African Journal of Mining, Entrepreneurship and Natural Resource Management (AJMENRM)
ISSN: 2706-6002 Volume 1, Issue 2 (April 2020), PP 115-124
www.ajmenrm.ttu.ac.ke/

Land use/land cover classification for the iron mining site of Kishushe, Kenya: A feasibility study of traditional and machine learning algorithms

M. Siljander[1*], N. Adero[2,5], F. Gitau[2], E. Nyambu[3,4]
[1]*Department of Geosciences and Geography, University of Helsinki, Finland*
*(*Corresponding author: mika.siljander@helsinki.fi)*
[2]*Taita Taveta University, School of Mines and Engineering, Kenya*
[3]*Taita Taveta University, School of Science and Informatics, Kenya*
[4]*Taita Taveta County Government, Kenya*
[5]*Technische Universität Bergakademie Freiberg, Department of Surface Mining, Germany*

Abstract: *Motivated by the need to enhance the precision of land use/land cover classification for mining environments challenged by rapid anthropogenic and natural changes, we analysed multispectral Sentinel 2A satellite data using four different classifiers: Maximum Likelihood Classifier (MLC), Support Vector Machine (SVM), Random Tree (RT) and Random Forest (RF). Using adjusted training sample sizes drawn from the Kishushe iron ore mining site in Taita Taveta, Kenya, we conducted image analysis and compared the classification accuracies of the four methods, confirmed further by ground truthing. The study met the main objective of evaluating and comparing the performance of the traditional Maximum Likelihood classifier with the three machine learning algorithms of Support Vector Machine (SVM), Random Trees (RT), and Random Forest (RF). Eight land use/land cover classes were generated from each of the four classifications performed in R statistics software for RF and in ArcGIS 10.7 for RT, MLC and SVM methods. Random Forest (RF) method delivered the best overall accuracy at 74.63 % with a Kappa value of 0.67. Random Trees (RT) method came second at 72.64 % with a Kappa value of 0.64. The overall accuracy of the SVM method was 58.21 % with a Kappa value of 0.46 and for the MLC method, the overall accuracy was 57.21 % with a Kappa value of 0.45. These results confirmed that machine learning classifiers outperform traditional classifiers. The study also confirmed that for robust land use/land cover classification, it is essential to have quality training data as the quality can have large and considerable effects on classification results. Since the reliability of land use/land cover (LULC) maps derived from remotely sensed data for mining sites depends on accurate classification, this study gives evidence-based recommendation for adopting machine learning algorithms in satellite image analysis and classification to support environmentally sustainable decisions and informed policy direction for sound mine planning and monitoring.*

Keywords- *Accuracy, earth observation, land use/land cover change, Maximum Likelihood, Random Forest, Random Trees, Support Vector Machine.*

1. INTRODUCTION

Globally, there has been a rising awareness of, and political goodwill towards, the universal sustainability agenda. This growing awareness concerns both the United Nations SDGs and region- or country-specific sustainability goals. Research on mining and sustainable development must, therefore, look into the resource goals in the food-water-energy nexus interconnecting SDGs 2, 6, 7; the environmental goals including ecosystems and natural capital; and the issues affecting land, climate, and oceans interconnecting SDGs 13, 14, 15 ([1]). The African Union Agenda 7, on environmentally sustainable and climate resilient economies and communities, and Kenya Vison 2030's social pillar on ensuring a clean, safe and healthy environment, are also critical to environmentally responsible mining ([2]; [3]). Recent advances in satellite remote sensing have availed higher resolutions for environmental studies, a substantial improvement over the limited military applications that characterised the first artificial satellites launched in 1957 and 1958 by Russia and the USA, respectively. Digital Earth Africa, a recent initiative for earth observation services in Africa which was widely publicised at the 3rd RCMRD International Conference & 4th AfriGEO Symposium held in Nairobi, August 12–16, 2019, is availing analysis-ready geodata to serve environmental research interests in Africa. Artificial Intelligence (AI) and Machine Learning (ML), both key features of the wave of digitalisation and the Fourth Industrial Revolution (Industry 4.0 or 4IR), continue to leverage intelligent spatial decision support systems by learning from the big data generated from satellite-based imaging sensors. Industry 4.0 has been reflected as Mining 4.0 in the modern extractives industry.

TAITAGIS, a Finnish-Kenyan capacity building project in geoinformatics hosted at Taita Taveta University in Kenya's coastal mineral belt, has been promoting research which applies GIS and remote sensing techniques to advance environmentally responsible mining. Kenya's new Mining Act of 2016 has been referred to as Africa's most modern and progressive mining law, especially because of its provisions for empowering local communities, land and property-owners, and artisanal miners as legitimate participants in decision making on land acquisition, benefits sharing, and capacity development. This Kenyan case study is justified as part of the key research contributions providing new intelligent tools in support of the precise spatial metrics required to operationalise progressive mining regulations at community-wide scales in a coherent and systemic manner [4]. The iron ore mining site in Kishushe area of Kenya's coastal mineral belt of Taita Taveta was selected as a suitable case study for the main research purpose of testing, evaluating, and comparing the performance of traditional and machine learning algorithms. Four different classification methods were compared in evaluating land use/land cover features across the mining site. These methods were: Maximum Likelihood Classification (MLC); Support Vector Machine (SVM); Random Trees (RT); and Random Forest (RF).

The specific objectives of this study were to: (1) evaluate and compare four image algorithms to distinguish between eight different land use/land cover classes based on Sentinel 2A image; and (2) compare the different prediction maps obtained from MLC, SVM, RT, and RF. The different methods were applied to multi-sensor Sentinel-2 data to classify eight different land categories using the same training sites from Wanjala iron mining site in Kishushe. Sentinel 2A imagery was used in this study because it is freely available and has been proven to be appropriate for land use/land cover classifications.

The hypothesis of this paper is that modern machine learning algorithms are more reliable than traditional methods in classifying satellite imagery across spaces with complex land use/land cover dynamics such as mining areas. The second section of the paper presents the literature survey, beginning with the background and scope of the study setting in Kenya. The third section presents the research methodology, delving into the details of pre-processing imagery, training data, classification, and accuracy assessments. The fourth section presents the results, which are then discussed in detail. The last section presents the conclusion and outlook for further research.

2. LITERATURE SURVEY

Minerals and metals have been central to all the stages of human civilisation, with iron ore production trends growing in tandem with industrial developments. With rapid technological advances in the modern world, there is an ever-increasing strategic significance of minerals. However, mineral extraction has various ecological, environmental and social-economic impacts as evident in Taita Taveta County, Kenya ([4]; [5]; [6]; [7]). It is therefore, of high priority to improve their sustainability to avert the hazards caused during extraction. Accurate spatial information is key to sustainable mining practices and compliance monitoring. Geomonitoring is essential in this respect, hence satellite remote sensing and the spatial analysis tools availed by geographic information systems (GIS).

Various studies across the world show that land clearance, mineral prospecting, mineral exploration and mine development have resulted in large-scale land use/land cover changes across the world ([4]; [5]; [8]; [9]; [10]; [11]). Mining activities can change entire landscapes, surface mining particularly. Surface mining involves accessing shallow ore deposit by removing the soil or rock layers above it and moving the extracted overburden materials to nearby areas. Such activities alter land cover by reducing vegetation cover and accelerating desertification, with large-scale surface mining activities further affecting soil fertility, ecosystems, and regional biodiversity ([8]; [9]; [10]; [11]; [12]; [13]). Mining activities have also raised concerns on degrading water quality [14; [15]] and air quality ([16]). and often also public health. In this sense, unregulated mining activities have widespread, direct and indirect impacts on sustainable development, which can further be explored globally within the framework of Sustainable Development Goals (SDGs). These facts motivate the deployment of reliable earth observation and geodata processing techniques to classify and determine the precise spatial extents of various land use/land cover categories across mining sites. Though surface mining areas are distinguishable from satellite images, the detection and determination of the sizes of these areas through computation has been facing several challenges. To overcome these computational challenges, detailed analysis of the spectral signatures of various land cover classes is needed to provide distinguishable characteristics of surface mining areas as they include not only mine surface elements, but also waste rock dumps, tailings dams, water storage ponds, access roads, milling and processing infrastructure, and housing for workers ([12]; [17]). The complex characterisation of mining areas is evident in the usual co-existence of natural surface features such as forests together with zones of cleared land for development and other competing or conflicting interests and activities such as biodiversity conservation and agriculture. Therefore, it is important to precisely discriminate the land-use categories types and determine the spatial extents of land use/land cover characteristics of surface mining sites.

Several studies have compared variances between different machine learning (ML) algorithms in classifying the land use/land cover (LULC) using remotely sensed satellite images. Srivastava et al. [18] used maximum likelihood classifier (MLC), Support Vector Machine (SVM) and Artificial Neural Networks (ANN) to perform classification on three Landsat images and compared their accuracies using ground-truthed data. Results showed that Artificial Neural Networks (ANN) had the highest accuracy, followed by Support Vector Machine (SVM) and lastly, Maximum Likelihood Classifier in the last place (MLC). For example, research conducted by Garai and Narayana [19] applied unsupervised classification to land use/land cover change detection utilizing unsupervised classification method to find a reduction of 4.7 % in vegetative cover, being the result of coal mining. In Ghana, Basommi et al. [20] used maximum likelihood classifier to analyse land use/land cover changes in a mining area to determine a significant reduction in vegetative cover. Normalized Difference Vegetation Index (NDVI) has also been applied into land use/land cover changes for open cast mining ([9]). Support vector machine SVM algorithm is most widely used due to its capabilities of training the classifier with less training sample data ([21]). In the recent years, Random forest (RF) machine learning algorithm has widely been used by many researchers to perform image classification ([22]; [23]; [24]; ([25]). A study made by Pelletier et al. [26] compared classification results in the south of France, and it proved that RF had a higher overall accuracy compared to SVM. However, performance of a machine learning classification algorithm method is highly dependent on the study region land use/land cover types chosen and the satellite data used in study ([11]).

3. MATERIALS AND METHODS

The study area in Kenya, Kisushe, has interesting location features and a long history of land-mining-society conflicts which informed its choice for this research. The applied geospatial and space technology products, conceptual framework, and software services were so selected as to meet the need for land use/land cover classification maps, which are the basis for the shared visual understanding essential to transparent and evidence-based decisions.

3.1. Study area

3.1.1. Geographical and geological setting

The study area, Kishushe, is flat and located in the Kenyan coastal mineral belt of Taita Taveta County, Taita Sub-county, as shown in Figure 1. It lies between 380 10’ 1.45’’ E and 380 11’ 23.22’’ E, and between 30 14’ 26.30’’ S and 30 15’ 27.30’’ S. Kishushe borders Tsavo East National Park to the north and is part of the region’s geologically mineral-rich Mozambique belt with gemstones and industrial minerals ([27]; [28]). The mining of iron ore has been taking place in this area since early 2008 ([28]). In Figure 2 is shown a high-resolution true-colour image of the mining site.

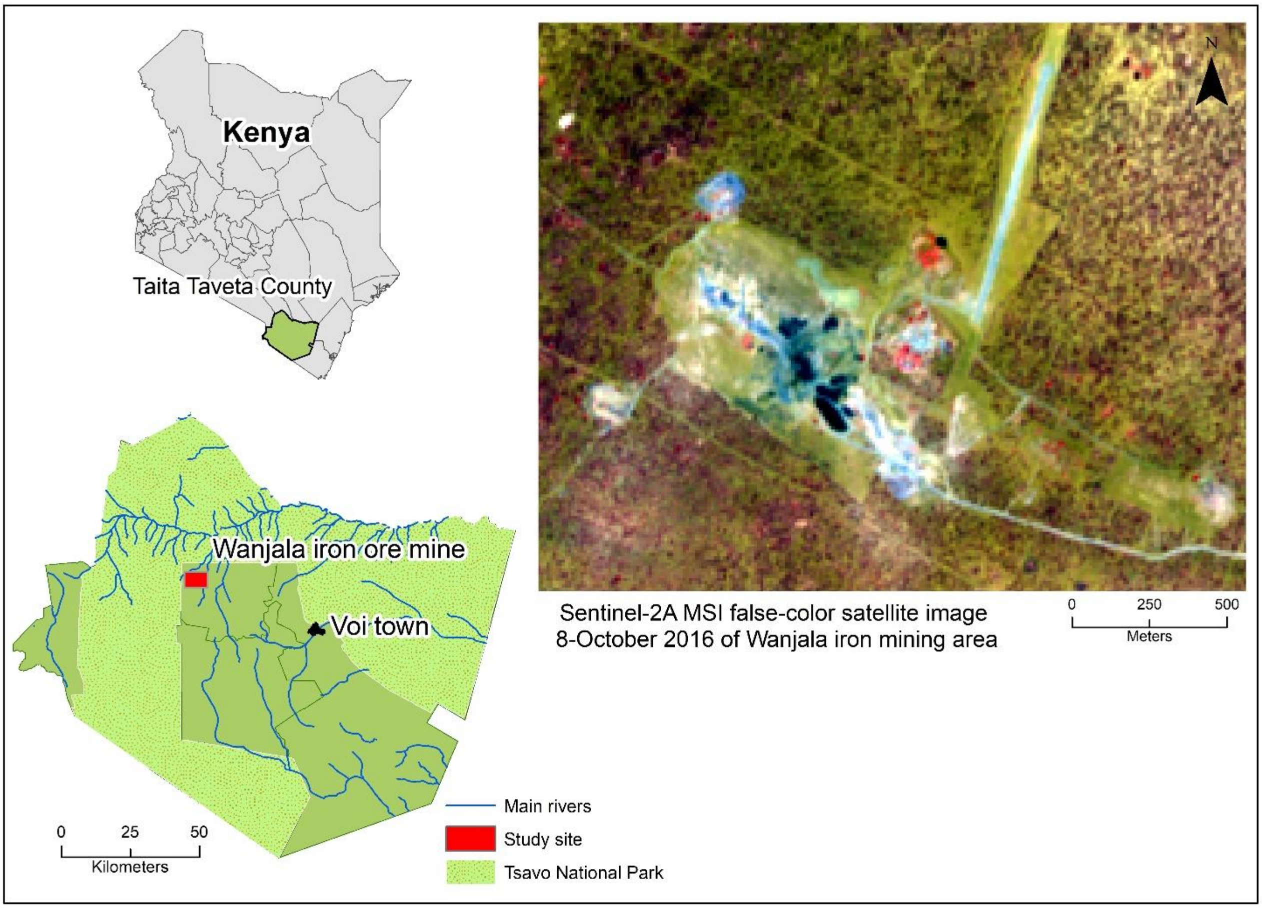

Figure 1 Location of Wanjala iron ore mine in Kishushe, Taita Taveta County, Kenya.

3.1.2. Iron ore production in Kishushe

The main part of the study area, Kishushe, consists of iron ore mine of medium grade (<62% Fe). This mining site commands the leading position as the only large-scale iron ore mine in Kenya. The iron ore occurs in both alluvial and vein forms. This situation favours open-pit mining methods which, however, impact heavily on vegetation and lead to serious land degradation ([29]; [30]). The study area was initially zoned for ranching. It had tall trees and good undergrowth for the community to graze and collect grass. The sudden change of land use from ranching to mining has significantly impacted the land cover and the ranching community rights, leading to constant conflicts between the local community and the mining company, Wanjala Mining Company [31].

3.1.3. Chronology of conflicts due to iron ore mining in Kishushe

For many years now, Kenya's only large-scale iron ore mine in Kishushe has been faced with long-standing land-related disputes. This scenario represents the competing and conflicting "mining and non-mining land use" interests in Taita Taveta County, which need urgent redress through integrated spatial modelling and information systems to avail shared visual evidence and "actionable location-based intelligence" ([5], [6]). The chronology of key conflicts in Kishushe can be traced back to the 1960s, when disputes between local communities and the colonial government started and led to a decline in the erstwhile normal production of iron ore in the 19th and part of the 20th century.

From the 1990s to 2010, the increasing value of iron ore and rising community awareness led to rising conflicts, pitting local leaders and local community against the mining company, with subsequent arrests. Over the 2010–2014 period, conflicts over prospecting rights, pegging, and claims arose [32]. A public inquiry conducted in 2016 by the Kenya National Commission on Human Rights [27] and Taita Taveta University established prevalent conflicts over land ownership between the local community and the mining company, resulting in demonstrations and stoppage of company trucks from ferrying iron ore from the mine site. Boundary disputes also arose in early 2014, between Kishushe Cooperative Society and Wanjala Mining Company [27]. In 2019, demonstrations escalated leading to the closure of the mining operations due to worsening conflicts between the company and the local community over workers' welfare and environmental impacts [33].

Sustainable solutions to these disputes depend on the transparency of the shared visual evidence and active participation of all stakeholders in decision making, which is the value addition a map-based research of this kind provides.

Figure 2 True colour image showing an aerial view of Wanjala iron mining site land use/land cover [34].

3.1.4. Overview of the study methodology

The workflow adopted, as shown in Figure 3, consisted of the following steps: (1) satellite image data pre-processing; (2) training area data creation; (3) data classification, where four different classification algorithms were applied; (4) assessing the classification accuracies; and (5) comparison and mapping of the classification results. Each step has been explained in detail hereafter.

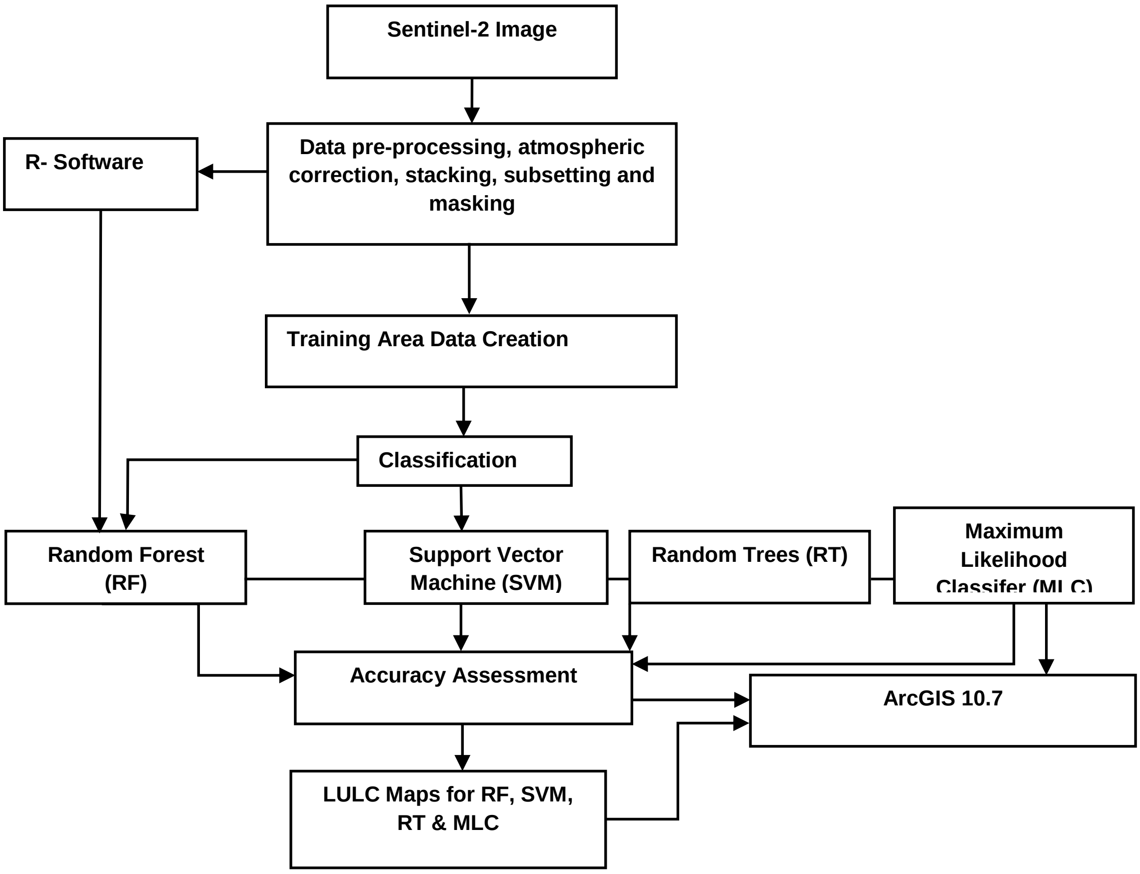

Figure 3 Procedural workflow from pre-processing data to comparison of classification results.

3.1.5. Satellite image data and pre-processing

The available Sentinel-2 satellite image Level-1C (S2MSI1C-2016-10-08) 2A_MSIL1C_20161008T074232_N0204_R049_T37MDS_20161008T075021 for the study area was obtained from the European Space Agency's (ESA) Sentinel Scientific Data Hub. We processed reflectance images from Top-Of Atmosphere (TOA) Level 1C S2, to Bottom-Of-Atmosphere (BOA) Level 2A by using Sen2Cor v2.8 standalone processor from the ESA ([35]). Level 1C Sentinel was georeferenced to WGS 1984, UTM Zone 37 N coordinate system. For this study, we only used the bands with a spatial resolution of 10 m namely: band 2-Blue, band 3-Green, band 4-Red, band 8-NIR, respectively. The other steps in pre-processing of the Sentinel (BOA) satellite image included layer stacking, subsetting and masking, which were all conducted in R software with raster and RStoolbox package [36].

3.1.6. Training data

Prior knowledge of the study area was obtained by visiting the study site and collecting training samples using a handheld GPS receiver for the key land-use classes in the mining area. We used Google Earth Pro [34] to collect and digitise additional training samples with visual interpretation of the available very high-resolution image (16th June 2016). We developed a classification scheme based on the prior knowledge of the study site where eight training classes were identified, namely: bare ground, cleared ground, green vegetation, iron ore stockpile, open ground and open cut pits, open water, waste rock and sand dumps, and savanna grassland. A five-metre buffer was calculated to all the training points (n=161) to gain more pixels for training the image with the different classifiers. In order to reduce possible error due to the non-random selection of training points, another point data set (n=201) was created using Create random points tool in ArcGIS and interpreting the land use/land cover classes with Google Earth. This second data set was used for land use/land cover classifications accuracy assessment.

3.1.7. Classification approach

In total, 161 training samples were used for the classification of Sentinel image and the same training sites were used for all the four different classification algorithms. In this study, one traditional classification algorithm, the Maximum Likelihood Classifier (MLC), and three machine learning methods: Support Vector Machine (SVM), Random Forest (RF), and Random Trees (RT) were used and their performances were compared. Three of the classification algorithms (MLC, SVM, and RT) were performed in ArcGIS (ESRI, Redlands, CA, USA) with default parameter settings and one classification algorithm (RF) was computed in R statistical software version 3.6.1. ([37]). We intentionally compared RT and RF classification methods to see if the classification algorithm performance has differences between software. The classifiers are explained briefly, subsequently.

3.1.7.1. Maximum Likelihood Classifier (MLC)

The Maximum Likelihood Classifier is the most common method used for remote sensing supervised classification. It is a parametric statistical method where the analyst first supervises the classification by identifying representative areas called training areas and the classification process is a standard pixel-based technique using a multivariate probability density function of classes ([38];[39]). In this study, MLC was trained and classified using ArcGIS 10.7 software with default parameters.

3.1.7.2. Support Vector Machine (SVM) Classifier

Invented by Vladimir N. Vapnik, [40] “Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. As a nonparametric method, it requires no a priori knowledge of the structure and statistical distribution of the data being analysed and incorporates a regularisation technique which enables it to achieve relatively stable solutions with respect to other classification algorithms when addressing least square problems with noisy data ([41]). For classification, SVM is a discriminative classifier formally defined by a separating hyperplane. The algorithm outputs an optimal hyperplane, which categorises new examples. In two-dimensional space, this hyperplane is a line dividing a plane in two parts where in each class lies in either side. SVM separates the classes with a decision surface that maximises the margin between the classes ([40]; [21]). We trained the model with SVM classifier in ArcGIS 10.7 with default parameters.

3.1.7.3. Random Forest (RF) Classifier

Random Forests (RF) is an ensemble learning classifier, based on constructing a multitude of decision trees, choosing random subsets of variables for each tree, and using the most frequent tree output as the overall classification ([42]). Random Forests are widely used in remote sensing applications as it corrects for decision trees' habit of overfitting to their training set. A detailed review of RF in remote sensing can be found in ([43]). In our study, RF was computed using the randomForest (v.4.6-14) package [44] in R statistics software ([37]).

3.1.7.4. Random Trees (RT) Classifier

Random Trees is a supervised machine-learning classifier based on constructing a multitude of decision trees. Random Trees is basically the same classification method as Random Forest because the Random Trees classifier uses [42] Random Forest algorithm; however, it is applied to ArcGIS software. In this study, we used default parameters to train the Random Trees classifier.

3.1.8. Accuracy Assessment

Independent validation data set were applied to investigate the classification accuracy. We calculated the Overall Accuracy (OA) and the Kappa Coefficient (κ) [45] based on the confusion matrices of different classification techniques. The accuracy of each classification method: Maximum Likelihood (MLC), Support Vector Machine (SVM), Random Trees (RT) and Random Forest (RF) was determined by performing an accuracy assessment. In this study, reference data that included 201 points, was randomly generated for the study area for eight land use/land cover classes in ArcGIS 10.7 software.

The values of each classified raster in different methods (MLC, SVM, RT, and RF) were extracted to points. This joined the raster values of different classification classes to the random reference points. The frequency of each randomly extracted points was then calculated to determine how frequently the predicted and the true reference values occurred in different classification classes. Finally, a pivot table was created for the error matrix of each classification. The Overall Accuracy (OA), producer’s and user’s accuracy and the Kappa Coefficient (κ) [45] values for different classes were analysed to evaluate the accuracies of the classification.

4. Results and Discussion

4.1. Results

The results of land use/land cover (LULC) classification accuracies obtained by the methods for the eight different classes are presented here.

4.1.1. Accuracy Assessment Results

The best overall accuracy was found for the Random Forest (RF) method as 74.63 % with a Kappa value of 0.67. The overall accuracy was 72.64 % with a Kappa value of 0.64 for the Random Trees (RT) method. The overall accuracy of the SVM method was 58.21 % with a Kappa value of 0.46. For the MLC method the overall accuracy was 57.21 % with a Kappa value of 0.45. In Table I is the presentation of the summary of the overall accuracies of all the land use/land cover classes and Kappa values of the four classification algorithms.

Table I Summary of overall land use/land cover classification accuracies and Kappa values among the four classification algorithms: MLC, SVM, RT, and RF.

	MLC	SVM	RT	RF
Overall accuracy (%)	57.21	58.21	72.64	74.63
Kappa statistic	0.45	0.46	0.64	0.67

In order to better understand the classification confusions between land use/land cover types, error matrices are provided for different classification methods in Tables II, III, IV, and V.

Table II Error matrix for land use/land cover using Maximum Likelihood algorithm.

LULC Classes	Iron Ore Stockpile	Open Water	Green Vegetation	Waste Rock and Sand Dumps	Cleared Ground	Open Ground and Open Cut Pit	Bare Ground	Savanna Grassland	Totals
Iron Ore Stockpile	**3**	0	2	2	0	0	0	**0**	7
Open Water	0	**1**	0	0	0	0	1	0	2
Green Vegetation	0	0	**18**	1	1	0	0	1	21
Waste Rock and Sand Dumps	0	0	0	**7**	0	1	0	0	8
Cleared Ground	0	0	2	7	**5**	0	3	1	18
Open Ground and Open Cut Pit	2	0	1	9	0	**14**	4	0	30
Bare Ground	0	0	0	0	0	0	**6**	0	6
Savanna Grassland	0	0	40	1	5	0	2	**61**	109
Totals	5	1	63	27	11	15	16	63	201
Producer's Accuracy	60.0%	100.0%	28.6%	25.9%	45.5%	93.3%	37.5%	96.8%	
User's accuracy	42.9%	50.0%	85.7%	87.5%	27.8%	46.7%	100.0%	56.0%	
Total true Value	115		**Overall Accuracy**		57.21%				
Total Reference Points	201		**Overall Kappa Statistics**		0.45				

Table III Error matrix for land use/land cover (LULC) using Support Vector Machine algorithm.

LULC Classes	Iron Ore Stockpile	Open Water	Green Vegetation	Waste Rock and Sand Dumps	Cleared Ground	Open Ground and Open Cut Pit	Bare Ground	Savanna Grassland	Totals
Iron Ore Stockpile	**3**	0	3	2	0	0	0	0	8
Open Water	0	**1**	0	0	0	0	0	0	1
Green Vegetation	0	0	**11**	0	1	0	0	0	12
Waste Rock and Sand Dumps	1	0	0	**13**	0	1	0	0	15
Cleared Ground	0	0	3	8	**5**	0	3	1	20
Open Ground and Open Cut Pit	1	0	0	2	0	**12**	0	0	15
Bare Ground	0	0	0	0	0	2	**10**	0	12
Savanna Grassland	0	0	46	2	5	0	3	**62**	118
Totals	5	1	63	27	11	15	16	63	201

Producer's Accuracy	60.0%	100.0%	17.5%	48.1%	45.5%	80.0%	62.5%	98.4%
User's accuracy	37.5%	100.0%	91.7%	86.7%	25.0%	80.0%	83.3%	52.5%
Total true Value	117	**Overall Accuracy**		58.21%				
Total Reference Points	201	**Overall Kappa Statistics**		0.46				

Table IV Error matrix for land use/land cover (LULC) using Random Trees algorithm.

LULC Classes	Iron Ore Stockpile	Open Water	Green Vegetation	Waste Rock and Sand Dumps	Cleared Ground	Open Ground and Open Cut Pit	Bare Ground	Savanna Grassland	Totals
Iron Ore Stockpile	**3**	0	1	2	0	0	0	0	6
Open Water	0	**1**	0	0	0	0	0	0	1
Green Vegetation	0	0	**46**	0	1	0	0	0	47
Waste Rock and Sand Dumps	1	0	1	**14**	0	3	0	0	19
Cleared Ground	0	0	1	7	**5**	0	2	4	19
Open Ground and Open Cut Pit	1	0	0	1	0	**11**	3	0	16
Bare Ground	0	0	0	1	0	1	**7**	0	9
Savanna Grassland	0	0	14	2	5	0	4	**59**	84
Totals	5	1	63	27	11	15	16	63	201
Producer's Accuracy	60.0%	100.0%	73.0%	51.9%	45.5%	73.3%	43.8%	93.7%	
User's accuracy	50.0%	100.0%	97.9%	73.7%	26.3%	68.8%	77.8%	70.2%	
Total true Value	146	**Overall Accuracy**		72.64%					
Total Reference Points	201	**Overall Kappa Statistics**		0.64					

Table V Error matrix for land use/land cover (LULC) using Random Forest algorithm.

LULC Classes	Iron Ore Stockpile	Open Water	Green Vegetation	Waste Rock and Sand Dumps	Cleared Ground	Open Ground and Open Cut Pit	Bare Ground	Savanna Grassland	Totals
Iron Ore Stockpile	**3**	0	1	2	0	0	0	0	6
Open Water	0	**1**	0	0	0	0	0	0	1
Green Vegetation	0	0	**54**	0	1	0	0	1	56
Waste Rock and Sand Dumps	0	0	1	**10**	0	2	0	0	13
Cleared Ground	1	0	1	8	**5**	0	2	4	21
Open Ground and Open Cut Pit	1	0	0	1	0	**11**	2	0	15
Bare Ground	0	0	0	2	0	2	**8**	0	12
Savanna Grassland	0	0	6	4	5	0	4	**58**	77
Totals	5	1	63	27	11	15	16	63	201
Producer's Accuracy	60.0%	100.0%	85.7%	37.0%	45.5%	73.3%	50.0%	92.1%	
User's accuracy	50.0%	100.0%	96.4%	76.9%	23.8%	73.3%	66.7%	75.3%	
Total true Value	150	**Overal Accuracy**		74.63%					
Total Reference Points	201	**Overal Kappa Statistics**		0.67					

The average values of producer's and user's accuracy are presented in Table VI. Random Forest (RF) achieved the highest overall accuracy, followed by the Random Trees (RT) and Support Vector Machine (SVM) methods. The Maximum Likelihood (MLC) method had clearly a lower overall accuracy than all the other methods.

Table VI: Average of producer's and user's accuracy for different land use/land cover (LULC) classification algorithms.

Method	Average of producer's and user's accuracy
Random Forest (RF)	69.13%
Random Trees (RT)	69.11%
Support Vector Machine (SVM)	66.80%
Maximum Likelihood (MLC)	61.50%

4.1.2. Mapping performance comparison of the four classification algorithms

The results of the classification of the land use/land cover characteristics for the year 2016 with four different algorithms (MLC, SVM, RT, and RF) is shown in Figure 4. As indicated earlier, the LULC maps had eight classes and the savanna grassland dominates in each of the classifications. From Figure 4, it can also be seen that there exist some clear differences in different algorithm map classifications, and this will be discussed in detail in the discussion section.

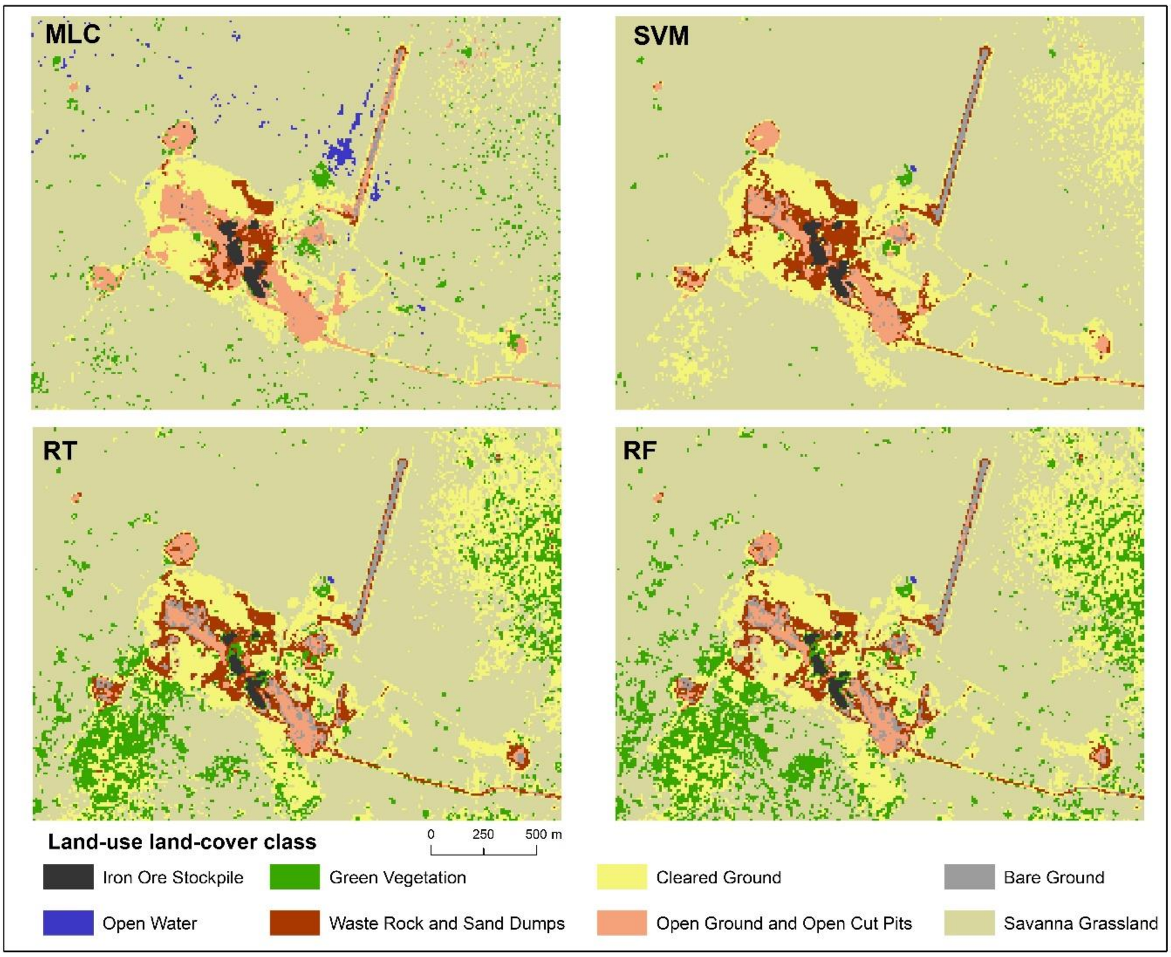

Figure 4 land use/land cover classifications for four different algorithms including Maximum Likelihood Classifier (MLC), Support Vector Machine (SVM), Random Tree (RT), and Random Forest (RF).

4.2. Discussion of Results

We discuss here the results obtained with emphasis on evaluating and comparing the accuracy assessments.

4.2.1. Accuracy assessment and map comparison

Information obtained from land use/land cover (LULC) analysis may be required for making informed decisions, policy making and/or administrative purposes. With the respective spatial details, the information from LULC analysis may be crucial for environmental protection and spatial planning, hence the need for it to be as accurate as possible. This is made possible by carrying out accuracy assessments or validation to prove the reliability of the resulting data and information to various users [46].

In this study, accuracy assessment results illustrated that open water, green vegetation, and savanna grassland classes could be accurately identified having average producer's and user's accuracy of 100%, 91.1%, and 83.7% using Random Forest; and 100%, 85.4%, and 81.9% using Random Trees classification methods for the Sentinel 2 image. For Support Vector Machine (SVM) and Maximum Likelihood Classifier (MLC), the LULC classification performed differently. We established the greatest differences in the classification of green vegetation and cleared vegetation, which were predicted as savanna grasslands and bare ground using the SVM and MLC (Figure 4). Figure 4 shows that the traditional MLC clearly overestimates open water class whereas all the three machine learning classifiers were able to detect the open water class satisfactorily. In Figure 5, there is only a small watering pond. MLC estimated open water to be all around the study area, which was not the true scenario on the ground after confirmation by ground truthing.

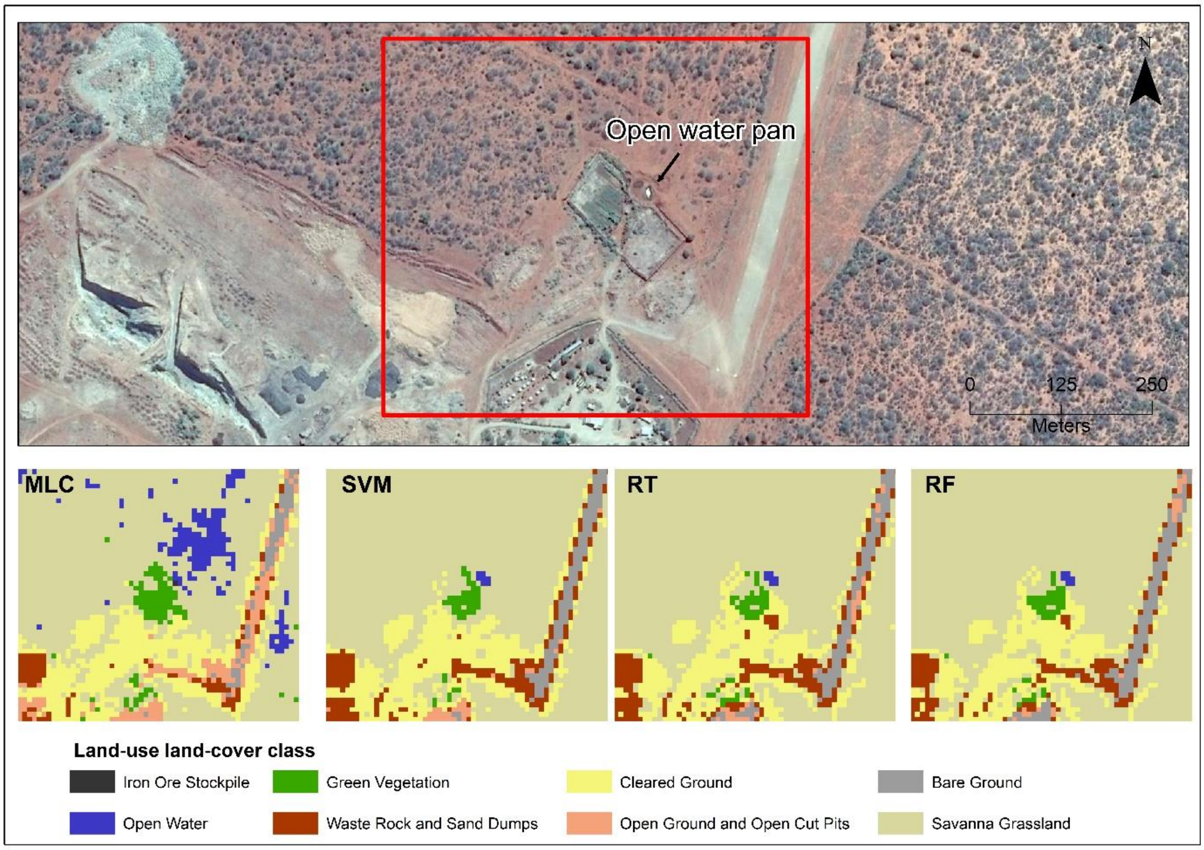

Figure 5 Maximum Likelihood Classifier (MLC) identifying open water classes where there were none, in contrast to the right classification made by the three machine learning algorithms: Support Vector Machine (SVM), Random Tree (RT), and Random Forest (RF).

From Figure 6, it can be seen that both RT and RF classification algorithms predicted large areas as green vegetation (verified with Normalized Difference Vegetation Index – NDVI) in contrast to MLC and SVM, which predicted little green vegetation. This outcome is possibly due to a mixed pixel challenge as portrayed in MLC and SVM methods. Therefore, savanna grassland appearing to have similar landscape characteristics and similar spectral response to green vegetation was predicted more in SVM and MLC. RF and RT proved to produce more accurate results compared to MLC and SVM methods for all the related land use/land cover classes. It should be noted, however, that SVM has widely been applied successfully to bitemporal forest-cover change studies with higher accuracies than decision trees as noises in the training data increased ([47]). This study used only pixel-based classification approach; however, it has been shown in many studies that object-based image classification can outperform pixel-based classification as objects can approximate heterogeneous real-world features better than arbitrary pixels. In addition, objects, instead of pixels, also reduce local spectral variability and the 'salt and pepper' effect (see e.g. [48]; [49]; [50]).

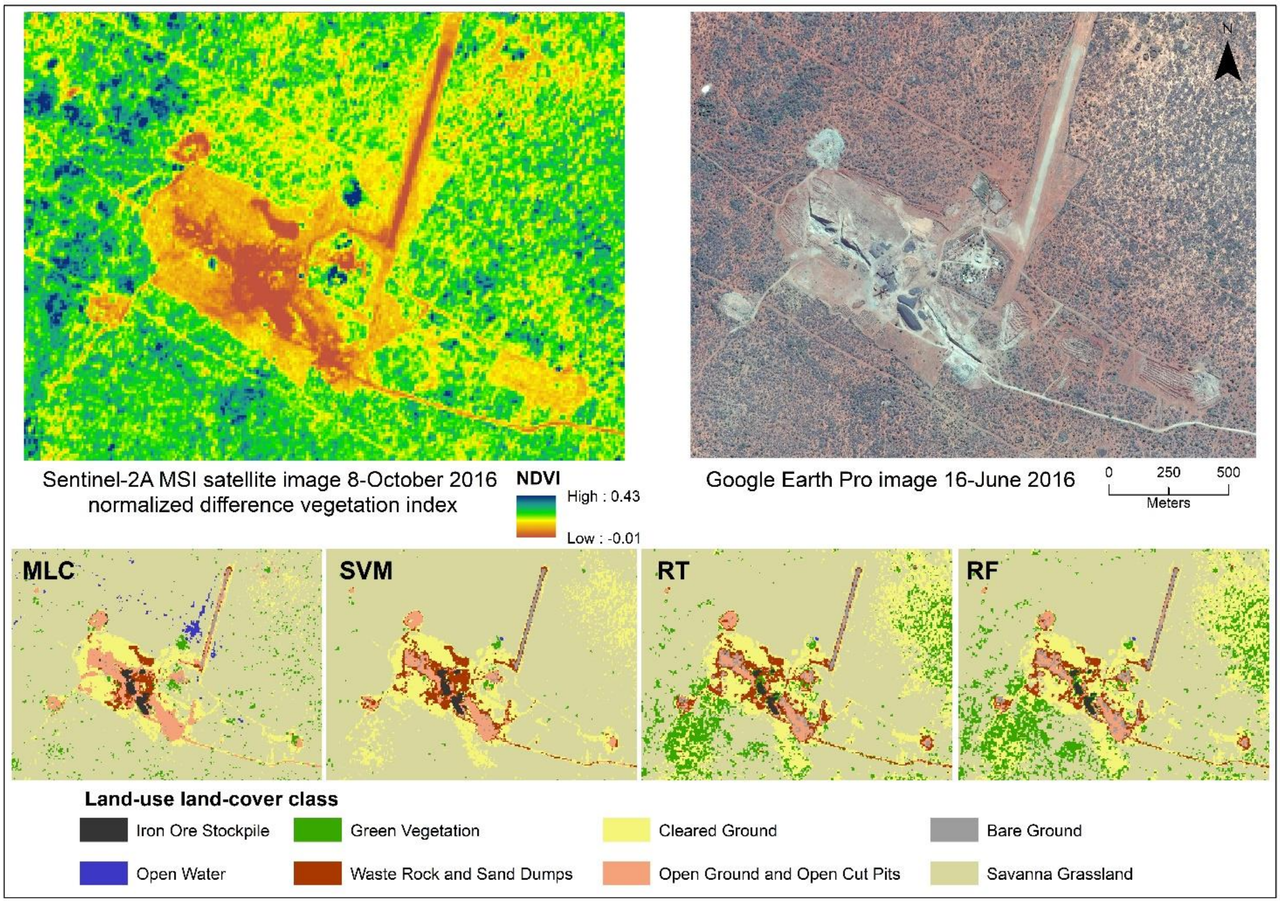

Figure 6 Maximum Likelihood Classifier (MLC) and Support Vector Machine (SVM) predicting much less green vegetation than Random Tree (RT) and Random Forest (RF) verified with normalised difference vegetation index (NDVI) map interpretation.

5. CONCLUSIONS

Earth observation services through satellite remote sensing and computer-based image analysis capabilities have undergone profound advancements since the late 20th century. The current velocity of changes in spatial data and information generation technology through remote sensing and image interpretation, leveraged by computer algorithms, continues to improve land use/land cover mapping and analysis for georeferenced and decision-ready results. For sound decision support, however, the different methods of land use/land cover classification must be evaluated for reliability, as this study has demonstrated by applying and comparing four different classification methods to a case study of a mining site in Kishushe, Taita Taveta County, Kenya.

The study met the main objective of evaluating the performance of the traditional Maximum Likelihood Classifier (MLC) with the machine learning algorithms of Support Vector Machine (SVM), Random Trees (RT), and Random Forest (RF). The feasibility of the four approaches as applied to the selected iron mining site in Kenya was performed using Sentinel 2A satellite image data. Eight land use/land cover classes were generated from each of the four classifications performed in R statistics software for RF and in ArcGIS 10.7 for RT, MLC and SVM methods. Random Forest (RF) method delivered the best overall accuracy at 74.63 % with a Kappa value of 0.67. Random Trees (RT) method came second at 72.64 % with a Kappa value of 0.64. The overall accuracy of the SVM method was 58.21 % with a Kappa value of 0.46 and for the MLC method, the overall accuracy was 57.21 % with a Kappa value of 0.45.

Intentionally, the study did not analyse land use/land cover change over time, but only analysed a one-time Sentinel satellite image for October 08, 2016. The reasons for this were, firstly, we did not have up-to-date training data from the study site, and it is well known that land use/land cover can rapidly change across mining areas. In Figure 7 is shown a Change Vector Analysis (CVA) map of the study area that characterises dynamic changes in multi-spectral space by a change vector over two multi-temporal Sentinel 2A imageries of 2016 and 2019. Visible in Figure 7 are clear changes in black iron ore stockpiles as well as expansion of cleared ground among other changes in land use/land cover. Secondly, this study focused on the mapping and accuracy performance of different classification algorithms instead of analysing the actual land use/land cover change. The study found that there exist clear differences in overall accuracy and Kappa values between different classification algorithms. This study also established clearly that machine learning algorithms outperform the traditional maximum likelihood classifier. By exploring the influence of the training sample size, it is established that to achieve accurate classification, the main area of concern is the quality of training data.

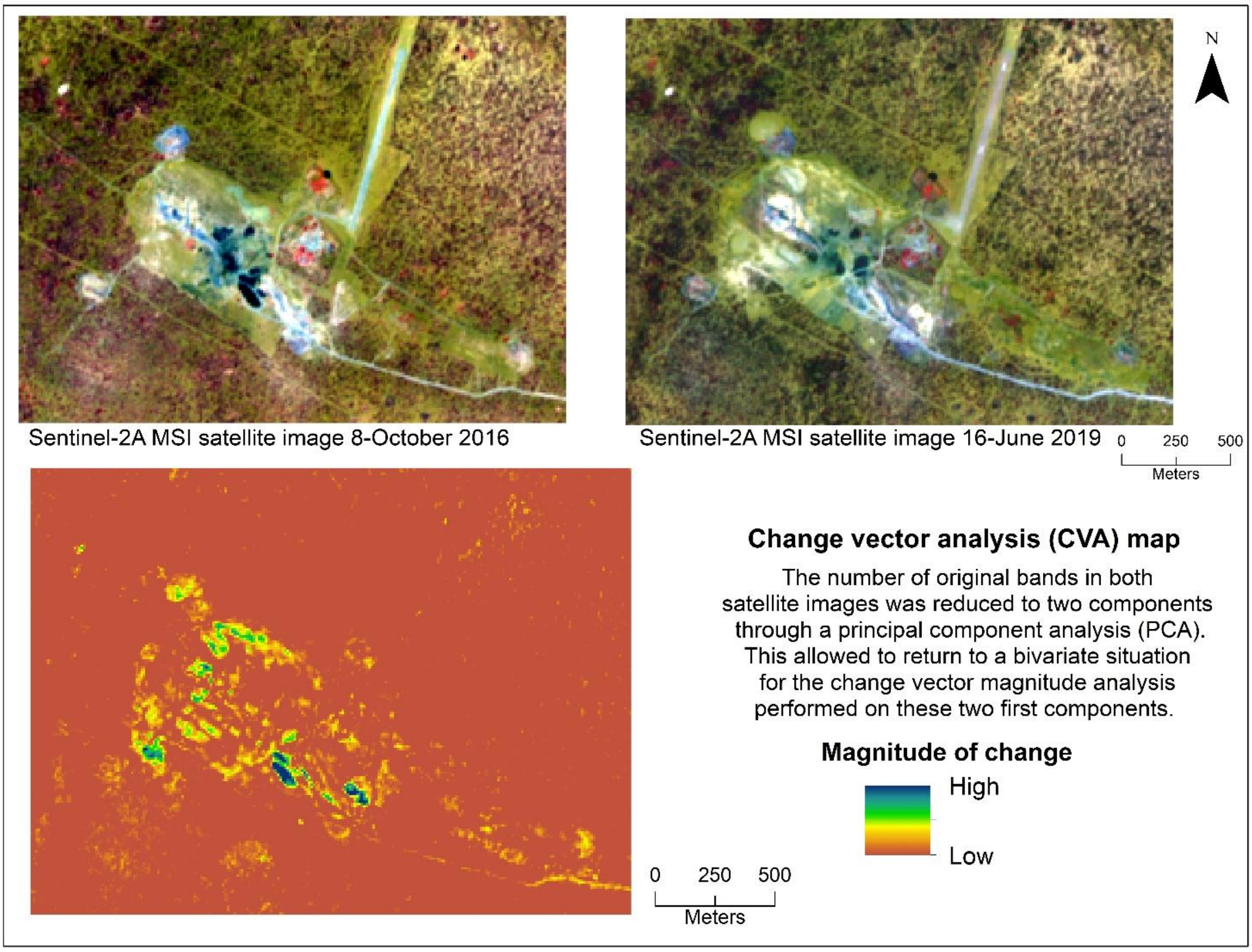

Figure 7. Sentinel 2A images of the study area for 2016 and 2019 and Change vector analysis (CVA) map showing the magnitude of change between the two images.

In future studies, we recommend using R software as it produced the most accurate results in this case study. As this study adopted the default parameters in running all the algorithms, improved results could be achieved by fine-tuning the classification algorithms. This is a ready possibility in the open-source R statistics software as opposed to ArcGIS, which is more limited in this aspect. We also recommend that object-based image analysis (OBIA) classification methods should be explored alongside pixel-based classification methods to gain the most accurate classification results for mining area mapping and monitoring.

Author contributions

The authors of this paper worked together under TAITAGIS, a Finnish-Kenyan integrated Geoinformatics education and training project (from Certificate Level in GIS to postgraduate specialisation pathways in GIS for Environmental Monitoring and GIS for Mining & Geology). The project is hosted at Taita Taveta University, Kenya, and draws students from the greater East African region. Mika Siljander and Nashon Adero coordinated the overall conceptualisation, structure, drafting of the paper, and the associated quality control. Mika Siljander also pre-processed the data and performed the classification analyses. As part of his PhD research on mine planning, Nashon Adero also provided an overview of developments in the mining sector and linkages to SDGs. Francis Gitau, a CEMEREM Scholar, pursuing his MSc in Mining Engineering conducted most of the accuracy assessments of image analysis, which produced results for discussion and provided additional literature to inform detailed interpretations. Emmanuel Nyambu conducted most of the field data collection from the actual mining site, ground truthing, and mapping of the study area as part of his MSc research under TAITAGIS.

Acknowledgements

The authors would like to acknowledge the funding from the Ministry for Foreign Affairs of Finland for the TAITAGIS project coordinated by the Finnish National Agency for Education, HEI-ICI programme (http://cimo.innofactor.com/programmes/hei_ici_index/programmes/hei_ici/projects/taitagis). Taita Taveta University and CEMEREM is acknowledged for providing the enabling research environment and scholarly platform for presenting these findings for further reviews towards future improvements.

REFERENCES

[1] UNSSC Knowledge Centre for Sustainable Development, 2019. "The 2030 Agenda for Sustainable Development, " Accessed June 2019 from https://www.unssc.org/sites/unssc.org/files/2030_agenda_for_sustainable_development_kcsd_primer_en.pdf
[2] A. Union, "African Union Linking Agenda 2063 and the SDGs, 2019," 2019. https://au.int/en/agenda2063/sdgs.
[3] GOK. "The Kenya Vision 2030.," *Gov. Repub. Kenya*, 2007.
[4] N. J. Adero, "Operationalizing progressive mining regulations: Lessons from Kenya on integrating geospatial models into system dynamics," in *2nd International Convention on Geosciences and Remote Sensing, Las Vegas, USA*, 2017, p. 34.
[5] N. J. Adero, "Redressing the nexus of human rights and mining in Kenya using geospatial models," in Proceedings of the BHT 2018: Future Materials – Safe Resources Supply – *Circular Economy and the 3rd Interdisciplinary Colloquium and PhD Conference on the Social Responsibility of Science and Scientists*, Freiberg, Germany, pp. 12–13.
[6] N. J. Adero, "Optimising mine planning: Applying the integrative power of spatial models and system dynamics under divergent land-use scenarios," *in 2nd International Conference of the Regional Centre for Mapping of Resources for Development (RCMRD) on "Space Science for Sustainable Development*, Nairobi, Kenya, 15–17 August, 2018b, p. 68.
[7] M. A. Mwakumanya, M. Maghenda, and H. Juma, "Socio-economic and environmental impact of mining on women in Kasigau mining zone in Taita Taveta County," *J. Sustain. Min*., 2016, doi: 10.1016/j.jsm.2017.04.001.
[8] A. T. Lima, K. Mitchell, D. W. O'Connell, J. Verhoeven, and P. Van Cappellen, "The legacy of surface mining: Remediation, restoration, reclamation and rehabilitation," *Environ. Sci. Policy*, vol. 66, pp. 227–233, 2016, doi: https://doi.org/10.1016/j.envsci.2016.07.011.
[9] R. Latifovic, K. Fytas, J. Chen, and J. Paraszczak, "Assessing land cover change resulting from large surface mining development," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 7, no. 1, pp. 29–48, 2005, doi: https://doi.org/10.1016/j.jag.2004.11.003.
[10] M. Kamga et al., "Evaluation of land use/land cover changes due to gold mining activities from 1987 to 2017 using landsat imagery, East Cameroon," *GeoJournal*, May 2019, doi: 10.1007/s10708-019-10002-8.
[11] Y. Vasuki, L. Yu, E.-J. Holden, P. Kovesi, D. Wedge, and A. H. Grigg, "The spatial-temporal patterns of land cover changes due to mining activities in the Darling Range, Western Australia: A Visual Analytics Approach," *Ore Geol. Rev.*, vol. 108, pp. 23–32, 2019, doi: https://doi.org/10.1016/j.oregeorev.2018.07.001.
[12] J. Mukherjee, J. Mukherjee, D. Chakravarty and S. Aikat, "A Novel Index to Detect Opencast Coal Mine Areas From Landsat 8 OLI/TIRS, " *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(3), 2019. 891–897.
[13] N. Demirel, M. K. Emil, and H. S. Duzgun, "Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery," *Int. J. Coal Geol.*, vol. 86, no. 1, pp. 3–11, 2011, doi: https://doi.org/10.1016/j.coal.2010.11.010.
[14] A. J. Miller and N. P. Zégre, "Mountaintop removal mining and catchment hydrology," *Water* (Switzerland). 2014, doi: 10.3390/w6030472.
[15] F. Nippgen, M. R. V. Ross, E. S. Bernhardt, and B. L. McGlynn, "Creating a More Perennial Problem? Mountaintop Removal Coal Mining Enhances and Sustains Saline Baseflows of Appalachian Watersheds," *Environ. Sci. Technol*., 2017, doi: 10.1021/acs.est.7b02288.
[16] D. L. Mummey, P. D. Stahl, and J. S. Buyer, "Microbial biomarkers as an indicator of ecosystem recovery following surface mine reclamation," *Appl. Soil Ecol.*, vol. 21, no. 3, pp. 251–259, 2002, doi: https://doi.org/10.1016/S0929-1393(02)00090-2.
[17] T. Werner, A. Bebbington, and G. Gregory, "Assessing impacts of mining: Recent contributions from GIS and remote sensing," *Extr. Ind. Soc.*, vol. 6, Aug. 2019, doi: 10.1016/j.exis.2019.06.011.
[18] P. K. Srivastava, D. Han, M. A. Rico-Ramirez, M. Bray, and T. Islam, "Selection of classification techniques for land use/land cover change investigation," *Adv. Sp. Res.*, vol. 50, no. 9, pp. 1250–1265, 2012, doi: https://doi.org/10.1016/j.asr.2012.06.032.
[19] D. Garai and A. C. Narayana, "Land use/land cover changes in the mining area of Godavari coal fields of southern India," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, no. 3, pp. 375–381, 2018, doi: 10.1016/j.ejrs.2018.01.002.
[20] P. L. Basommi, Q. Guan, and D. Cheng, "Exploring Land use and Land cover change in the mining areas of Wa East District, Ghana using Satellite Imagery," *Open Geosci.*, vol. 7, no. 1, pp. 618–626, 2015, doi: 10.1515/geo-2015-0058.
[21] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens*., vol. 66, no. 3, pp. 247–259, 2011, doi: https://doi.org/10.1016/j.isprsjprs.2010.11.001.
[22] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random Forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006, doi: https://doi.org/10.1016/j.patrec.2005.08.011.
[23] D. Rocha *et al.*, "Random Forests in the Supervised Classification of Multidimensional Images of the Tetrazolium Test," *J. Agric. Sci.*, vol. 11, p. 115, Sep. 2019.
[24] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *International Journal of Remote Sensing*. 2018, doi: 10.1080/01431161.2018.1433343.
[25] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, 2016, doi: 10.1016/j.gsf.2015.07.003 [26] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, 2016, doi: 10.1016/j.gsf.2015.07.003.
[26] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, 2016, doi: 10.1016/j.rse.2016.10.010.
[27] Kenya National Commission on Human Rights. "Public Inquiry Report on Mining and Impact on Human Rights: Taita Taveta County,

" 2017. Nairobi: Top Quality Converters Co., 1–152. ISBN: 978-9966-040-43-5.

[28] S. M. Maranga, A. K. Bett, and K. Ndeto and G. Bett, "Kenyan Iron Ore: Mining and prospects of Processing," *In Proceeding of the International Conference on Science and Technology of Ironmaking and Steelmaking*, December 16-18, 2013, CSIR-NML Jamshedpur.

[29] P.K., Keoko, G. K. Bett, K. Gilbert, B. K. Rop," The Implications of Geology and Structure on Iron Ore Mining at Wanjala Mine , Taita-Taveta County, Kenya." Proc. Sustain. Res. Innov. Conf. 6 - 8 May, 2015 70–73.

[30] F. Gitau and N. Mutua, "Spatial Modelling of Dynamics of Land Use Land Cover Due to Mining Activities in Taita Taveta County." p. 24, 2019.

[31] M. Mwandawiro, "Mining in Taita Taveta County: Prospects & Problems," Heinrich Böll Stiftung, East and Horn of Africa All. 2011.

[32] J. Obiri, "Extractive Industries For Sustainable Development In Kenya Final Assessment Report," United Nations Development Programme (UNDP), Nairobi, Kenya, 2014.

[33] GOK. "Parliament of Kenya, Senate.," *The Hansard*., 41(2), 2019, 1–35.

[34] Google Earth Pro 7.3.2.5776, May 14, 2016. Wanjala 03°15'01.03''S, 38°10'43.96''W, Eye alt 1.5 km Image@2019 CNES, Airbus, viewed August 27, 2019. http://www.google.com/earth/index.

[35] M.-W. Uwe, L. Jerome, R. Rudolf, G. Ferran, and N. Marc, "Sentinel-2 Level 2a Prototype Processor : Architecture , Algorithms and First Results," *ESA Living Planet Symp*. 2013, Edinburgh, UK, 2013.

[36] B. Leutner, N. Horning and J. Schwalb-Willmann, "RStoolbox: Tools for Remote Sensing Data Analysis. R package version 0.2.6.", 2019, https://CRAN.R-project.org/package=RStoolbox

[37] R Core Team, "R Core Team (2019). R: A language and environment for statistical computing.," R Found. Stat. Comput. Vienna, Austria. URL http//www.R-project.org/., 2019.

[38] O. Hagner and H. Reese, "A method for calibrated maximum likelihood classification of forest types," *Remote Sens. Environ*., 2007, doi: 10.1016/j.rse.2006.08.017.

[39] P. M. Mather, "A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data," *Int. J. Remote Sens.*, vol. 6, no. 2, pp. 369–376, Feb. 1985, doi: 10.1080/01431168508948456.

[40] V. Vapnik, "Statistical learning theory. " New York: John Wiley & Sons, 1998.

[41] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens*., 2002, doi: 10.1080/01431160110040323.

[42] L. Breiman, "RANDOM FOREST," *Mach. Learn*., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004.

[43] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016, doi: https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[44] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, 2002.

[45] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.

[46] S. S. Rwanga, and J. M. Ndambuki, "Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS." *International Journal of Geosciences*, 08(04), 2017. 611–622. https://doi.org/10.4236/ijg.2017.84033

[47] C. Huang et al., "Use of a dark object concept and support vector machines to automate forest cover change analysis," *Remote Sens. Environ.*, 2008, doi: 10.1016/j.rse.2007.07.023.

[48] T. G. Whiteside, G. S. Boggs, and S. W. Maier, "Comparing object-based and pixel-based classifications for mapping savannas," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 13, no. 6, pp. 884–893, 2011, doi: 10.1016/j.jag.2011.06.008.

[49] Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogramm. Eng. Remote Sensing*, 2006, doi: 10.14358/PERS.72.7.799.

[50] P. Hurskainen, H. Adhikari, M. Siljander, P. K. E. Pellikka, and A. Hemp, "Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes," *Remote Sens. Environ*., 2019, doi: 10.1016/j.rse.2019.111354.