuracy of the classifier; instead, it brings confusion. Fewer $Ntree$ increases the confidence in features used to build the decision trees and reduces the correlation between the electorates [45]. In contrast, $Ntree$ can be as big as possible, while $Mtry$ should be equal to the square root of the number of features [47]. There is no standardized way to adjust the parameters for RF models. Therefore, it depends on the discretion of the analyst and the nature of the study. In this study, the RF classifier from the “*smileRandomForest*” library was utilized. To optimize

the model, Ntree was adjusted to 25, *Mtry* was set to three, which is equal to the square root of explanatory features, and each node was adjusted to have 2 leaves, as detailed in Table 5. Each decision tree received a share of 0.5 % from the original sample data for OOB, nodes per tree were set not to exceed 1000. The optimization of the parameters was obtained using 10-fold cross-validation.

Support Vector Machine

The SVM algorithm is a supervised vector-based classifier that has been used for both regression and classification [48]. Different types of SVM models have been widely used for different remote sensing applications which include LULC, water quality, and vegetation mapping. SVM models use discriminative methods by learning the distribution patterns of training data. It uses linear or nonlinear lines (known as hyperplanes) to optimize the boundaries between the vector data. The development of SVM models was designed as a binary classifier that only deal with linear data [49]. The method used the principle of finding the best marginal line (hyperplane) between the two features in the high-dimensional space, then the optimal hyperplane serves as the decision boundary that separates the two features. This principle generally works well when applied to linear data.

However, it is challenging to separate nonlinear data using the linear SVM approach. Because the inherent design of its model tends to not completely separate the training dataset, leading to classification errors [50]. After numeracy attempts to develop SVMs suitable for classifying nonlinear data, the joint effort from SVMs and kernel functions was the viable way to separate the nonlinear data. The primary role of the kernels function in SVM models is to mobilize the training data to a new dimensional space, allowing the marginal lines to find the optimal hyperplane to separate the vectors [51]. The most popular kernel functions for the classification of complex remote sensed data include the radial basis function (RBF) and polynomial kernel [52]. These two kernels have been reported in the literature for their ability to produce accurate classification results; for example [53], used RBF-SVM to monitor dynamic land use changes in Urmia Lake Iran, as did [54] in urban expansion mapping using a polynomial kernel function. [51] compared the performance of the kernels, and they found RBF to be slightly more accurate than the polynomial kernel in delineating the sub-catchment.

As there is no rule of thumb in kernel selection, the analyst’s discretion is used based on the nature of the study. Therefore, for the current study, the polynomial kernel was selected based on trying an error approach. The applicable SVM type for polynomial kernel in remote sensing data analysis is support vector classifier (SVC) which allows the cost (C) of non-separable vectors (C-SVC). Thus, it is also known as soft margin because it requires only the best hyperplane and accepts some separation errors as defined in the C threshold; if value C is high, it can lead to overlifting. In this way, C-SVC guarantees the results of which their integrity is highly dependent on the training data and parameter tuning of the model (Table 5). The C-SVC model has been used in different studies for LULC classification and achieved meaningful results, such as [18].

The C-SVC from the “*libsvm*” library in GEE was used. The kernel and C were set to polynomial and 10 respectively, as the main parameters to control marginal lines and error threshold. Additionally, hyperparameters were also tuned includes a degree, gamma, and coef0 to achieve the optimum hyperplane, as shown in Table 5.

Table 5. Hyperparameter tuning of the MLA models.

Classifier	Classifiers	Parameters	Parameter adjustments
NB	<i>ee.Classifier.smileNaiveBayes()</i>	λ	0.000001
RF	<i>ee.Classifier.smileRandomForest()</i>	Ntree	25.00
		Mtry	null
		MinLeafPopulation	2.00
		MaxNodes	1000.00
		BagFraction	0.50%
SVM	<i>ee.Classifier.libsvm()</i>	kernelType	POLY
		svmType	C_SVC
		coef0	0.3
		degree	1.00
		cost	10
ANN	<i>MLPClassifier (ANN)</i>	activation	RELU
		hidden_layer	1.00
		nuerons	10

Artificial Neural Network

ANN is a non-parametric algorithm with a complex network of interconnected neurons. The network is composed of an input phase, one or more intermediate phases, and an output phase. In remote sensing applications, the input layer has neurons that serve as receptor sites for data from each variable. They pass the data to the intermediate phase via channels. Each channel has a numeric value assigned, known as weight, which is the factor of the inputs to be transmitted to the neurons of the intermediate phase [55]. Neurons in the intermediate phase house the numeric values called bias, which are added to the inputs. Then, the total neuron values are subjected to the activation function, which determines the neurons that are going to contribute to the inputs of the next phase [56]; all the activated neurons transmit the inputs to the next phase; the process is called forward propagation. The inputs propagated through the network to the output phase, where the inputs are assigned to a corresponding class [57]. In addition, the misclassified data is backpropagated through the network model iteratively for revision by adjusting the weights. In this study, the feed-forward propagation ANN model was used, as defined:

$$\text{Output} = aS(\sum_{i=1}^n w_i w z_i + b) \quad (2)$$

Where w_i and z_i are weights that are carried by the channels, b is the bias as the remainder of the training samples, i is the number of neurons in the intermediate phase.

According to [58], ANN models, in general, tend to do data lifting. The network size and the number of neurons to be used are very important parameters to consider to achieve the best results; the large complex network can lead to over lifting of the input data [59]. The number of intermediates was set to 1 phase with 10 iterations. The nonlinearity was solved by the RELU activation function, as detailed parameter tuning is presented in Table 5.

2.6. Accuracy Assessment

The most used accuracy assessment standard for land use mapping is the Kappa coefficient [39]. The Kappa assesses the status of classification results against the randomly assigned values. The nature of the Kappa values ranges from -1 to 1 . These kappa values rank the classification accuracy of the classified image, Where the value is less than zero, no agreement, greater than ≤ 0.40 fair, ≤ 0.75 satisfactory, and ≤ 1 indicates good performance [60]. The Kappa can be defined as follows:

$$K = \frac{P(A) - P(E)}{1 - P(E)} \quad (3)$$

Where $P(A)$ is the overall accuracy and $P(E)$ chance agreement, which is given by the sum of the predicted values and actual values for each class an agreement by chance [61].

Note: j = land use class of interest (e.g. water bodies), will be used as an example to determine the accuracies of the class of interest.

The overall accuracy represents the sum of correctly labelled pixels for all the classes in the input image as per provided test sample sites and is mathematically expressed as a fraction of correctly labelled pixels and the total number of the test sample sites [62].

$$OA = \frac{\sum_{\text{correctly labelled pixels}}}{\sum_{\text{pixels in test sample sites}}} \quad (4)$$

PA is the number of pixels that the test sample of the given class (j) agreed to labelled pixels as class j in the input image. It is given by the sum of correctly labelled pixels of a class j , divided by the sum of correctly labelled pixels and the pixels claimed to not belong to class j , as adapted from [39]:

$$PA_j = \frac{\sum_{\text{correctly labelled pixels}_j}}{\sum_{\text{correct labelled pixels}_j + \text{mislabelled as class}_j}} \quad (5)$$

The user's accuracy is determined by the number of correctly labelled pixels of class divided by the sum of the pixels class claimed to belong to class j in the test samples, as proposed by [63].

$$UA_j = \frac{\sum_{\text{correctly labelled pixels}_j}}{\sum_{\text{correct labelled pixels}_j + \text{labelled as class}_j}} \quad (6)$$

As an effort to assess the performance of NB, RF, SVM, and ANN models, the aforementioned accuracy metrics will be calculated from the confusion matrix of each model, as presented in Figure 7.

3. Results

Figure 3 illustrates the result of the L8-OLI variable collinearity test, as discussed in section 2.3. One variable of the highly correlated pairs was removed. For example, ultra-blue and blue bands show a linear spectral relationship of 0.98, as shown in red to light red cells in Figure 3. In contrast variables showed the least correlation or distinct spectral variation were retained to facilitate the classification process. As a result, in a total of eleven variables tested, six multi-spectral bands such as visible (blue, green, and red), NIR, and SWIR (1 and 2) bands were selected as explanatory variables for the LULC classification along with three spectral indices (NDVI, NWDI, and BSI). To demonstrate the unique spectral information of each variable per class, the spectral variability was computed, as illustrated in Figure 4.

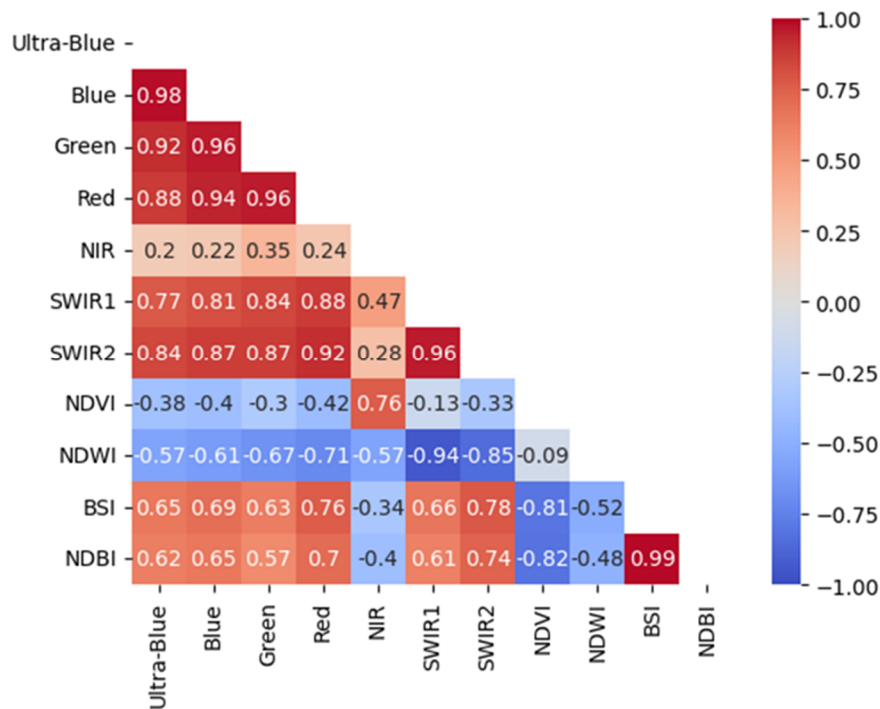


Figure 3. The variable collinearity heatmap. Blue to red shows increasing correlation relationship and vice versa.

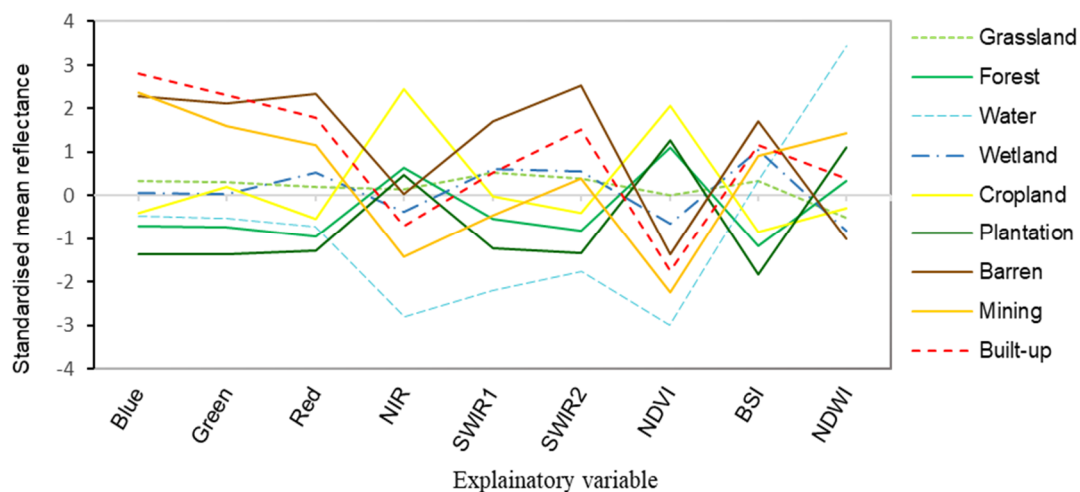


Figure 4. Standardised average reflectance values of each land use class.

3.1. Mapping and Spatial Extent LULC Classes

This study successfully mapped nine (9) LULC classes covering about 4444 km² of the uMngeni River catchment using four MLA strategies (NB, RF, SVM, and ANN). The classification results are shown in Figure 5 and graphical representation in Figure 6. The MLAs show estimates of varying extent but agree to class geographic position in the study area. The catchment is dominated by grassland vegetation, covering of 1819.31 km² being RF and 1206.16 km² ANN. Natural Forest is the second dominant class, with an area cover of 1200.13 km² determined by NB and 762.2 km² ANN. This is followed by planted forest, with an estimate of 738.9 km² and to 296.5 km² by NB.

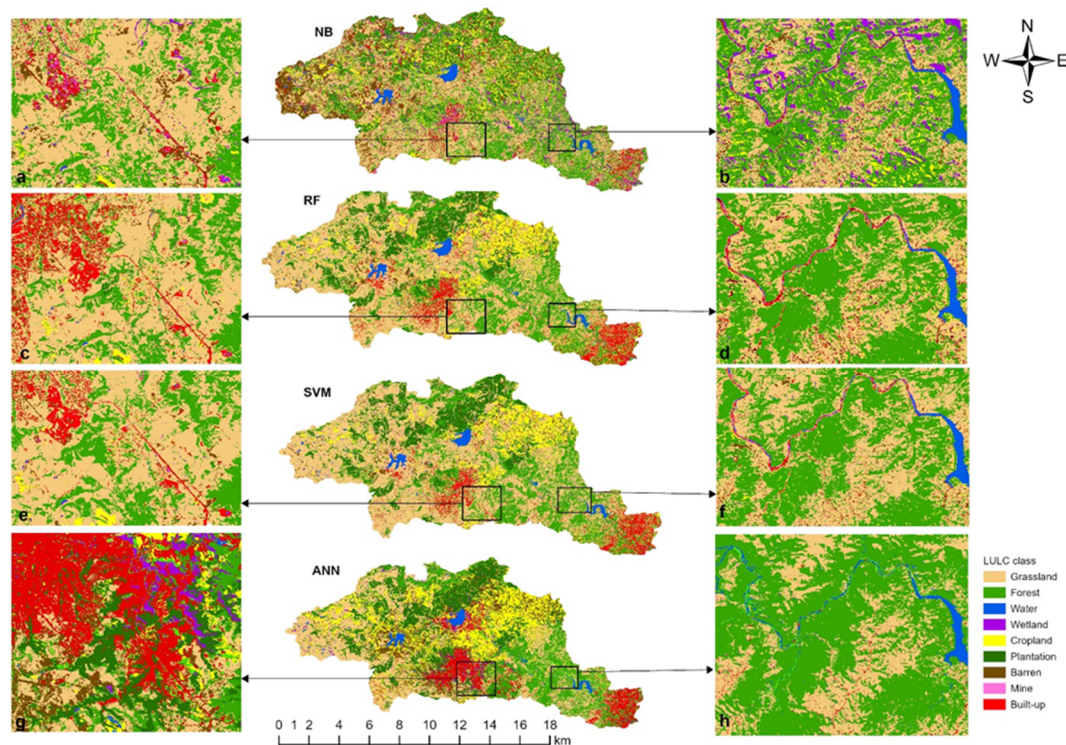


Figure 5. UMngeni River catchment LULC maps produced by NB, RF, SVM, AND ANN models with detailed small portions in (a) to (h).

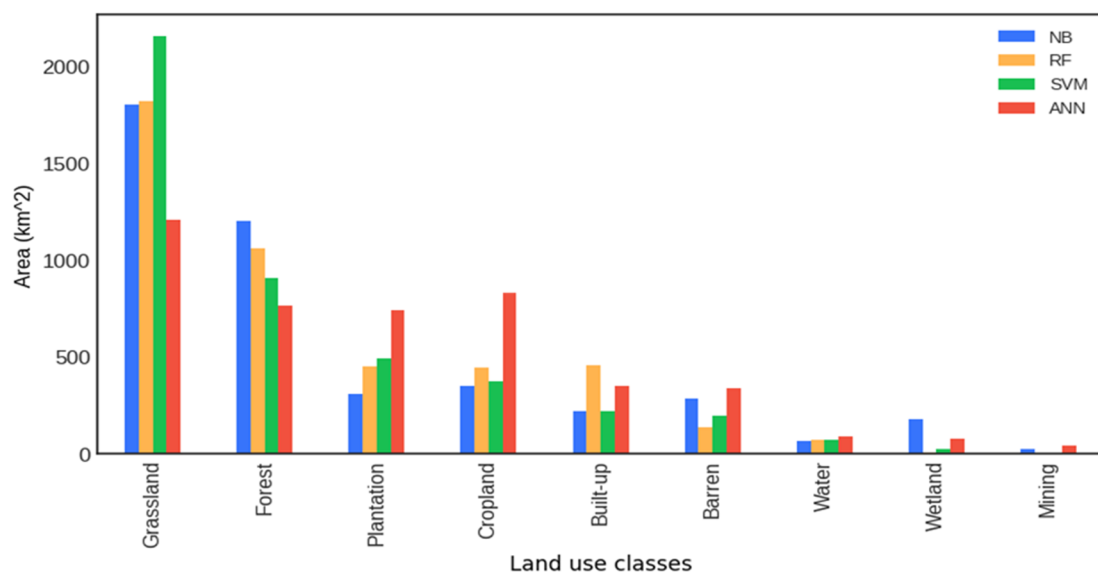


Figure 6. The spatial distribution of land use classes produced by each MLA.

Plantation, as the third dominant class, shows slightly variations of area cover in NB, SVM, RF (Figure 6). Except distinct area coverage of 738.88 km^2 recorded by ANN model. ANN showed the significant areas extent of 831.27 km^2 in the cropland class, in contrast, RF and SVM reported the statistical insignificant for cropland class cover at 372.73 km^2 and 440.29 km^2 , respectively. Built-up and Bare soil area cover vary across the four MLAs, ranking them in the fifth and sixth dominant classes, respectively. The Water, wetland, and mining classes showed the least area coverage. In this category, the mining class represents the least coverage of the total study area, with area cover at a maximum of 45.04 km^2 determined by ANN (Figure 6).

Despite the agreement on the trend of the LUL class area extent, it was challenging to discriminate between the tiny built-up class members in the matrix of grassland or trees and bare soil areas. In addition, confusion between built-up and mining was noticed. Some mining pixels were

classified as built-up or the other way around. Similarly, pixels of the mountain bare rock areas were claimed to either belong to mining or built-up class. Furthermore, a spectrum of pixels forming the irregular lines between the edges of different land cover parcels (e.g. cropland and forest) was also observed. The narrow river paths were represented by a variety of pixels from other classes, which were dominated by the built-up class as shown in Figure 5. (b) and (d). It is also important to report the error of omission of some of the small attributes which include barren, wetlands, and ponds at the sub-pixel level.

The geographic positions of thematic land use classes are consistent across four MLAs (Figure 5). In the matrix of grassland, as it dominates the catchment by 40.99%, the cropland and barren land are observed in the proximal distance from the water bodies and in the periphery of the upper parts of the catchment. Meanwhile, natural forest land cover class cover a large portion of the middle reaches extending to the lower regions, with about 0.08 % of mining classes on the forest banks. The plantation is scattered around the upper region with higher density in the far north of the upper region. The distribution of 10.27% (Table 6) of the built-up class is observed mainly in the lower region of the study area with dense areas towards the uMgeni river estuary (Figure 6). The wetland class covered tiny patches (0.12%) in the dam shorelines and visible at the tip of the upper region.

Table 6. Thematic land use classes area cover report by RF¹.

LULC CLASS	Area in km ²	Area in %
Grassland	1819.31	40.99
Forest	1058.55	23.85
Built-up	455.69	10.27
Cropland	446.49	10.06
Plantation	440.29	9.92
Barren	134.68	3.03
Water	74.81	1.69
Wetland	5.29	0.12
Mining	3.35	0.08
Total	4438.46	100.00

¹ These are the results (km² and percentage) for class area cover that was used in the discussion section since RF produced higher accurate maps than SVM, NB, and ANN.

3.2. Comparison of Machine Learning Algorithms Mapping Accuracy

The accuracy matrices of the models were generated using 30% of the sample size (Table 4). The matrices of the MLAs (Figure 7) provide the statistical report on the agreement between reality on the ground and maps produced. The OA, Kappa, UA, and PA were used to assess the accuracy of the models. According to the confusion matrix reports, the OA and Kappa were 64.38% and 0.58 for NB, 97.02% and 0.96 for RF, 89.74% and 0.88 for SVM, and 87% and 0.86 for the ANN, respectively. The overall performance of the models was generated from the accuracy of the individual classes.

The classification accuracy for the land use classes varies significantly across the four MLAs as presented in Figure 7. RF and SVM have well-classified the built-up class at the threshold of >85.60 % for both UA and PA. The NB model struggled to separate built-up and mining classes. Hence, there is observable swapping between the two classes. Similarly, the ANN model showed a high commission mining pixel to built-up. Thus, NB only achieved 48.54% and 49.39% for UA and PA for built-up, respectively. NB tends to confuse cropland with plantations, in contrast, RF, SVM, and ANN successfully extracted the cropland pixels with high percentages of metrics (Figure 7). All the MLAs delineated the surface water, plantation, forest, and barren with fairly high confidence. Except NB model that commissioned some natural forest pixels to plantations class.

It is worth reporting that only the RF model provided meaningful results for mapping uMngeni river catchment classes with a small area extent, which includes barren, mining, and wetland classes. The SVM model experienced the error of omission in mining and wetland pixels. In contrast, ANN model mapped mining and wetlands classes with observable overfitting which highly correlates with the decline of buildings, grassland, and forest classes (Figure 5). NB significantly overlifted the barren and wetland classes at the expense of cropland and plantation land cover classes. Consequently, these accuracy metrics form the basis of the overall classification performance of each model.

NB													RF												
Reference data													Reference data												
	1	2	3	4	5	6	7	8	9	Sum row	PA%		1	2	3	4	5	6	7	8	9	Sum row	PA%		
Predicted classes	1	1797	140	0	14	50	0	328	97	157	2583	69.57	1	2554	9	0	0	7	0	13	0	0	2583	98.88	
	2	50	1132	0	409	582	61	0	0	0	2234	50.67	2	22	2093	0	0	98	21	0	0	0	2234	93.69	
	3	0	0	694	2	0	1	0	0	0	697	99.57	3	0	0	696	0	1	0	0	0	0	697	99.86	
	4	8	9	1	103	0	0	2	8	0	131	78.63	4	8	1	0	121	1	0	0	0	0	131	92.37	
	5	93	342	0	6	2429	400	11	1	2	3284	73.96	5	60	10	0	0	3210	0	4	0	0	3284	97.75	
	6	18	629	0	3	595	1585	27	0	2	2859	55.44	6	11	37	1	0	1	2801	8	0	0	2859	97.97	
	7	244	11	0	0	0	0	1171	3	239	1668	70.20	7	64	0	0	0	1	0	1597	0	6	1662	96.09	
	8	5	0	3	73	0	0	9	71	76	237	29.96	8	5	0	3	4	0	0	1	133	91	237	56.12	
	9	56	3	2	85	0	0	62	252	449	909	49.39	9	4	0	0	0	0	0	7	0	898	909	98.79	
Sum column	2271	2266	700	695	3656	2047	1610	432	925			Sum column	2728	2150	700	125	3319	2822	1630	133	995				
UA 79.13 49.96 99.14 14.82 66.44 77.43 72.73 16.44 48.54													UA % 93.62 97.35 99.43 96.80 96.72 99.26 97.98 100.00 90.25												
OA% 64.10													OA% 85.00												
Kappa 0.58													Kappa 0.82												
SVM													ANN												
Reference data													Reference data												
	1	2	3	4	5	6	7	8	9	Sum row	PA%		1	2	3	4	5	6	7	8	9	Sum row	PA%		
Predicted classes	1	2330	54	0	3	24	3	156	11	2	2583	90.21	1	2100	111	0	8	285	2	2	68	7	2583	81.30	
	2	75	1932	0	0	186	38	2	0	1	2234	86.48	2	119	190	0	2	37	166	0	10	0	524	36.26	
	3	0	0	697	0	0	0	0	0	0	697	100.00	3	0	0	685	0	0	0	0	11	1	697	98.28	
	4	19	1	3	96	1	0	0	11	0	131	73.28	4	24	6	0	1011	0	0	0	0	0	1041	97.12	
	5	134	52	0	6	3080	0	11	0	1	3284	93.79	5	184	13	0	0	3051	0	36	0	0	3284	92.90	
	6	16	86	0	1	8	2737	11	0	0	2859	95.73	6	0	24	0	0	6	2820	9	0	0	2859	98.64	
	7	297	0	0	0	4	0	1289	0	78	1668	77.28	7	167	0	0	0	38	0	1258	0	205	1668	75.42	
	8	17	2	3	18	0	0	24	47	126	237	19.83	8	17	0	0	0	0	0	0	74	146	237	31.22	
	9	24	1	0	5	0	0	60	27	792	909	87.13	9	45	0	0	0	2	0	1	73	788	909	86.69	
Sum column	2912	2128	703	129	3303	2778	1553	96	1000			Sum column	2656	344	685	1021	3419	2988	1306	236	1147				
UA' 80.01 90.79 99.15 74.42 93.25 98.52 83.00 48.96 79.20													UA % 79.07 55.23 100.00 99.02 89.24 94.38 96.32 31.36 68.70												
OA% 89.18													OA% 87.50												
Kappa 0.87													Kappa 0.86												

Figure 7. Confusion matrix and accuracy metrics for NB, RF, SVM, and ANN.

3.3. Variable Importance of L8-OLI Explanatory Variables

The dimensional land cover classes on uMngeni river catchment mapping was achieved through contribution of nine different variables incorporated into the L8-OLI image. For such reason, it is worth to report the general contribution of each variable in the classification. Therefore, the variable importance was evaluated using the *explain()* function in GEE. As the results, the green and SWIR2 bands were reported to be the variables that contributed the most in labelling the classes, and the blue and green bands, in second place, contributed almost equally to accuracy and the classification of the data (Figure 8). With less variation, BSI, SWIR1, and NDWI were grouped as the third most important variables, while NIR and NDVI were determined to be the least important variables.

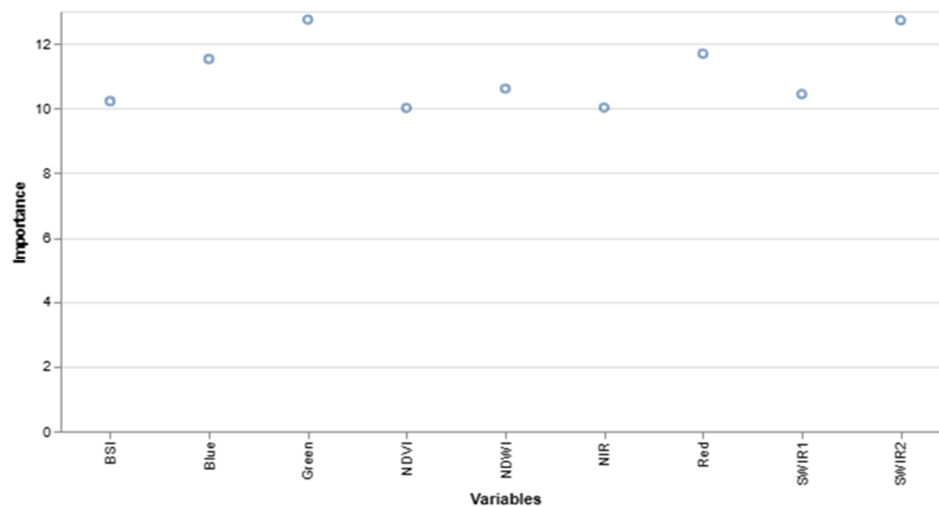


Figure 8. General variable importance.

4. Discussion

The four non-parametric MLAs (NB, RF, SVM, and ANN) were complemented in GEE to classify the nine land use classes. To unpack the status of the LULC and the spatial extent of individual classes in the uMngeni River catchment. The outcomes of this work will help to identify tools suitable for planning toward safeguarding freshwater resources. Also provides an insight into the application of MLAs and their respective abilities to explain the LULC in the undulating river catchment area.

The thematic classes were classified based on the use of the spectral properties from the training sites for each class. The grassland class was identified as the most dominant class. It is well distributed with an area extent of 40.99% of the total catchment area. The grassland class is made up of small shrubs, herbaceous, and different types of grass species. It forms the carpet-like layer on the surface of the soil with the roots embedded in the soil, which strengthens the topsoil from stripping during the fluvial period, increases the water infiltration rate, and plays a significant role in transpiration and photosynthesis. Thus, plays a crucial role in water quality and groundwater recharge.

The forest provides similar ecological functions as the grassland; it covers an average of 23.85% of the total catchment area, which makes it the second largest class after grassland. It is abundant in the middle regions, and partly scattered patches are observed in the lower and the upper regions. The areas covered by the forest class are characterized by woody plants for both open and closed canopies. Most importantly, it serves as carbon storage, indirectly saving water from degradation by mitigating climate change [7]. Therefore, the attempt to maintain the existing natural vegetation cover should be made and possibly promote initiatives that seek to expand the area cover of these classes. For example, the South African government targeting planting ten million trees in the next five years to rehabilitate the degraded green areas.

The cropland is the fifth largest class, which covers an area of 10.06% of the study area. The cropland includes commercial and subsistence farming. It is evenly distributed around the catchment, but homogeneous patches are observed in the north of the middle regions towards the tip of the upper regions (Figure 5). In the same region, about 9.92% of the plantation area cover was observed extending towards the south of the upper region. The plantations are characterized by planted wood trees for commercial reasons. The areas dominated by cropland fields and plantations are characterized by poor water quality due to the inputs applied to either maximize the yield or to control the insects [9]. The residuals tend to remain in the soil for a longer time as non-point pollutant. During rain periods pollutants get activated and splashed into the waterbodies via run-off and ultimately change the chemical composition of the water with implications for the water quality degradation. With the realization of the impact of agricultural inputs on waterbodies. The environmental managers will have to ensure control measures are in place to promote sustainable agriculture and timber.

Among the classes that account for the bigger share of the catchment area, the built-up class is the last at 10.27% distributed around the catchment, with dense patches in the lower regions and the southern part of the middle region. It is also necessary to highlight the visibility of the built-up class

around or along waterbodies. This class comprises a range of infrastructure types that serve different purposes such as residential areas, industrial, and economic corridors, towns, and public service infrastructures. The built-up class creates impervious surfaces with low water infiltration, increasing the water run-off, soil degradation, hiking of the non-point pollutants to the rivers, and ultimately increasing the stream rate flow with the implications for water quality degradation and hydrological structures. The increase of built-up structures goes together with the package of services such as wastewater drainage facilities which eventually discharge into the fresh waterbodies. Considering the detriment effects that rise from the built-up land use class, practical measures to control the urban sprawling must be developed and a balance between industrial activities and fresh waterbodies must be leveled.

Barren and mining pose similar threats to freshwater resources as built-up, they cover a relatively small portion of the catchment, 3.03%, and 0.08 % respectively, as shown in Table 6. Barren includes areas characterized by poor vegetation or completely bare soil. It is one of the main sources of water turbidity, it allows the topsoil particles to splash into the nearby rivers reducing the transparency of the water, which causes other aquatic species to struggle to survive. The mining close is the smallest land use class with snail trail-like structures and pits mining partly distributed randomly in the study area.

Water and wetland classes cover 1.69% and 0.12% of the uMngeni River catchment area, respectively. These two classes include dams, complex river networks, ponds, and perennial and non-perennial wetlands, member of the wetland class are likely found in the tip of the upper region and in the banks of the dams. They play an important role in promoting the functionality of ecological processes by naturally purifying the water providing the medium for aquatic organisms and supporting day-to-day human needs.

The quantitative analysis of the LULC in the catchment unpacked the scientific information and distribution of the land use types. This information will serve as a reference to the freshwater resource managers and environmental health specialists in developing management plans and policies for freshwater resources based on informed decisions.

It is crucial to assess the performance of the MLAs at the class level, as models produce different results in different classes [30]. In a view of that, only the ANN model confidently delineated the river path as part of the water class but could not pick up the small water bodies. In contrast, RF, SVM, and NB struggled to distinguish between small river paths and built-up class. The RF and SVM show a strong ability to map small water bodies. Similar results have been reported by [62], highlighting the efficiency of SVM in delineating small waterbodies. However, it performed poorly in the identification of the mining class members. NB and ANN on the other hand significantly demonstrated the swapping of the pixels between built-up and mining classes, while RF succeeded in separating these classes with a high user accuracy score. The spectral overlap among the set of classes that share similar optical properties and the sensitivity of the models to the training dataset can account for this error [29] found similar results, claiming that RF possesses a strong generalization technique for LULC analysis which makes it less sensitive to the small variations in input vector data. It is worth reporting the challenge encountered by all MLAs to delineate the wetland areas except the RF model. NB and ANN tend to overfit the wetland class, while SVM commissioned some of the wetland pixels to grassland. These findings were also observed in other studies, [20] suggested that this can be associated with composite signatures from the mixed pixels caused by the L8-OLI 30 m resolution and sparse vegetation within the wetland. In addition, this can also be influenced by the proportion of the wetland class and their relatively small training sample allocation. The accurate assessment at the class level provides the opportunity to put together the overall performance of the MLAs.

Overall, all the models performed well and produced more realistic LULC maps that resemble the reality on the ground. The findings of this study indicate that RF outperformed all the MLAs with slightly higher OA and Kappa values from those recorded by the SVM model as the second-best classifier, which produced higher accurate results compared to ANN and NB. The comparison of the RF and SVM has been debated in many previous studies. [18] found RF to be robust in assessing LULC using multiple MLAs, as did [64] in mapping the Cocco forest classes using various MLAs. They attributed the performance of the model to its less sensitivity to training data size and the advantage of the random multi-decision trees in generalization. In contrast, [7] found SVM to

outperform the RF class in mapping natural forest cover classes, highlighting that the performance of the model can be credited to its default parameters. [19] in evaluating MLAs for dynamic land use, found SVM to be superior to RF, they credited the performance of the models to the use of the kernel functions to optimize the hyperplanes to find the optimum margin to classify the complex land use classes.

An emerging MLA in the remote sensing field, Nave Bayes was found not suitable for analysis of the complex land use landscapes. Similar findings were also reported in other studies, [29] in the performance of multi-MLAs in multi-data sources, which found NB to be not suited for complex LULC analysis. A similar study by [30] also found similar results, highlighting the limitations in the inherent design of the model to deal with high dimensional data. [47], associated NB performance with the limitations of its model to limited training data. The NB classifier can be effective in large areas with more linear land cover pattern.

The results of our study proved the excellence of the RF model, which produced the higher accurate LULC analysis of the uMngeni River catchment. These results will fill the gap in the literature by identifying the suitable MLA for quantitative analysis of dimensional LULC in undulating watershed areas. In addition, this can be a useful tool for water resource managers and environmental health stakeholders to facilitate the development of freshwater resource management strategies by unpacking the visual and qualitative LULC information for the watershed areas.

The L8-OLI multi-spectral bands with incorporated indices provided a good performance in classification of uMngeni River Catchment. For such reasons, L8-OLI data was found to be reliable for analysis of the complex LULC in the undulating landscape of the catchment areas. However, it was found to be challenging to classify the pixels in the edges of the big parcels of class and build-up class members in the core of barren and/or grassland or partially masked by trees. These findings are consistent with findings reported by [20], who found that pixels in the transition zone between two parcels have been misclassified due to mixed-pixel composite signatures, as explained by [65,66]. [18] also reported that it is challenging to discriminate classes from others in L8-OLI 30 m resolution images due to mixed pixels. Other studies factored in the limitations of pixel-based algorithms, [25] associated mixed pixel effects with the restrictions of pixel-based models to go beyond the pixel level.

5. Conclusion

In nature, water catchment areas provide well-balanced ecosystems that promote a conducive environment for both flora and fauna to thrive. The uMngeni water catchment is one of the important watershed areas that provide such services in RSA. Remote sensing LULC mapping unpacked the visual and qualitative information that is necessary to facilitate the conservation of freshwater resources. The Landsat 8 OLI imagery provided reliable data for mapping the thematic land use class. The RF and SVM produced meaningful land cover maps with high accuracies of matrices. However, the RF model proved to be slightly superior to the SVM model. The high performance of these two models is facilitated by the hyperparameter tuning, their ability to handle huge complex data, and the ability to work in a limited training dataset. We conclude that RF in conjunction with Landsat 8 OLI data is the best MLA in mapping complex land use classes of undulating landscape of watershed areas.

Author Contributions: Conceptualization, O.B., A.R. and M.G.; Methodology, O.B., A.R. and M.G.; Formal analysis, O.B. and M.G.; Supervision, M.G.; Writing - original draft, O.B.; Writing - review & editing, O.B., A.R. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the South African National Space Agency (SANSA) (www.sansa.org.za).

Data Availability Statement: The surface reflectance Landsat 8 data can be accessed in Earth Engine data catalog at USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog | Google for Developers.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mekuria, W.; Gedle, A.; Tesfaye, Y.; Phimister, E. Implications of Changes in Land Use for Ecosystem Service Values of Two Highly Eroded Watersheds in Lake Abaya Chamo Sub-Basin, Ethiopia. *Ecosyst Serv* **2023**, *64*, doi:10.1016/j.ecoser.2023.101564.

2. Aneseyee, A.B.; Soromessa, T.; Elias, E. The Effect of Land Use/Land Cover Changes on Ecosystem Services Valuation of Winike Watershed, Omo Gibe Basin, Ethiopia. *Human and Ecological Risk Assessment* **2020**, *26*, 2608–2627, doi:10.1080/10807039.2019.1675139.
3. Pullanikkatil, D.; Palamuleni, L.G.; Ruhiiga, T.M. Land Use/Land Cover Change and Implications for Ecosystems Services in the Likangala River Catchment, Malawi. *Physics and Chemistry of the Earth* **2016**, *93*, 96–103, doi:10.1016/j.pce.2016.03.002.
4. Jourdain, D.; Namakando, N.; Mungatana, E.D.; Mirzabaev, A.; Njiraini, G. Revealing Salient Aquatic Ecosystem Services Bundles in the Olifants River Catchment, South Africa. *Wetl Ecol Manag* **2023**, *31*, 267–286, doi:10.1007/s11273-023-09915-6.
5. Lynch, A.J.; Cooke, S.J.; Arthington, A.H.; Baigun, C.; Bossenbroek, L.; Dickens, C.; Harrison, I.; Kimirei, I.; Langhans, S.D.; Murchie, K.J.; et al. People Need Freshwater Biodiversity. *Wiley Interdisciplinary Reviews: Water* **2023**, *10*, doi:10.1002/wat2.1633.
6. Nyathi, N.A.; Zhao, W.; Musakwa, W. Land Use Land Cover Changes and Their Impacts on Ecosystem Services in the Nzhelele River Catchment, South Africa. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *5*, 809–816.
7. Gyamfi-Ampadu, E.; Gebreslasie, M.; Mendoza-Ponce, A. Mapping Natural Forest Cover Using Satellite Imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sens Appl* **2020**, *18*, doi:10.1016/j.rsase.2020.100302.
8. Obaid, A.; Adam, E.; Ali, K.A. Land Use and Land Cover Change in the Vaal Dam Catchment, South Africa: A Study Based on Remote Sensing and Time Series Analysis. *Geomatics* **2023**, *3*, 205–220, doi:10.3390/geomatics3010011.
9. Namugize, J.N.; Jewitt, G.; Graham, M. Effects of Land Use and Land Cover Changes on Water Quality in the UMngeni River Catchment, South Africa. *Physics and Chemistry of the Earth* **2018**, *105*, 247–264, doi:10.1016/j.pce.2018.03.013.
10. Ouchra, H.; Belangour, A.; Erraissi, A. Comparison of Machine Learning Methods for Satellite Image Classification: A Case Study of Casablanca Using Landsat Imagery and Google Earth Engine. *Journal of Environmental & Earth Sciences* **2023**, *5*, 118–134, doi:10.30564/jees.v5i2.5928.
11. Akar, O.; Tunc Gormus, E. Land Use/Land Cover Mapping from Airborne Hyperspectral Images with Machine Learning Algorithms and Contextual Information. *Geocarto Int* **2022**, *37*, 3963–3990, doi:10.1080/10106049.2021.1945149.
12. Lekka, C.; Petropoulos, G.P.; Detsikas, S.E. Appraisal of EnMAP Hyperspectral Imagery Use in LULC Mapping When Combined with Machine Learning Pixel-Based Classifiers. *Environmental Modelling & Software* **2024**, *173*, 105956, doi:10.1016/j.envsoft.2024.105956.
13. Delogu, G.; Caputi, E.; Perretta, M.; Ripa, M.N.; Boccia, L. Using PRISMA Hyperspectral Data for Land Cover Classification with Artificial Intelligence Support. *Sustainability* **2023**, *15*, 13786, doi:10.3390/su151813786.
14. James, D.; Collin, A.; Mury, A.; Costa, S. Very High Resolution Land Use and Land Cover Mapping Using Pleiades-1 Stereo Imagery and Machine Learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2020**, *XLIII-B2-2020*, 675–682, doi:10.5194/isprs-archives-XLIII-B2-2020-675-2020.
15. Mahmoud, R.; Hassanin, M.; Al Feel, H.; Badry, R.M. Machine Learning-Based Land Use and Land Cover Mapping Using Multi-Spectral Satellite Imagery: A Case Study in Egypt. *Sustainability* **2023**, *15*, 9467, doi:10.3390/su15129467.
16. Bayas, S.; Sawant, S.; Dhondge, I.; Kankal, P.; Joshi, A. Land Use Land Cover Classification Using Different ML Algorithms on Sentinel-2 Imagery. In: 2022; pp. 761–777.
17. Parashar, D.; Kumar, A.; Palni, S.; Pandey, A.; Singh, A.; Singh, A.P. Use of Machine Learning-Based Classification Algorithms in the Monitoring of Land Use and Land Cover Practices in a Hilly Terrain. *Environ Monit Assess* **2024**, *196*, doi:10.1007/s10661-023-12131-7.
18. Loukika, K.N.; Keesara, V.R.; Sridhar, V. Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability* **2021**, *13*, 13758, doi:10.3390/su132413758.
19. Woldemariam, G.W.; Tibebe, D.; Mengesha, T.E.; Gelete, T.B. Machine-Learning Algorithms for Land Use Dynamics in Lake Haramaya Watershed, Ethiopia. *Model Earth Syst Environ* **2022**, *8*, 3719–3736, doi:10.1007/s40808-021-01296-0.
20. Dash, P.; Sanders, S.L.; Parajuli, P.; Ouyang, Y. Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data in an Agricultural Watershed. *Remote Sens (Basel)* **2023**, *15*, doi:10.3390/rs15164020.
21. Hughes, C.J.; De Winnaar, G.; Schulze, R.E.; Mander, M.; Jewitt, G.P.W. Mapping of Water-Related Ecosystem Services in the UMngeni Catchment Using a Daily Time-Step Hydrological Model for Prioritisation of Ecological Infrastructure Investment - Part 2: Outputs. *Water SA* **2018**, *44*, 590–600, doi:10.4314/wsa.v44i4.08.

22. Kusangaya, S.; Warburton, M.; Archer van Garderen, E. Use of ACRU, a Distributed Hydrological Model, to Evaluate How Errors from Downscaled Rainfall Are Propagated in Simulated Runoff in UMngeni Catchment, South Africa. *Hydrological Sciences Journal* **2017**, *62*, 1995–2011, doi:10.1080/02626667.2017.1349317.
23. Kaur, H.; Tyagi, S.; Mehta, M.; Singh, D. Time Series (2001/2002–2021) Analysis of Earth Observation Data Using Google Earth Engine (GEE) for Detecting Changes in Land Use Land Cover (LULC) with Specific Reference to Forest Cover in East Godavari Region, Andhra Pradesh, India. *Journal of Earth System Science* **2023**, *132*, doi:10.1007/s12040-023-02099-w.
24. Mandal, S.; Bandyopadhyay, A.; Bhadra, A. Dynamics and Future Prediction of LULC on Pare River Basin of Arunachal Pradesh Using Machine Learning Techniques. *Environ Monit Assess* **2023**, *195*, 709, doi:10.1007/s10661-023-11280-z.
25. Maviza, A.; Ahmed, F. Analysis of Past and Future Multi-Temporal Land Use and Land Cover Changes in the Semi-Arid Upper-Mzingwane Sub-Catchment in the Matabeleland South Province of Zimbabwe. *Int J Remote Sens* **2020**, *41*, 5206–5227, doi:10.1080/01431161.2020.1731001.
26. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens Environ* **2017**, *194*, 379–390, doi:10.1016/j.rse.2017.03.026.
27. Mashala, M.J.; Dube, T.; Ayisi, K.K.; Ramudzuli, M.R. Using the Google Earth Engine Cloud-Computing Platform to Assess the Long-Term Spatial Temporal Dynamics of Land Use and Land Cover within the Letaba Watershed, South Africa. *Geocarto Int* **2023**, *38*, doi:10.1080/10106049.2023.2252781.
28. Floreano, I.X.; de Moraes, L.A.F. Land Use/Land Cover (LULC) Analysis (2009–2019) with Google Earth Engine and 2030 Prediction Using Markov-CA in the Rondônia State, Brazil. *Environ Monit Assess* **2021**, *193*, 239.
29. Orieschnig, C.A.; Belaud, G.; Venot, J.-P.; Massuel, S.; Ogilvie, A. Input Imagery, Classifiers, and Cloud Computing: Insights from Multi-Temporal LULC Mapping in the Cambodian Mekong Delta. *Eur J Remote Sens* **2021**, *54*, 398–416, doi:10.1080/22797254.2021.1948356.
30. Palanisamy, P.A.; Jain, K.; Bonafoni, S. Machine Learning Classifier Evaluation for Different Input Combinations: A Case Study with Landsat 9 and Sentinel-2 Data. *Remote Sens (Basel)* **2023**, *15*, 3241, doi:10.3390/rs15133241.
31. Viana, C.M.; Girão, I.; Rocha, J. Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. *Remote Sens (Basel)* **2019**, *11*, 1104, doi:10.3390/rs11091104.
32. Gao, B.C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens Environ* **1996**, *58*, 257–266, doi:10.1016/S0034-4257(96)00067-3.
33. Nguyen, C.T.; Chidthaisong, A.; Kieu Diem, P.; Huo, L.-Z. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. *Land (Basel)* **2021**, *10*, 231, doi:10.3390/land10030231.
34. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. In Third ERTS Symposium; NASASP-351 I: Greenbelt, MD, USA, 1973; pp. 309–317.
35. Rikimaru, A.; Roy, P.S.; Miyatake, S. Tropical Forest Cover Density Mapping. *Tropical Ecology (India)* **2002**, *43*.
36. Carlotto, M.J. Effect of Errors in Ground Truth on Classification Accuracy. *Int J Remote Sens* **2009**, *30*, 4831–4849, doi:10.1080/01431160802672864.
37. Foody, G.M. Assessing the Accuracy of Land Cover Change with Imperfect Ground Reference Data. *Remote Sens Environ* **2010**, *114*, 2271–2285, doi:10.1016/j.rse.2010.05.003.
38. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S. V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens Environ* **2014**, *148*, 42–57.
39. Gandhi, Ujaval, 2021. End-to-End Google Earth Engine Course. Spatial Thoughts. <https://courses.spatialthoughts.com/end-to-end-gee.html>
40. Gyamfi-Ampadu, E.; Gebreslasie, M.; Mendoza-Ponce, A. Evaluating Multi-sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction. *Remote Sens (Basel)* **2021**, *13*, doi:10.3390/rs13051033.
41. Costache, R.; Tin, T.T.; Arabameri, A.; Crăciun, A.; Costache, I.; Islam, A.R.Md.T.; Sahana, M.; Pham, B.T. Stacking State-of-the-Art Ensemble for Flash-Flood Potential Assessment. *Geocarto Int* **2022**, *37*, 13812–13838, doi:10.1080/10106049.2022.2082558.
42. Breiman, L.E.O. Random Forests. **2001**, 5–32.
43. Zhang, Q.; Liu, M.; Zhang, Y.; Mao, D.; Li, F.; Wu, F.; Song, J.; Li, X.; Kou, C.; Li, C.; et al. Comparison of Machine Learning Methods for Predicting Soil Total Nitrogen Content Using Landsat-8, Sentinel-1, and Sentinel-2 Images. *Remote Sens (Basel)* **2023**, *15*, 2907, doi:10.3390/rs15112907.
44. Mather, P.M.; Koch, M. *Computer Processing of Remotely-Sensed Images: An Introduction*; John Wiley & Sons, 2011; ISBN 1119956404.

45. Cuypers, S.; Nascetti, A.; Vergauwen, M. Land Use and Land Cover Mapping with VHR and Multi-Temporal Sentinel-2 Imagery. *Remote Sens (Basel)* **2023**, *15*, 2501, doi:10.3390/rs15102501.
46. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.K.; Ha, S. Machine Learning Approaches to Coastal Water Quality Monitoring Using GOCI Satellite Data. *Glsci Remote Sens* **2014**, *51*, 158–174, doi:10.1080/15481603.2014.900983.
47. Camargo, F.F.; Sano, E.E.; Almeida, C.M.; Mura, J.C.; Almeida, T. A Comparative Assessment of Machine-Learning Techniques for Land Use and Land Cover Classification of the Brazilian Tropical Savanna Using ALOS-2/PALSAR-2 Polarimetric Images. *Remote Sens (Basel)* **2019**, *11*, doi:10.3390/rs11131600.
48. Yulianto, F.; Raharjo, P.D.; Pramono, I.B.; Setiawan, M.A.; Chulafak, G.A.; Nugroho, G.; Sakti, A.D.; Nugroho, S.; Budhiman, S. Prediction and Mapping of Land Degradation in the Batanghari Watershed, Sumatra, Indonesia: Utilizing Multi-Source Geospatial Data and Machine Learning Modeling Techniques. *Model Earth Syst Environ* **2023**, *9*, 4383–4404, doi:10.1007/s40808-023-01761-y.
49. Huang, C.; Davis, L.S.; Townshend, J.R.G. An Assessment of Support Vector Machines for Land Cover Classification. *Int J Remote Sens* **2002**, *23*, 725–749, doi:10.1080/01431160110040323.
50. Lantzanakis, G.; Mitra, Z.; Chrysoulakis, N. X-SVM: An Extension of C-SVM Algorithm for Classification of High-Resolution Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *59*, 3805–3815, doi:10.1109/TGRS.2020.3017937.
51. Srivastava, P.K.; Han, D.; Rico-Ramirez, M.A.; Bray, M.; Islam, T. Selection of Classification Techniques for Land Use/Land Cover Change Investigation. *Advances in Space Research* **2012**, *50*, 1250–1265, doi:10.1016/j.asr.2012.06.032.
52. Martins, S.; Bernardo, N.; Ogashawara, I.; Alcantara, E. Support Vector Machine Algorithm Optimal Parameterization for Change Detection Mapping in Funil Hydroelectric Reservoir (Rio de Janeiro State, Brazil). *Model Earth Syst Environ* **2016**, *2*, 138, doi:10.1007/s40808-016-0190-y.
53. Roushangar, K.; Aalami, M.T.; Golmohammadi, H.; Shahnazi, S. Monitoring and Prediction of Land Use/Land Cover Changes and Water Requirements in the Basin of the Urmia Lake, Iran. *Water Supply* **2023**, *23*, 2299–2312, doi:10.2166/ws.2023.132.
54. Saraf, N.M.; Lokman, M.F.; Abdul Rasam, A.R.; Hashim, N. Assessment of Urban Growth Changes in Klang District Using Support Vector Machine by Different Kernel. *IOP Conf Ser Earth Environ Sci* **2022**, *1051*, 012023, doi:10.1088/1755-1315/1051/1/012023.
55. Dahiya, N.; Gupta, S.; Singh, S. Qualitative and Quantitative Analysis of Artificial Neural Network-Based Post-Classification Comparison to Detect the Earth Surface Variations Using Hyperspectral and Multispectral Datasets. *J Appl Remote Sens* **2023**, *17*, doi:10.1117/1.JRS.17.032403.
56. Wu, J.L.; Ho, C.R.; Huang, C.C.; Srivastav, A.L.; Tzeng, J.H.; Lin, Y.T. Hyperspectral Sensing for Turbid Water Quality Monitoring in Freshwater Rivers: Empirical Relationship between Reflectance and Turbidity and Total Solids. *Sensors (Switzerland)* **2014**, *14*, 22670–22688, doi:10.3390/s141222670.
57. Chen, H.; Lin, X.; Sun, Y.; Wen, J.; Wu, X.; You, D.; Cheng, J.; Zhang, Z.; Zhang, Z.; Wu, C.; et al. Performance Assessment of Four Data-Driven Machine Learning Models: A Case to Generate Sentinel-2 Albedo at 10 Meters. *Remote Sens (Basel)* **2023**, *15*, doi:10.3390/rs15102684.
58. Deshpande, A.R.; Emmanuel, M. Context Based Recommendation Methods: A Brief Review. *Int J Comput Appl* **2016**, *14*–19.
59. Hafeez, S.; Wong, M.S.; Ho, H.C.; Nazeer, M.; Nichol, J.; Abbas, S.; Tang, D.; Lee, K.H.; Pun, L. Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-Ii Waters: A Case Study of Hong Kong. *Remote Sens (Basel)* **2019**, *11*, doi:10.3390/rs11060617.
60. Hua, A.K. Land Use Land Cover Changes in Detection of Water Quality: A Study Based on Remote Sensing and Multivariate Statistics. *J Environ Public Health* **2017**, *2017*, 5–7, doi:10.1155/2017/7515130.
61. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens Environ* **1991**, *37*, 35–46.
62. Kang, C.S.; Kanniah, K.D.; Mohd Najib, N.E. Google Earth Engine for Landsat Image Processing and Monitoring Land Use/ Land Cover Changes in the Johor River Basin, Malaysia. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS); 2021; pp. 4236–4239.
63. Story, M.; Congalton, R.G. Accuracy Assessment: A User's Perspective. *Photogramm Eng Remote Sensing* **1986**, *52*, 397–399.
64. Ashiagbor, G.; Asare-Ansah, A.O.; Amoah, E.B.; Asante, W.A.; Mensah, Y.A. Assessment of Machine Learning Classifiers in Mapping the Cocoa-Forest Mosaic Landscape of Ghana. *Sci Afr* **2023**, *20*, e01718, doi:10.1016/j.sciaf.2023.e01718.
65. Campbell, J.B.; Wynne, R.H. *Introduction to Remote Sensing*; Guilford press, 2011; ISBN 1609181778.
66. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons, 2015; ISBN 111834328X.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.