UNIVERSITY OF THE WITWATERSRAND

JOHANNESBURG

Title:	**Prospecting archaeological mining sites using high resolution imagery and advanced classification algorithms in Ziwa, Zimbabwe.**
Author:	**Andile Heritage Fani**
Student number:	**817580**
Supervisors:	**Professor Robert Thornton**
	Dr Stefania Merlo

Master report in Geographic Information Systems and Remote Sensing by coursework and research report

Abstract

Prospection of archaeological sites using remote sensing and machine learning technology is premised on the fact that the spectral signature of a similar archaeological feature is equivalent over a landscape defined by an archaeologist. Therefore, known sites can be used for the identification and location (prospection) of unknown sites. Archaeological prospection has been applied over different types of landscapes, initially with the aid of traditional field surveying, kites, aerial photography, satellites, and recently through drones mounted with high resolution sensors. Use of high-resolution satellite imagery from multispectral sensors on low earth orbit vehicles has taken the lead as it has scored success in fulfilling certain objectives in specific areas through its extensive surveying abilities and reduced costs in comparison with others. This study focused on the ability of advanced classification algorithms Random Forest (RF) and Support Vector Machine (SVM), methods that have proved effective for Iron Age (IA) archaeological sites in this region, and for prospecting at the archaeological site of Ziwa, an iron age site dated between 1500 and 1700A.D. found in the north-eastern highlands of Zimbabwe using free Sentinel-2 imagery. The performance of these machine-learning classification algorithms was assessed in order to ascertain which model performed better. Two independent accuracy assessments were done, one for SVM and another for RF using randomly selected test data that were collected at the Ziwa site. The results indicated a very high potential for using remote sensing methods in prospecting for archaeological mining sites as has been the case with the identification of kraals and vitrified dung (Thabeng *et al*. 2019). For the first set of classifications, satisfactory overall accuracies of 80.67% with a Kappa coefficient of 0.78, and 80.17% with a Kappa coefficient of 0.7724 were achieved for RF and SVM respectively. On comparing the performance of the two individual classifiers using the MacNemar's test, a z value of 1.692 was obtained with a p value $\leq$ 1.96. This indicated that there was no significant difference between the two classifiers in discriminating the land use and land cover (LULC) classes built-up-area, bare-land, cultivated fields, enclosures, grassland, metalliferous-surface, non-metalliferous-surface, pit-structures, rock-outcrop, terraces, tarred road, water body, wooded vegetation and stone walls (terraces, pit structures and enclosures). In the second set of classifications, satisfactory overall accuracies 80.25% with a Kappa coefficient of 0,77, and 80.62% with a Kappa coefficient of 0.78 were achieved for RF and SVM respectively. On comparing the performance of the two individual classifiers using the MacNemar's test, a z

value of 2.289 was obtained with a p value ≥ 1.96. This indicated that after aggregating classes there was a significant difference between the performance of the two classifiers in discriminating the land use and land cover (LULC) classes. The ore grinding sites (approximately 60 cm in diameter), smelting areas/furnaces (<1-1.5m in diameter) and slag (approximately 3-5cm in diameter) however could not be detected due to their size falling way beyond Sentinel-2 10m resolution. To address this problem, this study recommends use of LiDAR or finer resolution imagery in combination with object-based classification as this has the ability to segment very small features with distinct textures, shape and signatures. While RF and SVM clearly indicated the landscape of the archeological interest (Area of interest 'AOI' around 36 km^2) within the broader investigated area (surroundings of the AOI) and could discriminate between metalliferous surfaces and stonewall structures, it however did not precisely discriminate the individual archaeological features such as pit structures, enclosures and terraces within the 36 km^2 under investigation. Remote sensing based archaeological prospection may not guarantee discrimination of the finer details of archaeological remains but may be used to understand location of a possible site. Furthermore, archeological prospection using the pixel-based methods has shown the ability to discriminate features of different physical composition (metalliferous surface and stonewalls) while it falls short in discriminating different features made from the same material such as pit-structures, enclosures and terraces.

Table of contents

List of Tables

List of Figures

1 Introduction

Cultural landscapes are defined as the output of culture-as-an agent working on a landscape medium and human activity imbued with cultural values (Saur 1925; Taylor 2011). That is to say, cultural landscapes are the tangible and intangible heritage including living and non-living features which identify a group of people and the interaction with their environment (Schaich *et al*. 2010).

The documentation of cultural landscapes is an important part of heritage preservation and contributes to the safeguarding of a country's national identity (Soovali *et al*. 2003a; Remondino & Rizzi 2010). Documentation may be undertaken for a variety of reasons including protection, restoration, conservation, preservation, identification, monitoring, interpretation, management of historic buildings and cultural landscapes or sites (Haddad 2010; Remondino & Rizzi 2010).

Haddad (2011), noted that documentation tools for archaeological sites have changed significantly from the traditional manual and analogue methods to digital methods. The new methods of documenting archaeological sites include introduction of remote sensing technology that brought forth machine-learning algorithms that could be used in Land Use and Land Cover (LULC) mapping and archaeological surveying (Haddad 2011). LULC, a category mostly used in remote sensing applications, is defined as the physical composition and characteristics (e.g., grass, forest, and impervious surfaces) or human-related activities (e.g., residential, commercial, and transportation) of land elements on the Earth's surface (Cihlar 2000). Land cover refers to the type of feature present on the surface of the land. It refers to a physical property or material e.g. water, sand, potato crop, or asphalt. Land use relates to the human activity or economic function for a specific piece of land, e.g. urban use, industrial use or nature reserve. Another way to put it is that land cover is more dynamic than land use. Most vegetated areas (land cover) change over time. Land use as it is related to the human activity is more stable (Kerle *et al*. 2004).

Archaeology has experienced a massive growth in the adoption of remote sensing and GIS techniques in research and in the prospecting and documentation of cultural sites. Remote sensing in archaeology began with poles, kites and hot air balloons that were used to take aerial photographs (Campana 2017). Later, helikites made their debut (Campana 2017), together with drones, unmanned aerial vehicles (UAV), and satellites (Forte 2010; Sadr 2012; Campana 2017).

Satellite images have been used to identify crop marks, which are crucial in detecting buried archaeological sites and new unknown sites as demonstrated by Agapiou *et al*. 2014. Furthermore, Luo e*t al*. 2019 highlights that Remote sensing techniques have contribute to the photography of archaeological sites (photograeochology), cultural heritage monitoring and conservation through the use of multispectral images (spectroarchaeology) and semi-automatic detection, interpretation and visualization of archaeological sites. Use of hyperspectral data in conjunction with advanced algorithms is geared for future archaeological prospection (Giardino 2011). In Africa, there is still a great need for the use of (semi)automated remote sensing methods stemming from machine learning developments through aerial and spaceborne RS technologies This will prove vital to expanding the knowledge base of Africa's archaeological record (Davis & Douglas 2020; Klehm & Gokee 2020).

In this research, remote sensing was used to classify land use at an archaeological mining site in Ziwa, Zimbabwe. This is important because archaeological mining sites in the Ziwa region have been destroyed by anthropogenic activities such as fire-wood harvesting and artisanal mining (Chipangura *et al*. 2013). Documenting and prospecting of other similar sites will contribute to the protection and discovery of other potential similar sites nearby (Chipangura 2013) and will put the site on a global map in order to promote ecotourism, contributing directly to UNESCO goals.

Furthermore, archaeological landscapes allow orientation of a culture within space and time. The area of interest is recognized as a spiritual, symbolic and sacred area in which contacting ancestral spirits and petitioning for rain, and other rituals are performed (Chipangura *et al*. 2017). Prospecting in this archaeological terrain will result in the discovery of similar sites where people may continue practicing their religion and safeguard a culture that is slowly being forgotten or eroded over time.

Documentation is essential for various other heritage conservation activities including: protection and management of the site, restoration, conservation, preservation, identification, monitoring, interpretation, management of historic buildings and sites and cultural landscapes (Remondino & Rizzi 2013; Haddad 2010).

Most importantly the mining sites are important to Zimbabweans in general because they showcase the technological achievements of earlier cultural contexts.

1.1 Problem statement

Earlier methods of documenting and prospecting of archaeological landscapes proved labour intensive and costly (Suerbier & Eisenbeiss 2010). Moreover, data quality, accuracy and flexibility of these methods has not been entirely satisfactory. More flexible and efficient methods such as remote sensing have, therefore, been explored (Haddad 2011; Breeze *et al*. 2015). These include the use of multispectral imagery in the mapping and prospecting of archaeological sites for further investigation (Agapiou *et al*. 2014; Immitzer *et al*. 2016). While multi-spectral remote sensing was originally developed to scan geological surfaces for metals, hydrocarbons, forestry and agriculture, they have proved useful for archaeology too (Parcak 2009).

Multi-spectral images provide a broad range of spectral bands at varying resolution. This makes the detection of diverse archaeological features possible. (Agapiou *et al*. 2014). Furthermore, multi-spectral images can be classified using advanced algorithms yielding higher accuracies. The computing of a wide range of spectral indices as well as band ratios is made possible due to the presence of multiple bands. This in turn enables the detection of potential archaeological features in varying environments. This makes multispectral imagery ideal in Archaeological applications (Agapiou *et al*. 2014).

My research used remote sensing techniques for the mapping and prospecting of archaeological mining sites in Zimbabwe that existed during the Iron Age (1500-1700 AD). Features to be identified included pit-structures, metalliferous surface, slag, furnaces, enclosures and terraces. The research focused on using machine-learning classification techniques specifically, Random Forest (RF) and Support Vector machines (SVM) applied to freely available and relatively high Sentinel-2 imagery in order to map an archaeological mining site. Machine-learning techniques have been used on other types of archaeological sites previously, for example the identification of habitation sites, identification of kraals and dung deposits in the Shashe-Limpopo confluence area in northern Limpopo Province (Thabeng *et al*. 2019). It has never been attempted for mining sites in Zimbabwe.

This approach showed great potential in saving costs, improving accuracy, time efficiency and allowed us to assess which algorithms produced the best results for prospection of archaeological sites with limited time spent in the field.

1.2 Aim and objectives

This research sought to use remote sensing to prospect the remains of metallurgical activities in the area of Ziwa, Nyanga landscape, Manica Province in north-eastern Zimbabwe using advanced classification algorithms and Sentinel-2 imagery.

The objectives of the research were as follows:

1. To use Sentinel-2 free multispectral imagery and advanced supervised classification to construct a LULC map of the Ziwa landscape for the detection of archaeological features, including: pit-structures, metalliferous and metal-depleted surfaces, slag, furnaces, enclosures and terraces;
2. To compare the accuracy of support vector machine (SVM) and random forest (RF) classifier algorithms in classifying the various archaeological components of the Ziwa landscape in order to establish the most efficient, algorithm in archaeological mine prospection as well as finding more cost and time efficient ways to complete various forms of archaeological prospection.

1.3 Research questions

This project sought to answer the following questions:

1 Are there sufficient and identifiable ‘signatures’ based on differential wavelength reflectance to discriminate between specific types of archaeological terrains related to gold ore preparation and metallurgical activities in the designated landscape including:

   a. Gold ore grinding and processing areas where ore was crushed, sieved and the mineral was extracted using the furnaces giving slag and metalliferous byproducts such as those around identified grinding areas;
   b. Smelting/forging/melt-refining areas and other -temperature technologies surrounding furnaces or fire sites; and
   c. dumping grounds for metallurgical processing waste.

2 Can sufficiently accurate models be constructed based on known, and geo-located sites that can be used to identify (predict) new sites based on the model that has been constructed for the known sites?
3 Comparing Random Forest and Support vector machines, which classification algorithm best discriminates archaeological sites from non-sites?

2 Literature review

2.1 The archaeology of Ziwa

Ziwa is the name given to the archaeological landscape located in the Nyanga area in the north-eastern part of Zimbabwe (Chirikure & Rehren 2004). This research focused on a small area within this terrain of around 36km^2.

Archaeological work in Ziwa has yielded many archaeological findings dating to between 1500-1700A.D.and many researchers have explored the Nyanga regions (e.g. Summers 1958). A British archaeologist, Robert Soper, thought that the unique archaeological landscape at Ziwa was created by previous farming activities because it appeared that the lines of loose stone structures were made to form agricultural terraces on which crops could be grown (Soper 2002; Strassburger, 2015; Chipangura 2019). Terraces were generally narrow with widths between 1.5-3 m in steeper slopes and around 7-10 m in gentle slopes (Soper 2006). The claim has however received great criticism from scholars such as Kritzinger (2008, 2010, 2012, 2017) and Chipangura (2019a). As discussed in section 1 above, the site is of importance for archaeologists and the general public in Zimbabwe for a number of different reasons that range from archaeological knowledge, heritage management and identity building.

Soper built up a strong case for his hypothesis that Ziwa terrain as built by a prehistoric agricultural society. His archaeological work between 1993 and 1999 res ulted in the archaeological monograph *Nyanga: Ancient fields, settlements, and agricultural history in Zimbabwe* (Soper 2002a). He used aerial photographs, foot surveys and traditional excavation (Soper 2002a; Soper, 2006). He argued that the stone structures in the area were used by a people who were probably ancestral to the current Shona people in the area that occupied the highlands from 1500 to approximately 1700 A.D. (Soper, 2006). Soper claimed that the stone structures of Nyanga were the largest complex of ancient buildings in sub-Saharan Africa (Soper 2002a).

Kritzinger was among the first to challenge the notion the Ziwa site is an agricultural site. Contrary to previous researches, she suggested that the site was probably built in connection with gold mining rather than the growing of crops and rearing of livestock (Kritzinger 2010). Kritzinger supported her position through the discovery of slag scattered around the area as well as presence of metalliferous deposits at the bottom of the hill which could have been mine waste (Kritzinger, 2010). Agropastaralism could not be completely ruled out in this site according to the work presented by Soper (2002). The Ziwa site is located on a steep rocky terrain which generally lacks extensive wethering to produce deep soils. (Soper 2011). Thus, this environmnet would lead to poor soils which do not retain any minerals, theirfore according to Soper (2011), this could be used as evidence against the agricultural hypothesis. However, presence of terraces Mining and Agropastoralism could have existed together but this is still an uresolved debate.

She developed a strong model depicting how gold mining, processing and refinement could have been related to the site and its features (Kritzinger 2010; Chipangura 2019a). The terraced area within the site has been thought previously to have been used for crops, while the pits were used as cattle pens for miniature cattle, and the hard granite surfaces were thought to have been used as grinding sites for grains and seeds (Soper 1996; Kritzinger 2008). Ziwa was thought to make up part of the early Southern African Agricultural complexes, which makes it one of the key areas where farming and rearing of livestock was an important component of the community's economy. The idea held by Soper that the archaeological terrain is the result of modification for agriculture and pastoralism is to some extent justified (Chirikure & Rehren 2004),

Kritzinger (2012) and Chipangura (2018) have recently also shown that the ancient landscape of Ziwa is probably an ancient mining site which existed between 1500 and1700A.D. They supported the argument by showing that depressions on granite outcrop rocks within the sites were probably ore grinding platforms which were manmade depressions on the rock surface as a result of intense grinding of ore using grinding stones (Kritzinger 2012a). The evidence of this is the discovery of a large quantity of quartz chips and quartz debitage around the grinding sites which are angular and probably the result of mechanical fracture from grinding (Strassburger 2015). Furthermore, crucibles and furnaces were discovered within the site that were used for smelting the ore (Kritzinger, 2012).

There is evidence supporting the existence of both the mining and agricultural hypotheses on whether the Agricultural or mining interpretation is more supported than the other (or if both exist),is yet to be resolved. A metallurgical waste dump was formed when ground ore was discarded in valleys between ridges of rock. Drainage of rainwater caused an 'erosion delta' that has ridges and rivulets that form striations as heavier and lighter particles separated during erosion (Thornton & Chipangura 2017). This has created an erosion fan or delta within the Ziwa mountain site. Parts of these deltas are being mined by artisanal miners today (Thornton & Chipangura 2017). Ongoing research hypothesises that the deltas and striations represent metallurgical mine waste from gold and/or iron smelting and other metallurgical processing (Thornton & Chipangura 2017).

The debate amongst different scholars still exists (Kritzinger 2008; Chipangura *et al*. 2017). Chirikure & Rehren (2004) support the notion that Ziwa was a rich metallurgical site as distinguished by its gold and iron mining features, while continuing to concur with the agricultural hypothesis. They supported this notion on the basis of the discovery of remnants of slag, ores and furnace fragments within the site (Soper 2003; Chirikure & Rehren 2004). This debate is still ongoing

The well-constructed pits display the presence of drain inlets and drain outlets on the floor of the pits which have tested positive for heavy metal concentrate (Kritzinger 2012).

The most obvious features of the site are the terraces, which are made up of low lying loose, dry-stone 'walls' which are built on hillsides spreading along hundreds of kilometers north of Nyanga (Mupira 2011). These were explored by F. O. Bernhard, on his farm around 40km north of the Ziwa monument (Bernhard 1959). Other archaeologists also did work in the area including de Villiers (1969), Burrett (2006), and Mupira (2011).

There are large numbers of depressions on the granite surfaces which Soper (1996) thought were grinding sites/platforms for crops. (Strassburger 2015; Kritzinger *et al*. (2008); Kritzinger (2012a).

Experiments on the chemical components of the ores and slags showed significant levels of iron and gold (Chirikure & Rehren 2004). Additionally, crucibles and furnaces were discovered within the site and these were possibly thought of having been used for smelting the ore (Kritzinger 2012).

The archaeology of the area also includes technical ceramics that is, ceramics used for technical processes and not for home use (Chirikure & Rehren 2004) Pronounced representations of female figures on the furnaces within the site, such as breast and waist features embossed on the furnace structures (Chirikure & Rehren 2004) were a figurative way of likening the iron smelting work to pregnancy and the precious processed metals being the offspring (Chirikure & Rehren 2004). There is yet to be a conclusive answer on whether the agricultural versus the mining hypothesis is correct or if both exist.

2.2 A brief review of remote sensing in archaeology

2.2.1 Remote sensing in archaeological prospection

Remote sensing (RS) techniques have been used for a variety of applications in archaeology from monitoring (Parcak, 2007), management (Reid, 2016), surveying (De Laet *et al*. 2007), 3D scanning of features (Remondino & Rizzi 2010; Haddad 2011) and prospection (Thabeng *et al*. 2019). Scholars such as Pugin 2016, have even used RS technology for predicting potential rock art sites while Breeze *et al*. (2015) has used RS techniques in archaeological reconstruction of sites. These cases show the unique value of the application and the wide range of contexts in which RS has been used.

Earliest methods such as hand measurements for documenting and prospecting archaeological sites were not accurate and took a lot of time (Haddad 2011). Furthermore, maps often lacked a common coordinate referencing system (Haddad 2011). Data quality, accuracy and flexibility were not entirely satisfactory. Lack of consistent coordinate reference system meant that size of features were usually distorted (Haddad 2011). Furthermore, physical maps were not flexible as the scale on a physical map was static that is, it could not be changed or adjusted and this limited the amount of analysis that could be done. Techniques which were more accurate, quicker and robust were invented and RS partially met these needs (Haddad 2007, Cerra *et al*. 2018). RS made it possible to analyse spectral signatures, archaeological signatures, as well as algorithms together with data emanating from diverse technologies possessing different resolutions (Agapiou *et al*. 2014a). All these have contributed to the prospection, identification, management, and protection of archaeological sites. RS and AI technologies are emerging technologies that have developed over time as satellites, data, and software have become increasingly complex and available (Klehm & Gokee 2020; Luo *et al*. 2019; Lasaponara & Masini 2011).

There are also some challenges with its use as well. RS may be expensive to implement since sensors, drones, and data require as highly skilled personnel (Sauerbier & Eisenbeiss 2010) to operate. Flight technicians, pilots and transportation costs involved in this may not be feasible (Sauerbier & Eisenbeiss 2010, Eisenbeiss *et al*. 2005a). However, RS benefits generally override the costs and challenges.

One does not have to be physically in the field to acquire data because it can access it remotely (Sadr 2016). Physically unreachable areas due to wars and physically barriers can be investigated without being physically on site. RS data may be available free, such as 10 meter resolution Sentinel-2 data, while other data may be more costly, such as, for example, Worldview or SPOT data. The wide range of RS techniques might require less or advanced training and computing power than others. The acquisition and pre-processing of LiDAR data from capture by an UAV requires highly skilled professionals while downloading a recent Google Earth image might need the least amount of training.

De Laet (2007) showed that RS technology is useful in preparing a survey exercise before the field work as it directs the efforts to a more localised and potentially optimal location since observing the archaeological site from a field survey is limiting as the spatial structures and characteristics may not be apparent (De Laet *et al*. 2007).

One of the approaches has been the use of ‘crop and soil markers’ which is usually applied in areas with dense vegetation or distinct soil contrast to the rest of the non-site area and these have been extensively applied (Agapiou *et al*. 2014a; Gojda & Hejcman, 2012). Crop marks are defined as detectable changes in the vegetation regime and growth due to presence of a potential buried archaeological sites (Cerra *et al*. 2018, Thabeng *et al*. 2020). According to Cerra *et al*. (2018), cropmarks can either be negative as the result of stress within the vegetation due to buried site, or positive (Thabeng *et al*. 2020) due to a wet ditch or fertile soil. These changes are picked up through the use of vegetation indices and this is demonstrated by Cerra *et al*. (2018) and Agapiou *et al*. (2014a). Crop marks have been used since the very early days of remote sensing, from kite photography to black and white photographs and over time they have been used in prospection of potential buried archaeological sites with the aid of NDVI. as demonstrated by Cerra *et al*. (2018) as well as Fuldain González & Varón Hernández (2019). Soil marks are differences in humidity or soil filling which present variances with the surrounding surface soil colour (Cerra *et al*. 2018).

The Normalised Difference Vegetation Index (NDVI) has been applied by Cerra *et al*. (2018) in potential archaeological areas covered with dense vegetation and is computed through dividing the difference of the Near Infrared (NIR) and Red bands with their sum and factors that influence these values are outlined by Verhoeven (2012) and Pugin (2016). This calculation is used in the determination of the density of green vegetation on a piece of land. This index involves NIR (750-900nm) and Red (600-690nm) bands from the electromagnetic spectrum where NIR is strongly reflected by the chlorophyll in green plants whereas Red bands are strongly absorbed by the chlorophyll in vegetation. The difference of these two is a value that ranges between -1 to +1, values closer to +1 indicate a high possibility of dense green vegetation cover while values closer to -1 mean there is less green vegetation cover or in some cases there is presence of water. From this it can then be determined whether vegetation cover is sparse, dense or bare. NDVI has been used as outlined earlier for the determination of crop marks. The closer the NDVI value is to +1 the stronger the crop mark signal (Joseph & Jeganathan 2018; Masini & Lasaponara 2012).

Over time RS techniques have evolved due to the emergence of newer generation satellites that have better resolution and coverage. Sentinel-2 satellite data is free but and still useful due to its very high spectral resolution. Having been launched in 2015 by the European Space Agency, it has quickly gained popularity within the RS field as it is free, and is a multispectral sensor having 13 bands at a medium to high spatial resolution of 10 m. In a study to evaluate the potentials of free Sentinel-2 satellite data for archaeological perspectives, crop marks were investigated by Agapiou *et al*.(2014a, b) and he concluded that, the Sentinel-2 sensor better identified and separated crop marks from other classes and performed better than commercial satellites SPOT 2, Worldview 2 (8 bands at 0.46 m resolution), IKONOS (4 bands at 3.2 m) and free Landsat 4 (7 bands at 30 m resolution) and other unknown buried archaeological sites were also apparently revealed. Among the above sensors, Sentinel-2 has the widest range of spectral bands and a spatial resolution of 60 to 10 m and it is free multispectral satellite imagery. This makes it a desirable resource for archaeologists with limited budgets for looking into preliminary prospecting of archaeological sites. Furthermore, the imagery has very high spectral resolution and a medium to high 10m spatial resolution making it possibly to conduct extensive analysis on the data. Sentinel-2 is a constellation satellite meaning it has two satellites that is Sentinel-2A and 2B (Gatti & Bertolini, 2013; Yan *et al*. 2016). An industrial consortium led by Astrium GmbH (Germany) created the satellite system while Astrium SAS (France) was responsible for the multispectral sensor (Gatti & Bertolini 2013). The European space agency launched Sentinel-2A on the 23rd of June 2015 and Sentinel-2B on the 7th of March 2017 near Karurou in French Guiana. The 2 satellites are phased at 180° to each other in a sun-synchronous orbit at a mean altitude of 786km (Yan *et al*. 2016). Sentinel-2 mission is to continue acquiring high resolution (10m) multispectral images that was provided by SPOT (commercial satellite) and USGS Landsat satellites and record Landcover changes and Geophysical variables for the next 12 years (Yan *et al*. 2016). Sentinel-2 has been used in the prospection of archaeological buried sites (Agapiou *et al*. 2014). The multispectral characteristics of Sentinel-2 have been best used to distinguish crop marks compared to other existing satellite sensors as demonstrated by Agapiou (*et al*. 2014). Furthermore, it is a free resources in contrast to other high resolution multispectral options such as Worldview or IKONOS (Yan *et al*. 2016).To what extent, then, can we rely on vegetation indices, crop and soil marks for archaeological prospection and what parameters influences this approach? How do researchers proceed if a particular sensor does not cover the spectrum in which vegetation is sensitive?

Although crop marks have been widely used by researchers such as Agapiou *et al*. (2014), their use has been criticised by authors such as Alexakis *et al*. (2012) and Mills & Palmer (2007). Its results may be misleading or inconsistent because the results fluctuate based on environment as highlighted by Agapiou *et al*. (2013). One of the limitations is that environmental factors can affect the development of crop marks. Moreover, the crop marks may be an indicator of the parent material from which the soil has been derived (Agapiou *et al*. 2014b

Thus, it challenges the reliability of this approach and the results may be misleading to some extent. Winton & Horne (2010) acknowledged that crop marks were a direct result of the geology of the area or merely a phenophases within that period. In the application of this approach for archaeological prospection, all these must be taken into account.

Agapiou *et al*. (2014) utilised spectroradiometric measurements (the amount of radiation energy reflected or absorbed by a feature)as a control, as well as high resolution sensors, such as WV and IKONOS and calibrated field spectroscopy measurements to supplement the research (Alexakis, *et al* 2012; Agapiou *et al*. 2012). Agapiou's team also supplemented the research with the aid of simple ratios in conjunction with vegetation indices which gave better results in his prospection.

Cerra *et al*. (2018), endorsed the Modified Chlorophyll Absorption ratio index because the amount of chlorophyll in the surface of the land is positively correlated with the crop mark values.

Cavelli tested the effectiveness of hyperspectral data for prospection using a multispectral infrared camera with both the infrared and visible spectrums to identified buried archaeological structures in Italy (Cavelli *et al*. 2007). In areas where there was scarce crop marks the method suggested by Cavelli *et al*. (2007) made use of the sub-superficial archaeological structures which presented themselves as variations in the spectral properties of terrain surfaces. These were clearly traceable on the hyperspectral images as tonal anomalies due to the abnormal terrain surface. Using the separation index and Detection indexes to analyse the tonal differences in the different images, Cavelli, *et al*. (2007) found that the VIS – NIR ranges are most important for archaeological detection. Cerra *et al*. (2018) who did a similar study, used hyperspectral data rather than multispectral data (as used by Agapiou *et al.* 2014a) to identify buried archaeological relics.

The RS approaches are not without shortcomings and these may be minimised through supplements such as using higher resolution imagery, hyperspectral data and other recognised indices.

Cerra *et al*. (2018) found that archaeological sites do not exhibit specific spectral characteristics that can be exploited to efficiently detect crop marks in any remotely sensed image as each site presents unique geomorphic characteristics. Moreover, results showed that one site cannot be superimposed on another site due the limitations on crop marks, soil marks and indices. To get the best results, a preexisting map depicting the underground structures before they were buried must be available. Within that map, an optimal number of elements must be related to the features of interest in order for the computed statistics to be meaningful. Moreover, if the preexisting map is incomplete, then the method can be used only to investigate the features present in the incomplete map (Cerra *et al*. 2018). This has thus led to the disadvantage of using crop or soil marks as proxy indicators for presence of an archaeological site as they need to be verified through existing old maps/schematics that depicted that ancient site. In this research the Vegetation indices were not used instead, the spectral signatures of the unique features were considered.

It is evident that without a preexisting full map, a high-resolution image supplementary approaches such as the use of crop or soil marks and a pre-existing old map/schematic, the success of archaeological prospection may be limited. This research did not use indices, crop/soil marks, or higher resolution or hyperspectral data. Due to cost and the limits of an MA-level research project, this study made use of medium resolution multispectral imagery and crop marks, soil marks, and indices were not used.

Importantly, all the above studies have been done outside Africa. A few studies have been done in southern Africa such as Sadr & Rodier (2012); Sadr (2016); Pugin 2016 and Thabeng *et al*. 2019. This research aims to increases the RS footprint in Africa, and to support archaeological work in sites which have not been fully explored. Furthermore, the above researches demonstrate that that one methods may work in one context, but not all

Radar such as SAR and LiDAR, has been used for the prospection of archaeological sites by a number of scholars and unlike the above approaches it is not centered on spectral values and reflectance (Chen *et al*. 2015). This approach has advantages over the pixel-based approaches as it could be used in any weather and in areas where potential sites were immersed in water (SAR) or on sites on the surface but possible in a very dense vegetated area (LiDAR). Radar sensors can be mounted on multiple platforms such as ground based, airborne, UAV and spaceborne platforms. It enables multi-polarization, higher resolution, and can be an active sensor giving it the ability to penetrate vegetation and soil as well as its ability to assess the conditions of the soil such as the di-electric properties of sites. Chen *et al*. (2015) expands on the workings and properties of the SAR Radar Systems and reviews the stages of evolution of this technology. Synthetic-aperture radar (SAR) is an example of a spaceborne RADAR system, and the main features of this technology are the polarization, incidence angle, frequency and ability to penetrate even dense canopy (Lasaponara & Masini 2013). Furthermore, this type of approach has the ability to penetrate dense cloud cover thus could be used in cloudy areas, flooded and glaciated environments. Details on the working of this space-borne technology are outlined by Topouzelis & Psyllos (2012) and Du *et al*. (2015). According to Lasaponara & Masini (2013), the greatest advantages of SAR data is that the backscatter depends on the geometry, structure, orientation, and geophysical properties (Lasaponara & Masini 2013).

Light Detection and Ranging (LiDAR) surveys can be very pricey and costly (Davis & Douglas 2020). This technology entails provision of a 3D surface information of ground features using high frequency active laser beam mounted on any platform, emitting beams that strike the surface of the earth creating a point cloud data (Kuma Mishra & Zhang 2012). LiDAR is best used in areas with very dense vegetation as this technology has the ability to penetrate even areas in the forest floors. The most important portion of this technology involves the time taken for the beam signal (pulse) to be emitted from laser source, strikes the target feature and then returns (backscatter) (Kuma Mishra & Zhang 2012 and Davis & Douglas 2020)). The signal becomes complex as it is affected by the roll, pitch, yaw of the platform and the exact coordinates of the target source and these aspects are explained in detail by (Hesse 2010). This technology operates only when there is a change in height of the earth surface. Surfaces that are smooth cause scattering and hence the returning signal records a lower elevation (Kuma Mishra & Zhang 2012). An important attribute of SAR is that one can map archaeological sites and features in areas with dense vegetation cover, unlike satellite imagery.

For monitoring archaeological sites, radar interferometry has used high resolution and multi polarization. Radar data is very complex to work with and requires a clear understanding of how the conditions of the target surface affect the backscatter. Although high-resolution data is commercially available, the application of this technology in archaeology still needs attention and growth. (Chen *et al*. 2015).

In a pilot study to explore the potential of using non-destructive RS techniques, Keay *et al*. (2014) aimed to improve archaeological mapping by integrating geophysical surveys and aerial photos. Ground Penetrating Radar (GPR) and LiDAR as well as high resolution satellite imagery was utilised. The GPR system made it possible to identify the depth of archaeological deposits in the site to be established whereas Chen's spaceborne SAR system measured only the earth's surface. Furthermore, an electrical resistivity and tomography was utilised. This combination of techniques incorporated all the benefits from the different approaches.

Synthetic-aperture radar (SAR) is an example of a spaceborne RADAR system, and the main features of this technology are the polarization, incidence angle, frequency and ability to penetrate even dense canopy (Lasaponara & Masini 2013). Details on the working of this space borne technology are outlined by Topouzelis & Psyllos (2012) and Du *et al*. (2015). According to Lasaponara & Masini (2013), the greatest advantages of RADAR data is that the backscatter depends on the geometry, structure, orientation, and geophysical properties (Lasaponara & Masini, 2013).

Light Detection and Ranging (LiDAR) surveys can be very pricey and costly (Davis & Douglas, 2020). This technology entails provision of a 3D surface information of ground features using high frequency active laser beam mounted on any platform, emitting beams that strike the surface of the earth creating a point cloud data (Kuma Mishra & Zhang 2012). The most important portion of this technology involves the time taken for the beam signal (pulse) to be emitted from laser source, strikes the target feature and then returns (backscatter) (Kuma Mishra & Zhang 2012 and Davis & Douglas 2020)). The signal becomes complex as it is affected by the roll, pitch, yaw of the platform and the exact coordinates of the target source and these aspects are explained in detail by (Hesse 2010). This technology operates only when there is a change in height of the earth surface. Surfaces that are smooth cause scattering and hence the returning signal records a lower elevation (Kuma Mishra & Zhang 2012).

LiDAR has mostly been used in other parts of the world, although Karim Sadr and his team have had great success in using the technology of discover a "precolonial city" just south of Johannesburg (Sadr 2019a,2019b, 2019c, 2016, 2012; Hunt & Sadr 2014.

2.2.2 Archaeological prospection in Africa

There has been very limited work in the use of remote sensing techniques for archaeological prospection in Africa (with the exception of Davis & Douglas 2020; Thabeng *et al*. 2019; Klehm *et al*. 2019) when compared to the rest of the world and it requires the highest resolution data sets, hyperspectral data, the best software in the market for processing and visualisation (Davis & Douglas 2020). LiDAR surveys, UAV, GPRs are expensive but have been adopted by the rest of the world in facilitating archaeological prospection (Klehm *et al*. 2019). Limited training of potential archaeologists constrains technical proficiency required to execute some RS jobs. For instance, when compared to optical imagery, SAR data processing is characterized by high complexity in data processing interpretation and Africa has not kept pace with the use of UAVs, SAR, and LiDAR data for archaeological remote sensing (Lasaponara & Masini 2013; Opitz & Herrmann 2018).

There is however free and open source data that can be used to conduct archaeological prospection. Indeed, there has been a sharp increase in the awareness remote sensing methods which have enabled research to grow (Opitz & Herrmann 2018).

The beginnings of archaeological prospection and surveying are thought to have started in the early 1950s in most parts of the world (Sadr 2012). Early aerial surveys were done in archaeological site prospection using aerial photographs (Saumagne 1952; Key *et al*. 2014). Initially the aerial photographs were used for military mapping and surveying purposes and this technology made it to the archaeological fraternity only after WW2 and as early as 1960s and 1970s in the eastern half of Southern Africa (Leisz 2012; Sadr 2012).

In southern Africa early archaeological work involving prospection of sites was as early as the mid 1900s and involved use of photographs (Mason 1968; Seddon 1968). One of the applications of aerial photographs in archaeological prospection was the identification of Iron Age sites in Botswana (Denbow 1979). For instance, vegetative patterns in aerial photographs were observed to get a better understating of hilltop settlements (Denbow 1979; Huffman, 1986). MacQuilkan & Sadr (2010), compared Google Earth and aerial photographs of ancient stone wall ruins in South Africa. Although aerial photographs had a better resolution for the 25km^2 site than the Google Earth images, the Google Earth images could be viewed in oblique or vertical views. Aerial photographs are static and not freely available, and therefore the use of free imagery with a better resolution was used to enable sufficient resolution for studying the mining and metallurgical landscape.

Another trend in African archaeology is the focus on mapping geomorphic properties of landscapes with the aid of high-resolution satellites in a study of human occupation and landscape reconstruction in the Sahara (Davis & Douglas, 2020; Biagetti *et al*. 2017)

Use of automated modern algorithms as well as high resolution data has proved its worth (Thabeng *et al*. 2019; Klehm *et al*. 2019). LiDAR and hyperspectral data have been used lately in classification using OBIA. Radar is where archaeology in Africa is going (Klehm *et al*. 2019), however, it was not used in this particular research due to high costs involved instead free Sentinel-2 data was used.

2.3 Classification techniques in remote sensing

A complete and effective classification should include a suitable classification system, a representative training sample, and must present an acceptable classification accuracy. This is crucial in archaeological prospection because it ensures a more reliable, spatially and spectrally correct ground representation for effective analysis with results within an acceptable error margin (Lu & Weng 2007). Classification is achieved through two types of classification methods: unsupervised and supervised classification. The former is employed when there is absence of ground or field data while the latter is achieved through the use of training data sourced from the site (Aburru & Golla. 2007). Pixel-based classification algorithms and a confusion matrix are used to assess the performance of the algorithm (Breeze *et al*. 2015). Pixel based classification is considered to perform best in high resolution imagery such as was the case in De Laet *et al*. (2007) where IKONOS imagery was used. According to Biagetti *et al*. (2017) pixel-based image methods are limited and do not consider spatial photo interpretative elements such as texture, context and shape. Object based classification takes into consideration not only the spectral signature of the feature but also the tone, texture, shape, size orientation and spatial relationship of features during the classification process. (De Laet *et al*. 2007). The tools and processes of how object-based classification is done is outlined by De Laet *et al*. (2007). One problem with object-based classification is that most archaeological remans lack uniform shape unlike the modern features that we have (De Laet *et al*. 2007). Ancient archaeological features may be disoriented, eroded, vandalised or damaged due to natural causes. This may affect the texture, shape and orientation. For structures that are still intact OBIA may be used. De Laet's team used a multi spectral medium-resolution imagery and pixel-based classification was chosen.

2.4 Traditional and Advanced classification algorithms

In doing archaeological prospection, two major classifications have been used: traditional classifiers and modern advanced classifiers. Examples of traditional classifiers include Maximum likelihood (MLC), Artificial neural networks (ANN), K means, Minimum distance and Iterative Self-Organizing Data Analysis Technique (ISODATA), Parallelepiped and Spectral angle mapper, which were employed in a number of studies (e.g. Butt *et al*. 2015; Agapiou & Sarris 2018; Abburu & Golla 2015; Keeney & Hickey 2015; Mather & Tso, 2009). Detailed outline on the workings of the above classifiers have been published (Lu & Wang *et al*. 2007; Abburu & Golla 2015). Advanced Modern classifiers are not limited to Support Vector Machines and Random Forests and have been employed by scholars such as Thabeng *et al*. (2019). Traditional classifiers although relatively easier to implement when compared to modern advanced classifiers have been some shortcomings.

2.4.1 Traditional classification algorithms

In general, the traditional classifier MLC performs optimally only when the remotely sensed data has a Gaussian statistical distribution (Chutia *et al*. 2015; Biagetti *et al.* 2017). These classifiers also often assume a normal data distribution (Belgiu & Dragut 2016; Liu *et al*. 2011). Because data obtained remotely rarely possesses the above statistical properties, their performance and accuracy can easily be flawed (Biagetti *et al*. 2017). Traditional classifiers assume that all remotely sensed data has a gaussian distribution (Lu & Wang 2007). However, this attribute is not true for most archaeological sites as they can be more complicated than this as is the case with the Ziwa archaeological site.

Ziwa is a very complex landscape having multiple stone structures and has undergone intensive erosion (Thornton & Chipangura 2017). Thus, the above shortcomings would not be desirable for such a complex landscape needing a robust classifier which will not overfit. Data sets obtained using RS technology are hardly normally distributed thus traditional algorithms do not consider that archaeological data does not usually occur in a 'normal distributed pattern'

Another major issue with traditional classification algorithms is that they are affected by high input sample dimension thus leading to overfitting and poor performance and this was experienced by Tuia *et al*. (2011) who compared SVM and traditional kernel methods of classification. In the study done by Tuia *et al*. (2011), SVM achieved overall accuracy of 76.17% and an estimated kappa statistic of 0.711. In a survey of image classification methods, it was found that MML presented a 'noisy' classification as the classifier overfitted and therefore performed erratically (Lu & Wang 2007). Traditional classifiers have been found to be more effective in classifying low-moderate resolution but may fail to accurately classify high spatial resolution remote sensed data (Chutia *et al*. 2016). Furthermore, traditional classifying algorithms assume independence among the classification outputs (Tuia *et al*, 2011) because traditional classifiers only take into consideration similarity between input values and ignore the potential relationships between the different classes that would have been predicted (Tuia *et al*. 2011).

MLC also fails to predict when the features are intensively redundant or superfluous (Chutia *et al*. 2015). Thus, this traditional classifier failed to perform when the remote sensed data source had high resolutions. This was witnessed by Chen *et al*. (2006), when a comparison between pixel-based and object-based oriented classification was done on multispectral SPOT 5 Imagery with a resolution of 2.5m and found that there was an increase in spectral dimensionality which in turn produced inconsistent classifications as well as the 'salt and pepper effect' with an overall accuracy of 77.79% and kappa coefficient of 0.7010 (Chen *et al*. 2009).

Having highlighted the shortcomings of traditional classification there were also major successes in the use of these. This was evident in a research done by Klehm *et al*. (2019) where the traditional classifier ISODATA was used to classify high resolution multispectral imagery for building a predictive model of the settlement pattern of the Iron Age community in a site in Botswana. The research demonstrated the potential of image classification to identify sites difficult to locate in a traditional pedestrian survey (Klehm *et al*. 2019). In this study traditional classification algorithm yielded positive results. According to Klehm *et al*. (2019) breaking up the image with a cluster-based, unsupervised classification worked well as an objective guide when trying to select features (groups of spectrally similar pixels) over an unfamiliar site ad this could be the major reason why the Traditional classifier yielded positive results.

2.4.2 Advanced Modern Classification Algorithms

Although advanced modern machine learning algorithms are more complex to work with (Tuia *et al*. 2011), they do not make any assumptions on the distribution of data. They are better known as non-parametric classifiers while most traditional classifier are parametric (Tuia *et al*. 2011). Use of Advanced classification algorithms for prospection has just begun in African archaeology (Davis & Douglas 2020).

2.5 Random Forest and Support Vector Machines

In a study to review the utilization of the Random Forest (RF) classifier in remote sensing, Belgiu & Dragut (2016) found that RF could successfully handle high data dimensionality, and multi collinearity and was insensitive to over fitting but sensitive to sampling design. A detailed description of the RF algorithm and its workings is outlined by Breiman (2001) and Biagetti *et al*. (2017). RF classification has been applied to both multispectral, hyperspectral, and has been proven to be the most successful in supervised classification (Biagetti *et al*. 2017). Advanced modern classifiers can be used when much is already known about the data before any modern classification may be applied. The data must not necessarily assume a certain distribution pattern

(Biagetti *et al*. 2017). Furthermore, the data gathered in the field or remotely must be a representative sample making up 70% training and 30% testing samples (Breiman 2001; Belgiu & Dragut 2016).

To optimise the Random Forest algorithm, a few parameters need to be set up first (Chen *et al*. 2012). These variable parameters are named *mtry* and *ntree* in the R implementation of Breiman's random forest algorithm, and their optimum values are obtained by doing a grid search approach based on the OOB estimate of error as explained by Biagetti *et al*.2017).

The RF classifier algorithm is sensitive to training samples. Training samples must be statistically independent, class balanced, representative of the population and large enough to bear data dimensions (Belgiu & Dragut 2016). A bad sample would entail collecting too few samples or samples which are not representative; a good sample is widespread over a given area (Joseph & Jaganathan 2018). Sampling must not lead to the destruction/disturbance of fragile archaeological remains thus one was careful in the site. Bad sampling practices lead to the destruction or moving of archaeological features *in situ* (Tartaron 2003). Good sampling techniques also involves the capturing correct coordinates of features and choosing an ideal period to do the sampling (Tartaron 2003). In the case of this research, the sampling was done in summer where there was less vegetation cover and rains therefore the features were not obstructed.
Further investigations have focused on how the training sample affects the performance of the RF classifier and how data dimensionality affects the classifier (Dalponte *et al*. 2013; Belgiu & Dragut 2016).

RF has outperformed many advanced machine learning algorithms through speed, stability, (Belgiu & Dragut 2016), but requires some parameters to be configured. Another important aspect of the RF algorithm is the ease with which the parameters such as the number of subset trees and the input features may be adjusted (Chatiu *et al*. 2015).

RF is a desirable classifier as it does not require cross validation so as to obtain an unbiased estimate of the test error and does not over fit (Chutia *et al*. 2015). RF is a preferred classifier because it is sensitive to the contribution of each band in LULC classification using the Mean Decrease in Accuracy using through the OOB error estimate (Breiman 2001; Thabeng *et al*. 2020; Biagetti *et al*. 2017). However, because it is very sensitive to the sampling design and RF can lead to incorrect models being predicted (Belgiu & Dragut 2016). It is evident from the above case that, when using RF, archaeological researchers must ensure they use a correct sampling strategy and be consistent in the type of data they collect in order to avoid misleading models. In this case a stratified sampling technique was enacted in the collection of archaeological data.

SVM generally outperforms traditional classification algorithms such as MLC (Biagetti *et al*. 2017). As is the case with RF classifies, SVM is effective as it is not affected by data dimensionality and the amount of training data and performs swiftly even with limited training data (Chutia *et al*. 2015; Mountrakie *et al*. 2015). It has been proven to perform optimally even in cases where the training data sets are minimal. Performance becomes compromised if the number features exceeds the training samples. Moreover, selecting the optimal kernel parameters for this algorithm was found to be challenging (Chutia *et al*. 2015). To get the best model when using SVM, certain conditions and protocol need to be met as highlighted above. Sigma and gamma parameters need to be tuned into the most optimal conditions as presented by grid search so as to train the SVM model. This in turn enables the archaeological training data to be optimally modeled in order to increase accuracy.

Both SVM and RF were chosen in this study due to the above-mentioned merits and because they have proven to be equally reliable (Ghosh *et al*. 2004; Dalponte et *al*. 2013; Thabeng *et al*. 2020; Thabeng *et al*. 2019).

3 Materials and methods

3.1 Study site

Ziwa is located in the north-eastern part of Zimbabwe (Figure 1). The cultural landscape was named after a granite mountain named Ziwa meaning, 'know' in the *Manyika* language. Key archaeological elements within site include stone terraces, stone enclosures, pit enclosures, hill forts and passages, smelting furnaces, grinding places, clearance cairns and other important remains built during the 16th and 17th centuries (Chipangura *et al*. 2017).

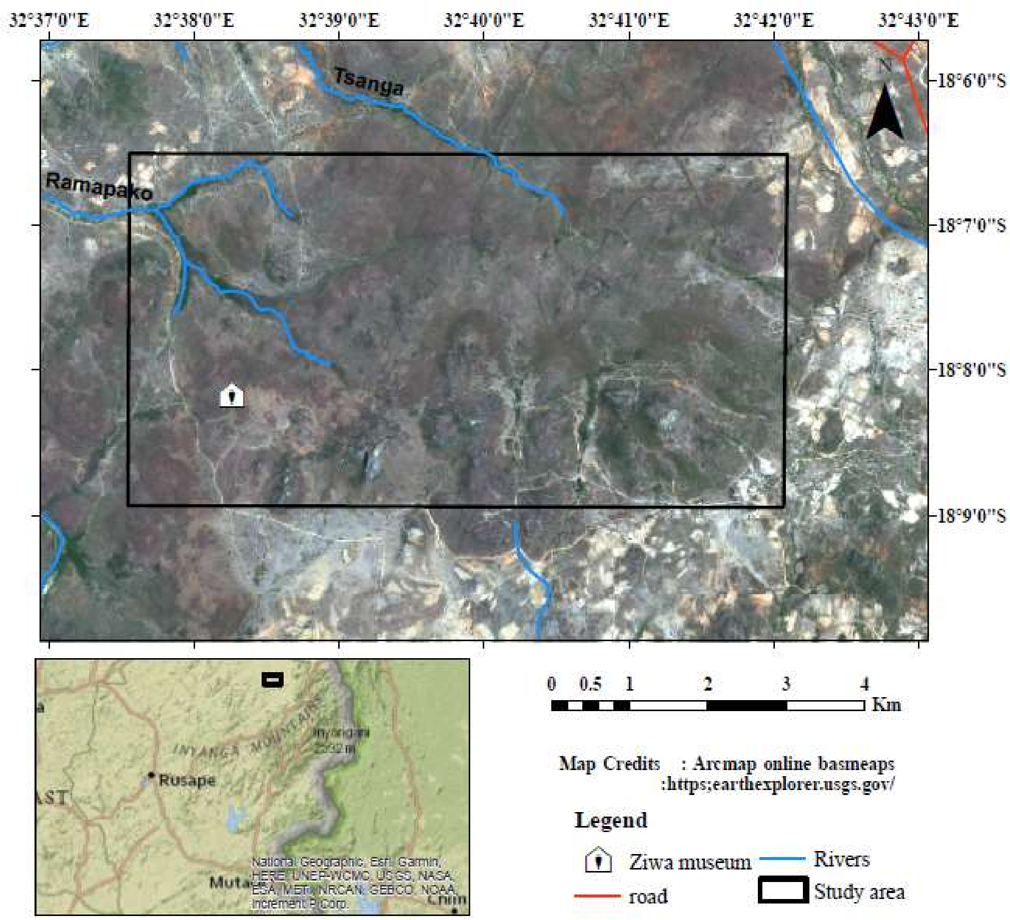


Figure 1. Map of the study area approximately 36 km^2 (the polygon in the lower figure is the footprint of the Sentinel image). This area was chosen due to prior knowledge of area acquired through existing foot surveys indicating potential sites within that boundary.

Ziwa was declared a national monument in 1946 and is being considered for world heritage listing (Shumba 2003). The site is a Late Iron Age site dated 1500-1700 A.D. and the area is known as a sacred shrine area (Shumba 2003)

Ziwa is predominantly made up of granite intrusions, visible as inselbergs, and sedimentary rock (Soper 2006). The granitic parent material gives rise to sandy soils which are generally highly permeable and lack the ability to store a lot of water (Soper 2006). The main geology of the area is the Precambrian Umkondo system composed of shales, quartzites and dolerites with highly leached soils (Jimu & Ngoroyemoto 2011).

The area receives an annual rainfall ranging from 741-2997mm with heavy dew and an average precipitation of 1200mm during the wet season between November and March (Jimu & Ngoroyemoto 2011). According to Jimu and Ngoroyemoto (2011), the mean annual temperature is at a minimum of 9-12^{0}C and a maximum of 25-28^{0}C. From August to November, the area is hot and susceptible to severe wildfires. The vegetation is mainly comprised of sub-montane and montane grassland (Jimu & Ngoroyemoto, 2011; Clark *et al.* 2017).

3.2. Materials

The research involved desktop work, accompanied by field work for the identification of training and testing data sets for classification purposes. The field work involved a foot survey of the site to assess any potential archaeological sites and a number of non-sites that would guide the classification within the boundary of the study site.

3.1.1 Introduction

Initially, the site and its geographic and cultural parameters were identified in collaboration with the primary ground-based archaeologists and anthropologists, Prof. Thornton and Dr Njabulo Chipangura. The key problem was defined, and the key outputs determined as follows: identification, classification and exploration of specific archaeological terrains in the selected area. The scope of the research was defined using GPS coordinates on GoogleEarth imagery in collaboration with the archaeologists. The field work involved a foot survey of the site to identify and assess potential archaeological features and a number of sites in order to guide the classification within the boundary of the study site. Soil samples were collected, brought back and analyzed at the University of Witwatersrand laboratory with the permission of the Zimbabwean Museums and Monuments Commission. Three Garmin eTrex 20 Global Positioning Systems (GPSs) were used to capture the relevant ground points. Desktop work involving digitising was done so as to supplement the validation and training data that was acquired during the field survey.

Following this, appropriate satellite imagery obtainable within the limits of the project budget was selected. Selection of appropriate software and machine-learning algorithms for the project followed subsequently. Google Earth imagery was used to construct sampling regions for training the algorithms. Google Earth (www.google.com/earth/index.html) is a free Google product and the specific version that was used in this research was Version 7.3. Google Earth is a popular geographical information application developed by Keyhole, Inc in the USA (Luo *et al*. 2018). This application was very useful in the digitization of more classes however due to the size and dense vegetation of some archaeological features (pit structures) they were not easily visible. Google Earth was used because imagery can be sourced at no cost and the data formats are easily accepted and manipulated by most geographic information software. Google Earth imagery also has some shortcomings. Google Earth has no spectral information and as it has only 3bands that is the Red, Green and Blue bands. Although its spatial resolution is high, Google Earth imagery has limited temporal resolution when compared to satellites such as the earlier Landsat missions. Furthermore, imagery for some parts of the world is restricted and its spectral and radiometric resolution allows only very limited RS analysis.

Digitizing on the Google Earth image was not an easy task, as some classes in question were covered by vegetation and others were overshadowed by shade caused by the mountains and dense canopy. To mitigate some of these challenges historic imagery was used and digitizing was employed to address issues of shade on areas with minimum or no shaded cover. Archaeological features under a shade or with dense vegetation were avoided.

Furthermore, an analytical software ENVI version 5.1 was used in this research to analyse the data. This tool is supported by an Integrated Data language (IDL), which is an array-based language (Chutia *et al*. 2015). This tool enabled the extraction of the different classes, spectral profiles from the multispectral image so that they could be compared.

3.1.2 Satellite imagery acquisition and pre-processing

Level 2A Sentinel-2 imagery data is available through USGS (United States Geological Survey Earth explorer) (https://earthexplorer.usgs.gov/). The satellite is owned by the European Space agency. This satellite was launched on the 23rd of June 2015 and hosts a Multispectral Instrument (MSI) with 13 Bands, a swath width 290Km, and spatial resolution of 10, 20 and 60 meters. It has a 5 day minimum global revisit time with twin satellites as well as a 10 m resolution (Immitzer *et al*. 2016). Sentinel satellites are part of the Corpenicus polar orbiting satellites which are sun-synchronous and phased at 180 degrees to each other (https://sentinel.esa.int/web/sentinel/missions/sentinel-2). The main objectives of these sensor is to monitor the changes on Earth's surface and atmospheric conditions and to aid this it has a wide swath width at 290 km. The Sentinel-2 characteristics are detailed in Table 1.

The chosen date for the acquired imagery from Sentinel-2 archives coincided with the time of the site visits (November 2018). This helped to support accurate classification and consistency.

The 13 spectral bands cover the Visible (VIS), Near Infrared (NIF) to the Short-wave Infrared (SWIR) at spatial resolutions ranging from 10 m, 20 m and 60 m. The Red, Green and Blue bands make up the natural color composite image (Louis *et al*. 2016). Bands at 60 m are Band 1 (443-

490 nm aerosol), Band 9 (945-1375 nm water vapor) and Band 10 (1375-1610 nm cirrus) and are ideally for atmospheric correction and cloud screening (Drusch *et al*. 2012, Louis *et al*. 2016). The majority of the bands with a 20 m spatial resolution are dedicated to vegetation classification (https://gdal.org/drivers/raster/sentinel2.html). Vegetation classification bands are within ranges 705-740 (band 5 Red Edge 1), 740-783 nm (band 6 Red Edge 2), 783-784 (band 7 Red Edge 3) and 865-945 (band 8A Red Edge 4 (Drusch *et al*. 2012). Bands 1610- 2190 and beyond make up the SWIR bands which are also at 20 m (Table 1). All these characteristics make Sentinel-2 imagery to be of sufficiently high spectral and spatial resolution which plays a big part in discriminating and prospecting the larger ancient archaeological mine features. Satellite imagery at ten or twenty meter resolution, is insufficient for identifying most archaeological sites and features.

Table 1. Characteristics of Sentinel-2 data. Source: https://scihub.copernicus.eu/.

Band Name	Band-Central wavelength	Band description	Band width (nm)	Spatial Resolution	Band range
Band	443nm	Aerosol	20	60	433-453nm
Band	490nm	Blue	65	10	458-523nm
Band	560nm	Green	35	10	543-578nm
Band	665nm	Red	30	10	650-680nm
Band 5	705nnm	RedEdge 1	15	20	698-713nm
Band 6	740nm	RedEdge 2	15	20	733-748nm
Band 7	783nm	RedEdge 3	20	20	773-793nm
Band 8	842nm	NIR	115	10	785-899nm
Band 8A	865nm	RedEdge 4	20	20	855-875nm
Band 9	945nm	Water Vapor	20	60	935-955nm
Band 10	1375nm	Cirrus	30	60	1360-1390nm
Band 11	1610nm	SWIR 1	90	20	1565-1655nm
Band 12	2190nm	SWIR 2	180	20	2100-2280nm

Cloud-free imagery dated December 7, 2018 covering over 10 000 km^2 inclusive of the study site as well as the greater part of Nyanga was acquired. This was sourced free of charge from the OpenScience hub (https://scihub.copernicus.eu/dhus/#/home). The date of acquisition correlated with the days the ground data was collected in Zimbabwe, furthermore this was an optimal season for the data collection since the rains had not started, and thus there was minimal vegetation cover. Minimal vegetation cover allowed a more thorough discrimination of archaeological features against non-features. Furthermore, minimal vegetation made it possible for proper determination of the spectral signatures of the features without contamination from vegetation cover. Obtaining an image during minimal vegetation cover allowed exposure of archaeological features and furthermore there was negligent cloud cover within this period and this is in line with the method that was undertaken by Thabeng *et al*. (2019).

Pre-processing of the image aimed to remove 'noise' and to correct of radiometric and spectral inconsistences which could compromise the analysis of the data for this specific project. Therefore, atmospheric correction of the image was sought to remove any disturbances that are caused by the interaction between sensor radiation and micro particles within the atmosphere. This was done using SNAP (Sentinel Application Platform) developed by ESA (European Space Agency). This software incorporates the plugin "Sen2Cor" that performs an atmospheric correction from Top of Atmosphere (TOA) to Bottom of Atmosphere (BOA) by adapting ATCO (Louis *et al*. 2016; Toming *et al*. 2016). The SNAP and sentinel tool boxes are available from http://step.esa.int/main/download whereas the Sen2Cor version 2.2.1 is available from: http://step.esa.int/main/third-party-plugins-2/sen2cor.

After pre-processing the Sentinel-2 image had 10 bands namely Bands, 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12 losing Band 1(coastal aerosol), Band 9 (water vapor) and Band 10 (SWIR Cirrus) which are used mostly for Hydrological research in Table 2. Thereafter resampling brought the image to the desired 10m resolution and the image was clipped to the area of interest.

In this research only Spectral bands were chosen for use in the prospection of archaeological sites This is consistent with the approaches employed by Sivitskis *et al.* 2019; Thabeng *et al*. 2019 and Abate *et al*. 2020. Agapiou *et al*. (2012) defines Vegetation Indices (VI) as indicators of vegetation stress or lack of which is associated with sub surface soil disturbance. Baghdad & Zribi (2016), highlight that a spectral band is a matrix of points defined by three dimensions, its coordinates and the intensity relating to the radiance. When archaeological features are refilled by soil, they usually retain water and nutrients thus plants grow more in these areas and stay green these may be easily identified using VI which enhance crop marks. (Lasaponara 2006; Cerra *et al*. 2018). VI enhance interpretation of multispectral satellite data as witnessed by Agapio *et al*. (2013) in a research to investigate the highlights and limitations of vegetation indices for the detection of a Neolithic tells in the Thessalian plain Greece where the majority performed dismally for archaeological prospection.

Although having showed positive results such as was in the case in Lasaponara & Masini 2006, Vegetation Indices do not come without short comings and thus they were not considered in this study. VI perform better in the presence of particular crops (barley, wheat, dense vegetation etc) and knowing the background of the soil is necessary as was the case in Agapiou *et al*. 2013. Ziwa the area is largely covered in sparse bush and shrubs as well as grass furthermore the background of the soil is still highly being contested between it being a mining or agricultural area, Thus Vegetation indices in this case could be misleading as the areas' history is still being highly contested. Crops such as barley and wheat cultivated over known archaeological sites as well as dense vegetation have been described as eliminating the error of the background due to their properties as they have a leaf area index over 3 (Agapiou *et al*. 2013). VI may be affected by leaf orientation, canopy structure thus may be very complicated and misleading (Agapiou *et al*. 2012. Unique VIs need to be developed for the identification of one plant variable thus for the case of Ziwa this is still yet to be determined due to the sparse distribution and diversity in the flora within the area (Agapiou *et al*. 2012). Unique VIs specifically for Ziwa are still yet to be developed bearing in mind its vegetation regime. VI give best results only when used in a particular phenophase of the crop or plant thus understanding of the phenophases of the plant/crop species is necessary (Agapiou *et al*. 2013). This entails that VIs are only effective only on certain periods of time of plant or crop growth thus it becomes a limitation. On other hand, Modern classification algorithms may be applied on any time period thus making spectral indices more convenient to apply on any season.

Depending on the season, VIs may give different results as was the case in a study done by Agapiou *et al*. 2013. They are developed to only enhance specific vegetation biophysical properties in a particular growth season of a particular plant species. VIs are affected by the spatial temporal variations of the atmos*p*here, soil brightness, environmental effects, shadows, soil colour, soil moisture making it easily. Thus, vegetation indices may be easily be misleading as their results oscillate depending on the season while spectral values in this research when applied on the various 'stone made' structures material give consistent results will give the same results in this case.

In the application of VIs better results are observed when high resolution hyperspectral data is used as was the case in Agapiou *et al*. 2012. Hyperspectral high-resolution data comes at a huge cost considering that the area of interest Ziwa study was over 36kms thus Free Sentinel imagery was used with spectral bands thus effectively cutting down costs.

Sahebjalal & Dashtekian (2013) highlight that use of spectral bands and classification of satellite imagery is a proper tool to derive Land cover use maps as well as important statistics Furthermore, they highlighted that using of spectral bands in classification provide satisfactory results in distinguishing different classes. One of the aims of this research was to statistically determine the performance of the modern advanced classification algorithms. Thus, VIs would not give the much-needed statistics such as the margins of error, Kappa, and accuracy assessment of the performance of the modern algorithms. On using VIs' it is crucial for one to determine the accuracy and error margin of their performance (Agapiou *et al*. 2013). For understanding the accuracy of VI they are usually run in groups so as to check how different results are. If there is very low relative differences between the indices results such as was the case inn Agapiou *et al*. 2012 several VIs were done and they showed low difference.

According to Agapiou *et al.* 2012, use of VI for the detection of crop marks lacks in accuracy and critical evaluation. For effective VI in satellite imagery, different indices are run thus a couple of images from different satellites with those specific bands is necessary thus more than one vegetation index is suggested to be used to enhance final results (Agapiou *et al*. 2012; Agapiou *et al*. 2013. Running of all these individual indices may be taxing and time consuming. On the other hand, using spectral bands and Artificial algorithms increases performance and time to do this is relatively less.

Furthermore, VI sometimes need to be supplemented by use of spectral information as was the case in Abate *et al*. 2020 where tasselled cap transformation and Principal component Analysis (PCA) was used to enhance the spectral information. Spectral bands need minimal supplement work for them to provide quick spectral information and the discrimination of materials.

Another disadvantage of VIs is that the values differ from one phenological stage to another ad 0thus cannot be used as a basis (Agapiou *et al*. 2012). Agapiou *et al*. (2012), in his research highlighted that VI are not normalised and may bring different results on the same phenological phase of plants as observed by Agapiou *et al*. 2013. Furthermore, the area of interest Ziwa had irregular vegetation patterns, that is, it had a host of different plant species which were in different phenophases. Use of VI was not going to be normalised as this changes over time due to environmental factors as explained by Agapiou *et al*. 2013. Spectral information is more normalised thus say for a specific material, the spectral signature is specific to that material. Furthermore, the type of material that was under investigation in the AOI determined led to the use of spectral information over the VI. The archaeological features within Ziwa are predominantly made of rock and the better way of discriminating them was with the use of spectral information which is more sensitive to the unique mineral components in each rock material. Unlike the VIs, spectral bands are more sensitive to the different archaeological features which are particularly sensitive to a certain portion of band range as demonstrated by Bassani *et al*. (2009).

3.1.3 Land cover categories and reference data collection

The site visit and surveying was done in Zimbabwe from the 5th to the 8th of December 2018 with the assistance of Dr. Njabulo Chipangura, Chief Curator of the Mutare Museum. Dr Njabulo is the manager in charge of oversight and management of the Ziwa Heritage site. He provided critical information on the location of important site features. Ground reference points of different categories of the features within the site were recorded. Pictures of the archaeological features were also taken (Table 3). The period chosen to conduct the field work had less rain and there was also less vegetation cover. This made it easy to identify some important classes under investigation such as pit structures, terraces and metalliferous surfaces.

Data were collected at both Nyahokwe and Ziwa within the large extent of the declared heritage area. The area is protected by the National Museums and Monuments of Zimbabwe under relevant legislation. Location of feature-rich sites, non-feature-rich sites and other important Land Use and Land Cover (LULC) classes used the stratified random sampling method that ensured the inclusion of all the key groups within a sample thus ensuring the acquisition of a representative sample. Thirteen classes were identified and are listed in table 2 below. Over 1200 samples were taken in the field. The samples were supplemented by digitization in Google Earth and a total of 5 362 were digitized. Furnaces, ceramics, grinding sites and grinding stones were also part of the archaeology in the area but were excluded from the classification due to their size which is smaller than the 10m resolution provided by the sensor.

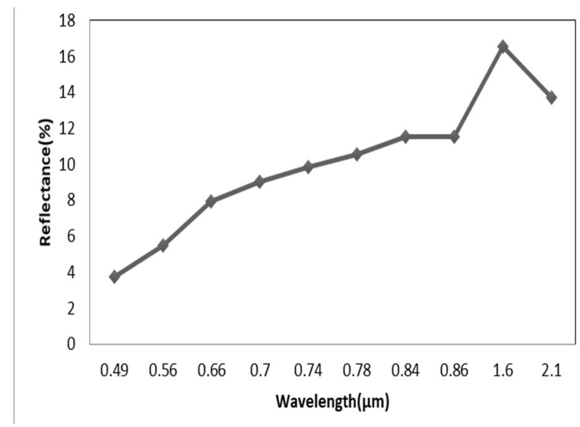
Two runs of classifications using Random Forest (RF) and Support Vector Machine (SVM) algorithms were done using the 5 361 points that were collected, the first using the initial set of categories, and the second using a modified set of categories. The first set of classifications included the initial thirteen classes, while the second round of classifications involved 11 classes to test for the impact of aggregated features on the overall classification as shown in Table 2 below.

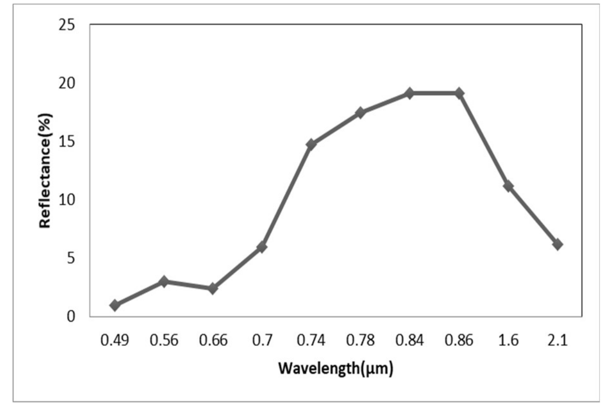
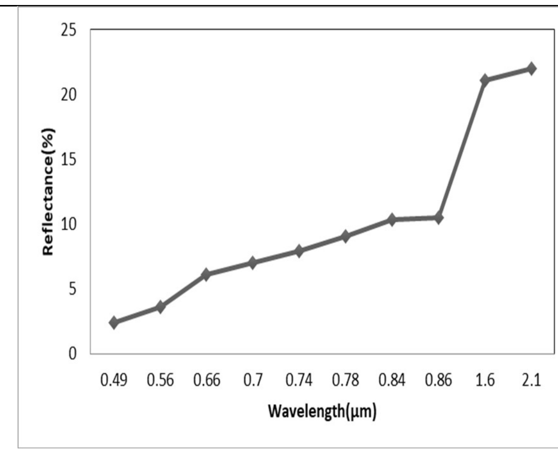
Table 2. List of classes that will be used for the supervised classification of the study area

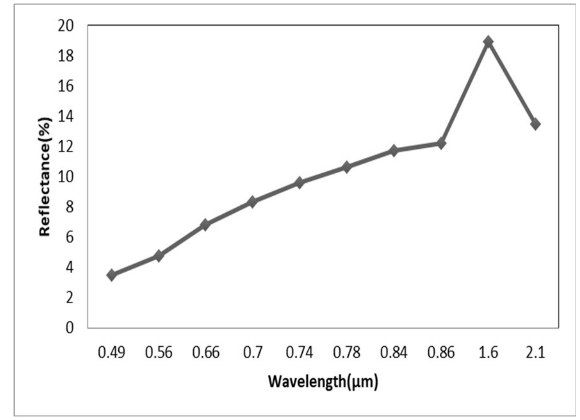
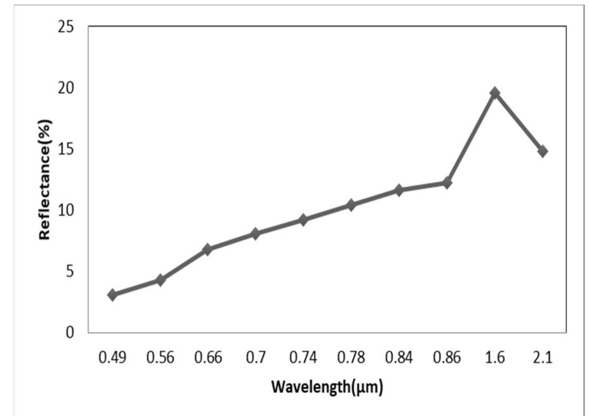
Class	Symbol	Feature	Number of points
Class 1	B	builtup_area	644
Class 2	Bare	bare_land	662
Class 3	C	cultivated fields	1266
Class 4	Enc	enclosures	189
Class 5	G	grassland	189
Class 6	M	metalliferous_surface	164
Class 7	Nm	non_metalliferous_surface	131
Class 8	Pit	pit_structure	42
Class 9	Ro	rock_outcrop	438
Class 10	T	terraces	262
Class 11	Tr	tarred road	90
Class 12	W	water body	908
Class 13	Wv	wooded vegetation	376
Class 14*	Sw	stone walls (terraces, pit structures and enclosures)	755

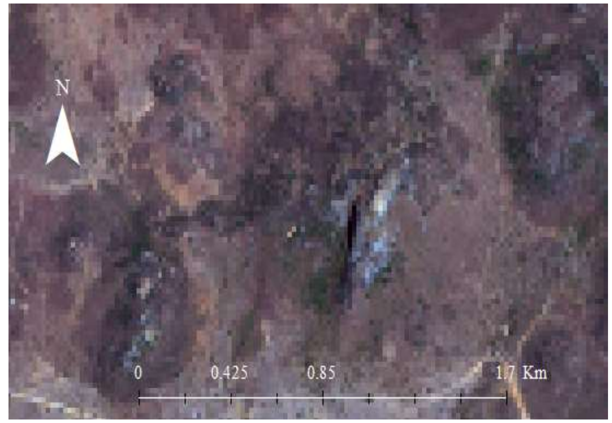
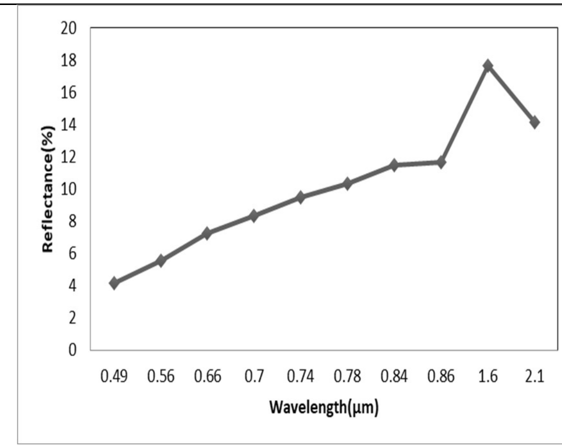
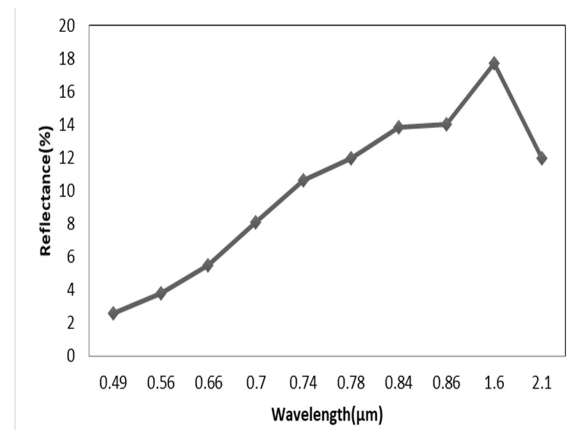
For the second set of classifications, classes named "pit-structures", "terraces" and "enclosures" were aggregated as one class: stonewalls (SW). This was done to improve the signal for these archaeological structures and terrains as they are all closely related on the ground and all made of the same rock materials. Subsequent runs proved that they could generate a better signal and this was implemented.

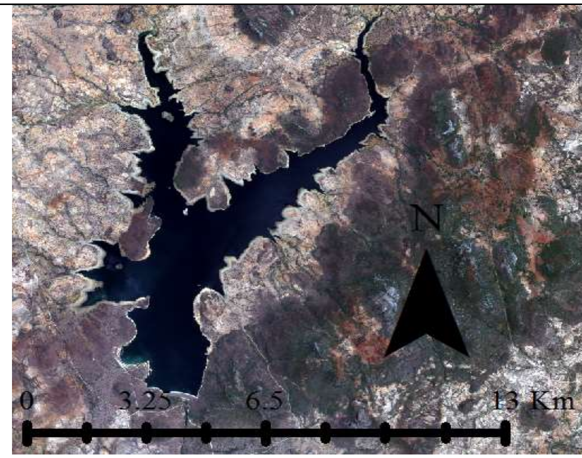
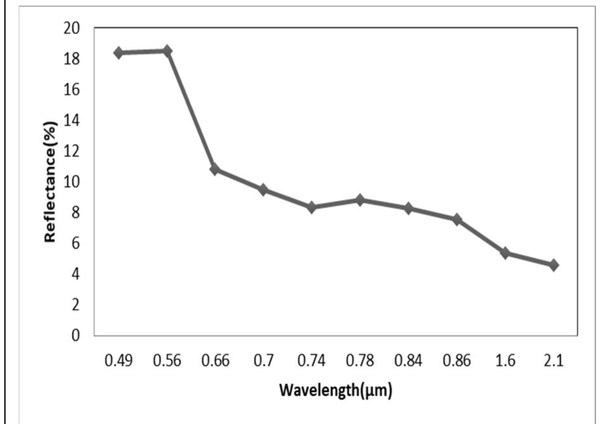
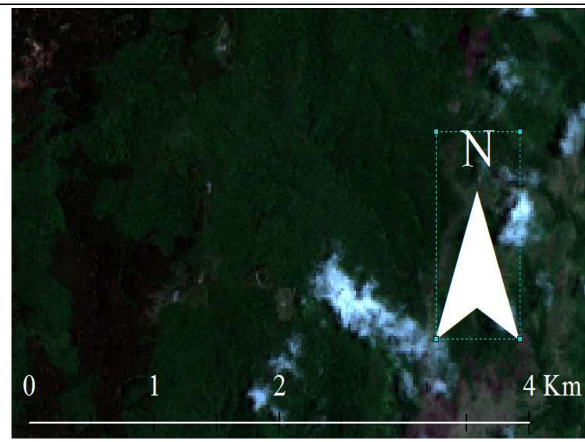
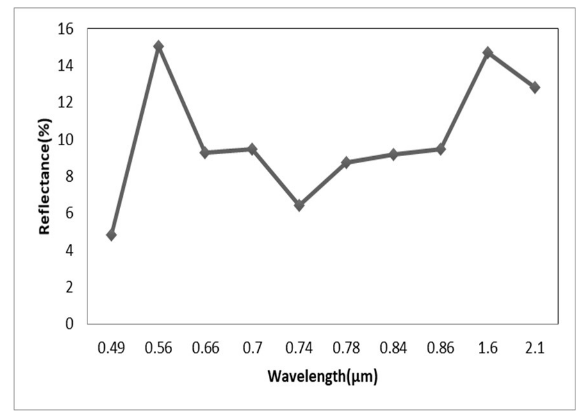
Table 3. Reference data for land cover land use classes. The mean reflectance curves were constructed using spectral data extracted from Sentinel-2 image using reference points

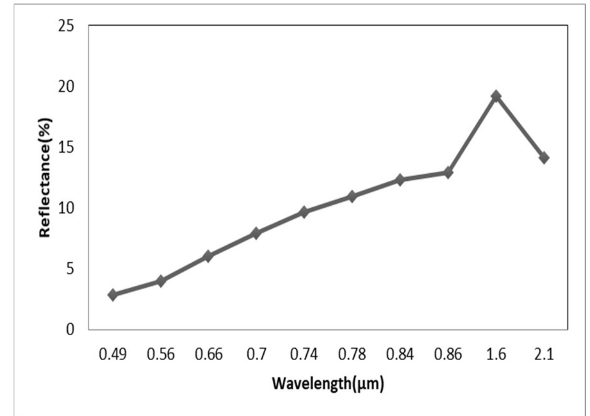
LULC class	Field observation photograph	Earth observation (as seen in Sentinel-2 imagery – 10m resolution)	Spectral reflectance curves
Built-up area	- 		

<table>
<tr>
<td>Bare Cultivated fields</td>
<td>Cultivated fields
</td>
<td></td>
<td>
</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td>
</td>
<td>
</td>
</tr>
</table>

Metalliferous surface Metalliferous surface			
Non metalliferous surface			

<table>
<tr>
<td>Rock</td>
<td></td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Pit structures</td>
<td></td>
<td>Not visible</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>Waterbody</td>
<td></td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Wooded vegetation</td>
<td></td>
<td>
</td>
<td>
</td>
</tr>
</table>

Terraces			Reflectance(%) Wavelength(μm) 0 5 10 15 20 25 0.49 0.56 0.66 0.7 0.74 0.78 0.84 0.86 1.6 2.1

3

3.1.4 Image Classification

After image correction and registration, the images were classified. 70% of the data was used for training while 30% was reserved for testing the model. The data was run using the first set of training points initially. This showed that several categories could be combined. The second run used the modified set with aggregated classes and the supervised classifications were done using two advanced, industry standard algorithms: Random Forest (RF) and Support Vector Machine (SVM). The first set of classes was classified using RF and SVM thereafter the second set of classes were classified using both RF and SVM.

These algorithms were implemented in R 3.4.1, using the "RSstudio" interface and programming environment and ENVI image analysis software. ArcGIS 10.0 was used for GIS tasks and mapping. RF and SVM algorithms have yielded satisfactory results in the past (Archer & Kimes 2008). Major merits of utilising these machine learning classifiers are that they are less sensitive to limited training points (Rodrieguez-Galiano *et al*. 2012b). This attribute was necessary in this research since there were limited ground data points collected due to the massive expanse of the archaeological landscape (Thabeng *et al*. 2019).

3.1.5 Random forest (RF) algorithm

Breiman & Cuttler (2001), developed the RF algorithm and described it an ensemble learning technique to improve the classification and regression trees (CART) by combining a large set of decision trees. Each tree contributes a single vote for the assignment of the most frequent class to the input data (Adam *et a*l. 2014). The implementation in the R library *ntree (number of trees)* (Breiman 2001) was used.

The bootstrap aggregating sampling method randomly selects two thirds of the original data for classification and the remainder (one third) for growing a tree (Archer & Kimes 2008; Thabeng *et al*. 2019). For optimal outputs, optimisation of conditions is possible through the systems' *mtry* function*,* OOBs and *ntrees* (Kulkarni & Sinha, 2013: Adam *et al*. 2014). During the process of growing trees at each node, the number of predictive variables considered at a node *mtry* is controlled by the user. This can be tuned up by the user to optimise performance of the model (Breiman 2001).

Furthermore, RF is a non-parametric classifier, having high accuracy and easily captures variation (Rodriguez-Galiano *et al*. 2012; Belgiu & Dragut 2016). Besides its ability to classify multi-spectral data, the RF algorithm measures importance of each variable (band input) in determining the classification outcome (Strobl *et al*. 2007). These variables include the mean decrease in accuracy, the Gini importance and the number of times each variable is selected by each tree ensemble (Strobl *et al*. 2007). The mean decrease in accuracy may be used to estimate which band contributes most to the distinguishing of different LULC classes (Thabeng *et al.* 2019). Higher numbers indicate the importance of the variable while lower values indicate the least important band contribution.

This technique is further explained in detail by Mutanga (Mutanga *et al*. 2012 and Kulkarni & Sinha 2013). Having said this, default optimising *ntree* and *mtry* values were used to optimise conditions within the R studio software suite package 3.4.1.

3.1.6 Support vector machines (SVM)

Boser, Guyon and Vapnik (1992), introduced machine learning classification algorithm to handle high dimensional data in remote sensing (Boser *et al*. 1992; Gualtieri *et al*. 1999). Adam *et al.* (2014) note that SVM is an advanced machine learning algorithm which forms part of the supervised classification technique where hyperplanes are drawn to divide each class. SVM uses a hyperplane to separate unlike classes using the maximum distance that will optimally differentiate the classes (Pal & Manther 2005). Support vectors are deduced through detecting the points closer to the hyperplane and thus are grouped as different classes avoiding errors (Pal & Manther 2005). The hyperplane is then used to make decisions on class separation (Petropoulos *et al*. 2012).

Furthermore, the algorithm uses the kernel method to perform classifications which is outlined by Gualtieri *et al*. (1999). These kernel functions include polynomial, sigmoid, radial basis function, and linear and are further explained by Ben-Hur & Weston (2010).

3.3 Accuracy assessment

Both sets of data for the two runs were split into training and testing data. From the whole sample the data was split on a proportion of 70% training and 30% testing data. The 70% data was used for doing the classification while the 30% was used in validating the results. This is a standard split used in RS accuracy assessment across most studies (Joseph & Jaganathan 2005)

To check the producer and user accuracy as well as any possible misclassification of data, and ascertain overall classification, an accuracy assessment was done where a confusion matrix was used. Accuracy assessment is based on the computation of the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and the Kappa (Kc) statistic through a confusion matrix (Congalton & Green 1999). The OA is the ratio of the number of validation pixels that have been correctly classified to the total number of validation pixels used for all classes and is expressed as a percentage (Adam *et al*. 2014). Kc is the proportion of correctly classified validation points after random agreements are removed and it expresses the extent to which the confusion matrix results are not obtained by chance or randomly (Petropoulos *et al*. 2012; Adam *et al*. 2014).

The results of the two different classifiers were compared for overall accuracy using the McNemar's test. This was used to test whether there was any significant difference in the way the two models performed. It is a non-parametric test built on a two-by-two confusion matrix (de Leeuw *et al*. 2006). It is computed through the formulae:

$$Z = \frac{f12 - f21}{\sqrt{f12 + f21}},$$

where $f\,12$ denotes the number of samples misclassified by the first classification algorithm but correctly classified by the second classification algorithm, $f\,21$ denotes the number of samples that are correctly classified by the first classification algorithm but misclassified by the second classification algorithm (Adam *et al*. 2014). A difference in the accuracy of the 2 individual classification algorithms is determined to be statistically significant ($p \leq 0.05$) if the Z value is greater than 1.96.

4 Results

4.1 Tuning of RF and SVM parameters

For the first RF trial, *mtry* values were set at 7 whilst the *ntree* was set at 500 while for the second trial, *mtry* values were set at 5 whilst the *ntree* was set at 500 (Figure 2). For the two RF trials, the OOB error rates were 19,57% and 18,86% for the first trial (without aggregating classes) and second trials (with aggregating classes) respectively. Two sets of RF classifications were conducted, thus both indepedant data sets had to be tuned independently. Grid search then gave 7/500 and 5/500 as optimal parameters for classifying the 2 sets of data. Grid search gave 0.01 and 10 as a combinations of both gamma and cost values bringing the most optimal prediction accuracy for SVM, these values are similar to the findings Thabeng *et al*. (2019) used.

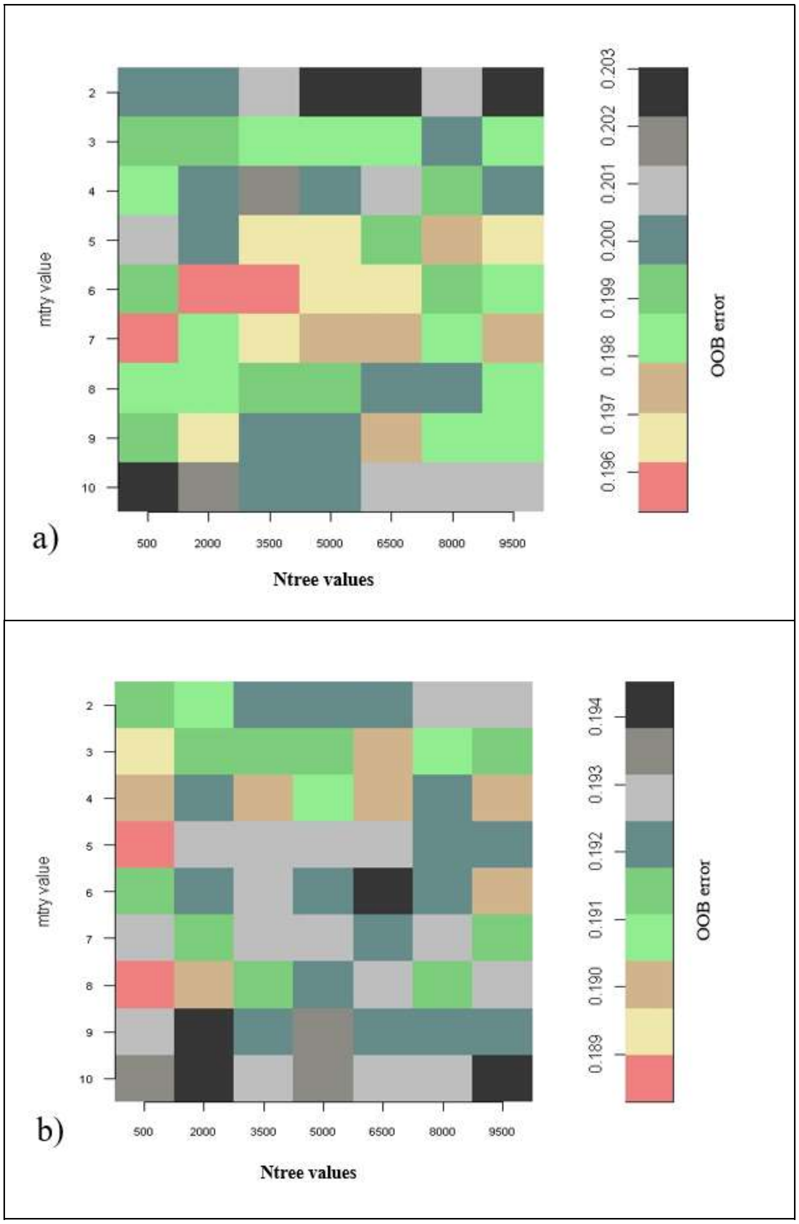


Figure 2. OOB errors of optimised *mtry* and *ntree* values using RF settings where: a) denotes the unmixed classification and b), the mixed/aggregated classes.

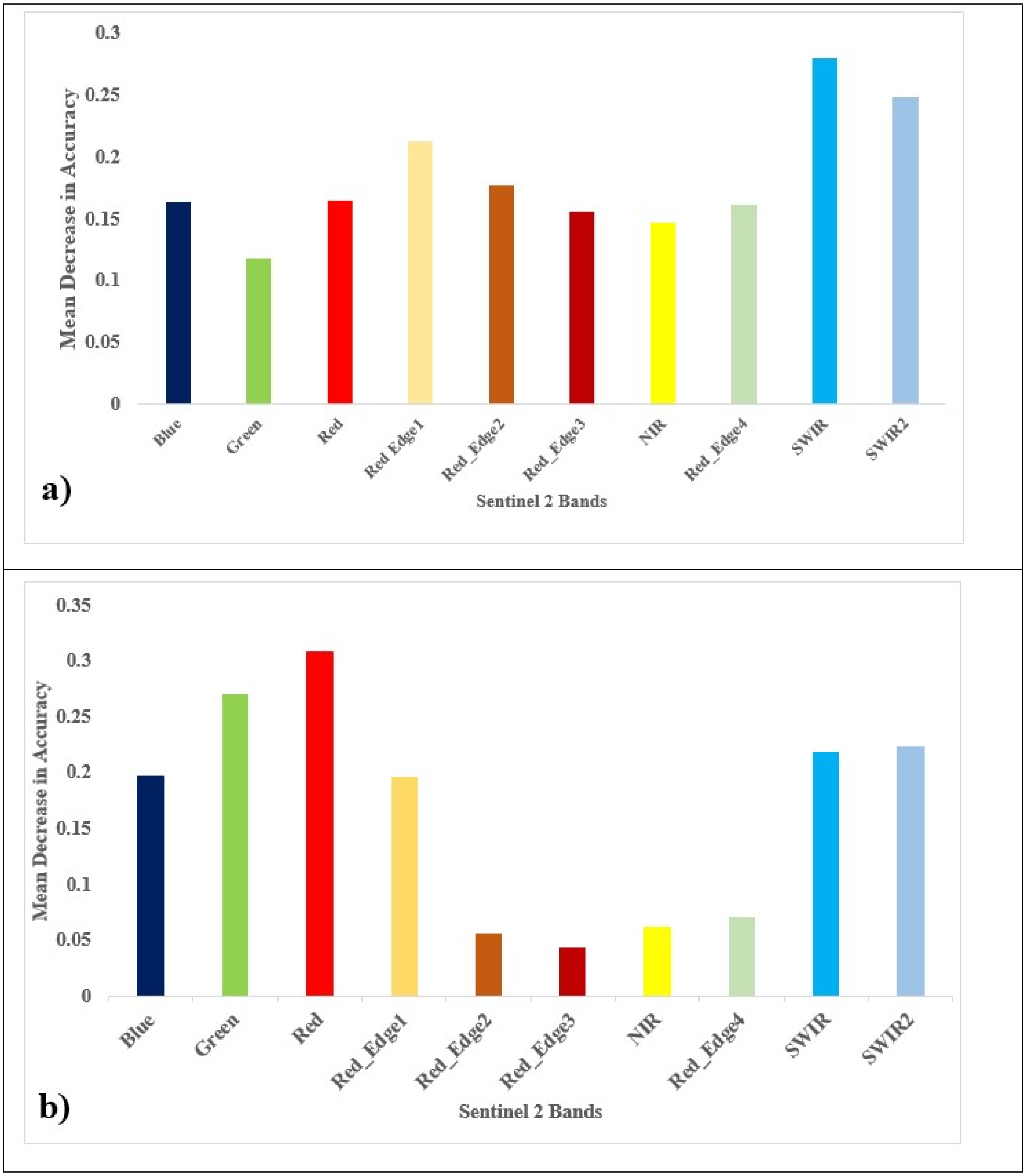


Figure 3. Histogram showing the overall importance of different Sentinel-2 bands in detecting the different LULC classes reported where a) is Trial 1 RF trial using individual classes and b) is Trial 2 using aggregated classes as indicated in Table 1.

Random forest was able to ascertain the importance of each band in discriminating the unique classes within the site (Figure 3). This was done using the RF model with an attempt to measure the importance of each variable through the use of mean decrease in accuracy as a measure of average decrease in prediction accuracy before and after the prediction (Thabeng *et al*. 2019). Important bands are those that had large values in the difference in prediction trees before and after classification. The most important bands that contributed to the overall classification and accuracy of the Random Forest model were the SWIR band (1565-1655) and SWIR2 band (2100-2280nm) for the first set of classification and then the Red (650-680nm), Green (543-578nm) and SWIR (1565-1655nm) and SWIR 2 band (2100-2280nm) for the second set of RF classification (Figure 3).

Having highlighted this, the RF model was also instrumental in discriminating the archaeological features (pit-structure, enclosures, terraces and metalliferous surfaces) from other LULC classes. The mean decreases in accuracy and RF model highlighted that for Trial 1 the SWIR, Red Edge 1 and SWIR 2 bands were most influential in the detection of enclosures. Among the Sentinel-2 bands, the SWIR, Blue and Red Edge 2 bands were crucial in detecting the metalliferous surface (Figure 4). Furthermore, the mean decreases in accuracy and RF revealed that Red Edge 2, Red Edge 1 and Red bands contributed the most in detecting pit-structures. Terraces were best discriminated from the rest of the other LULC through the Red Edge 2, SWIR and Red bands (Figure 4).

On aggregating the classes, enclosures pit structures and terraces, the Red Edge1, SWIR 2 and SWIR 1 bands were crucial in discriminating archaeological features from the rest of the other LULC (Figure 4.). In discriminating metalliferous surfaces from the rest of the classes, there was consistency with both trials as the same bands (SWIR, Blue, and Red Edge 1) in first Trial as well as in Trial 2 (Figure 4.).

Amongst the Sentinel-2 bands it is evident from this research that the Red Edge1, SWIR2 and SWIR bands were crucial in detecting stone structures that were related to the iron and smelting industry that existed in this site. Moreover, the metalliferous surfaces are in both trials easily discriminated through the combined contributions of the SWIR, Blue and Red Edge 1 bands and this indicates the important bands in understanding the spectral signatures of the archaeological features.

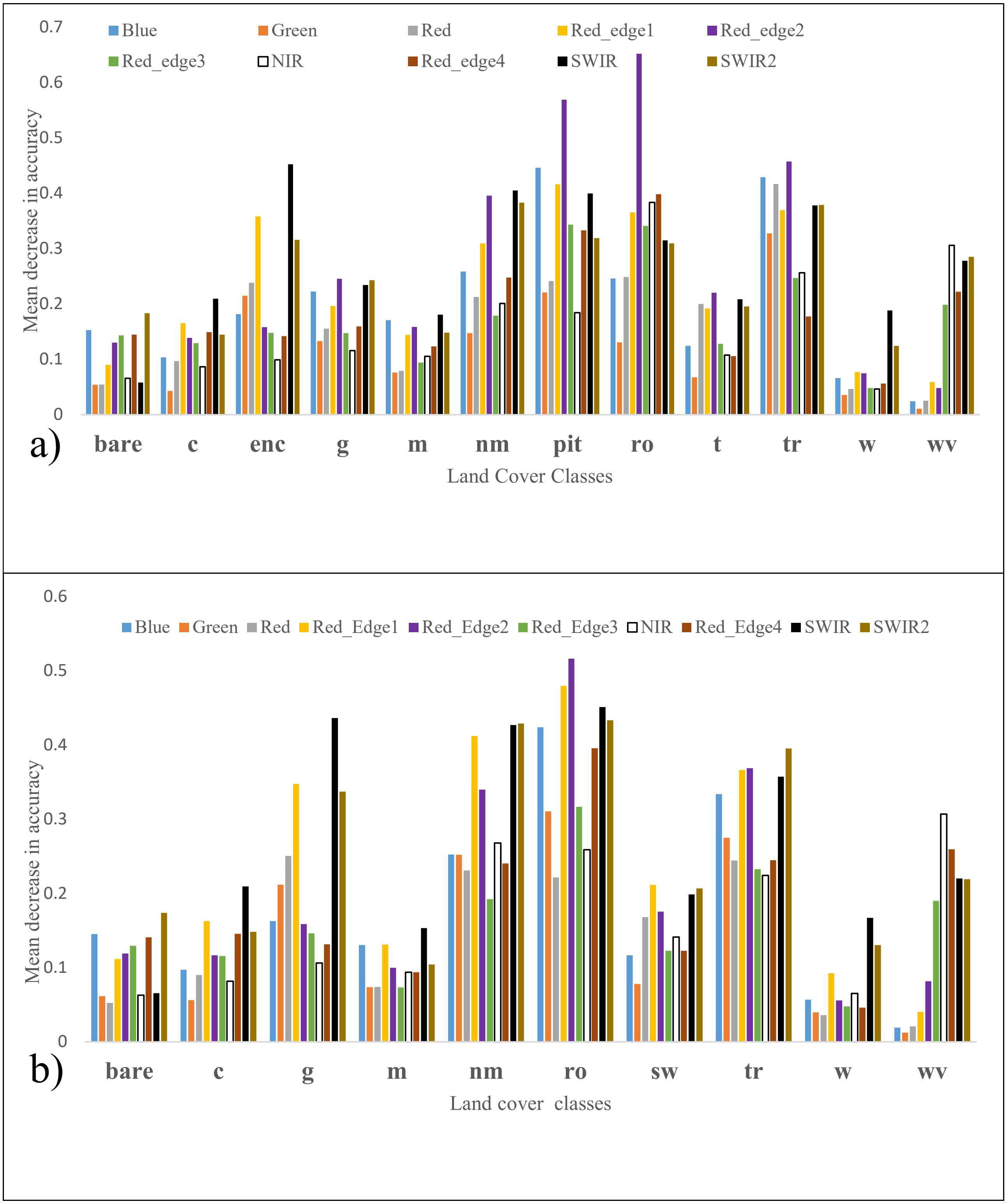


Figure 4. Histograms showing the importance of different Sentinel-2 bands in detecting each land cover classes : bare land (bare), cultivated land(c), grassland(g), (metalliferous_surface (m), non_metaliferous_surface (nm), rock_outcrop (ro), stone walls (sw), tarred road (tr), water_body (w) and wooded vegetation (wv), where a) is the initial classification and b) is the second RF classification.

4.2 Image classification

The Sentinel-2 image was classified on two trials using both RF and SVM so as to discriminate between built-up area, cultivated area, grassland, metalliferous surface, non-metalliferous surface, rock outcrop, stone walls (pit structure, terraces, and enclosures), water body and woody vegetation. In 'trial 1,' standalone classes included built-up area, cultivated area, grassland, metalliferous surface, non-metalliferous surface, rock outcrop, pit structure, terraces, enclosures, water body and woody vegetation (Figure 5). In 'trial 2', pit structure, terraces and enclosures were aggregated into one class, 'stone walls' (Figure 6)

The most prominent features of both the classified images was the pronounced presence of archaeological sites and metalliferous surfaces in the north-eastern part of the map (Figures 5 and 6). The results highlight the possibility that mining activities were concentrated and quite dominant on the north eastern side as there is a myriad of granitic rocks and boulders. Further instigations in this area could yield positive in roads for interest groups if a few issues be explored further. The reasons for exploiting a mountainous highland area, in the northeastern side whose geology is predominantly granitic for mining purposes should be investigated. Another difference between the outputs of the thematic maps was that the RF model had larger footprint covered by the archaeological features than the SVM model (Figure 8).

Another interesting finding is that the different algorithms seem to possess different strengths in discriminating the different classes. RF seems to predict more areas within the study area as being previously covered with archaeological mining features while SVM has a lesser area in the first trial (Table 3.1). Statistically, RF assigned 5.72km^2, 1.54km^2, 1.11km^2 and 10.37km^2 of the study area to enclosures, metalliferous surface, pit structures and terraces respectively in the first set of classifications. On the other hand SVM assigned 8.3km^2, 0.54km^2, 0.28km^2 and 8.24km^2 of the study area to enclosures, metalliferous surface, pit structures and terraces respectively in the first set of classifications (Table 3.1) In total RF assigned 18.74 km^2 while SVM assigned a total of 17.36 km^2 as possible archaeological sites.The aggregation of the three classes (pit structures, enclosures and terraces) seems to have had a significant impact on the performance of the two models. There are more archaeological features in RF than in SVM in trial 2 (Table 3.11). This is evident on table 3.1 where RF assigned 1.01 km^2 and 21.78km^2 of the study area to metalliferous surfaces and stone walls respectively. SVM on the other hand, assigned 0.59km^2 and 16.21km^2 of the study area to metalliferous surface and stonewalls respectively. In the second set of classifications, RF assigned more of the area to archaeological sites than SVM. This was a total of 22.79km^2 and 16.8km^2 for RF and SVM of the study area respectively (Table. 3.1).

Table 3.1 Area (km2) covered by different LULC classes as predicted by RF and SVM Classifier for both trials

	Trial 1		***Trial 2***	
Class	RF unaggregated	SVM unaggregated	RF aggregated	SVM aggregated
built-up area	0.17	0.40	0.12	0.23
bare land	3.93	4.71	3.22	4.60
cultivated fields	0.59	1.31	0.61	0.95
enclosures	5.72	8.30		
grassland	1.39	1.49	0.78	1.67
metalliferous surface	1.54	0.54	1.01	0.59
non metalliferous surface	1.55	1.18	0.83	1.31
pit structures	1.11	0.28		
rock outcrop	6.40	3.81	5.23	3.06
terraces	10.37	8.24		
tarred road	0.08	0.11	0.08	0.12
wooded vegetation	3.54	6.05	2.72	7.69
stone walls			21.78	16.21
water body	0.11	0.09	0.11	0.07

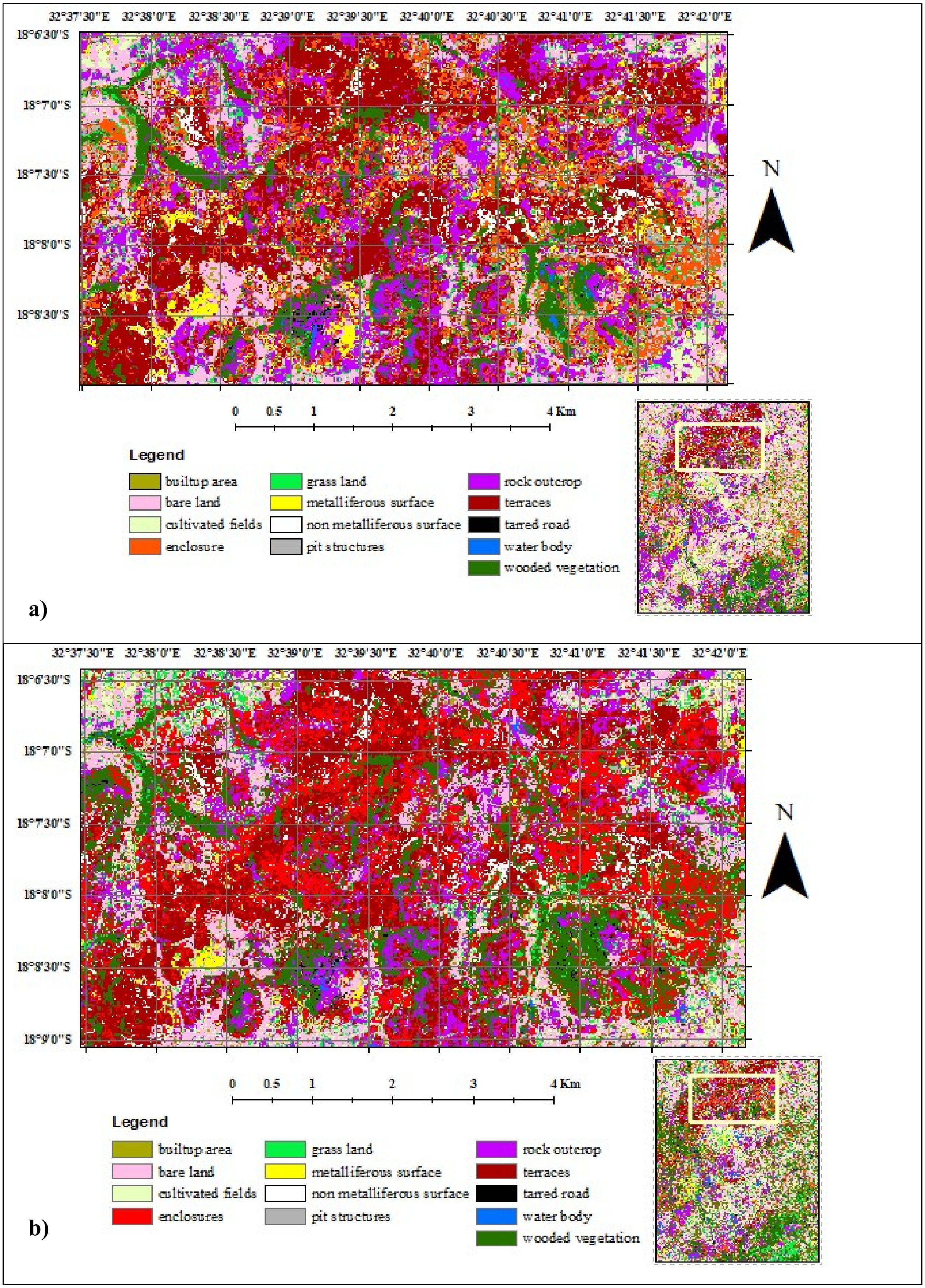


Figure 5. Classification maps obtained using Random Forest (a) and Support Vector Machine (b) using standalone classes in trial 1

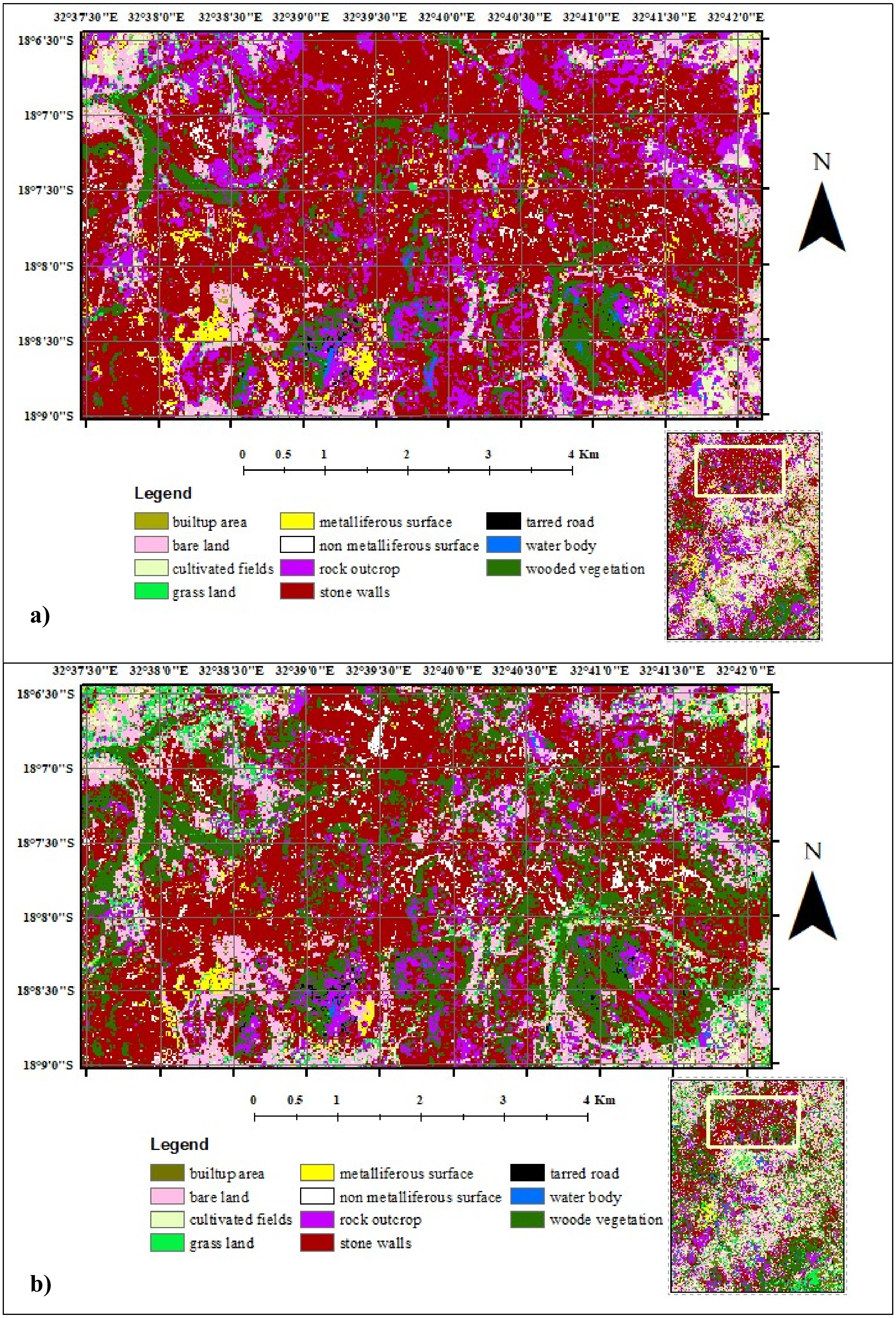


Figure 6. Classification maps obtained using Random Forest (a) and Support Vector Machines (b) using aggregated classes in trial

4.3. Accuracy assessment

The test data was used to assess the accuracy of both RF and SVM classifiers. From the two trials that were done, RF presented a higher overall accuracy than SVM. The table below summarises the results of the accuracy assessment of the 2 algorithms in the two trials respectively.

4.3.1 Trial 1

In the first trial 13 classes were taken into consideration using default RF and SVM parameters. All the classes were addressed as standalone classes without being aggregated. Overall accuracy (OA) achieved for RF was 80.6% and a Kappa coefficient of 0.78 (Table 4). Categories, enclosures, metalliferous-surfaces, non-metalliferous surface, pit structures, and terraces achieved 88.37%, 85.11%, 92.68%, 95.74% and 84.27% respectively for the UA while PA was pegged at 67.86%, 81.63%, 97.44%, 100% and 96.15% respectively (Table 4). On the other hand, the same 13 standalone classes SVM achieved an Overall Accuracy of 80.17% and a Kappa coefficient of 0.77 (Table 5). Categories, enclosures, metalliferous-surfaces, non-metalliferous surface, pit structures, and terraces achieved 82.08%, 90.48%, 88.10%, 97.83% and 88.757% respectively for the UA while PA was pegged at 69.64%, 77.55%, 94.87%, 100% and 91.03% respectively (Table 5). In discriminating thirteen LULC individual classes using the RF algorithm, the enclosure's class was confused with terraces, rock outcrop, metalliferous surfaces, bare land and cultivated areas classes (Table 4). On the other hand, on the same set of initial set of classifications, the SVM algorithm confused the enclosure's class with metalliferous surfaces, rock outcrop, and terraces (Table 5). The RF algorithm confused the metalliferous surface class with bare land, wooded vegetation and grassland (Table 4). SVM algorithm as well confused the metalliferous class with bareland and grassland (Table 5). RF confused the pit-structure class with the metalliferous surface and grassland classes while SVM confused the same class with only wooded vegetation (Tables 4 and 5). The class terraces was confused with bare land, enclosures, grassland, non-metalliferous surface and wooded vegetation classes by the RF algorithm (Table 4) while SVM confused the class terraces with bare land, enclosures, non -metalliferous surface, and wooded vegetation (Table 5). A great number of the above archaeological features were confused with other classes by both the classifies which then necessitated a second set of classifications which meant grouping the terraces, pit structures, enclosures as one class.

Table 4. Confusion matrix showing overall classification accuracy and kappa for discriminating 13 LULC individual classes using the RF algorithm

	built up area	bare land	cultivated fields	enclosures	grassland	metalliferous surface	non metalliferous surface	pit structures	rock outcrop	terraces	tarred road	water body	wooded vegetation	Total	UA
built up area	**123**	26	16	4	5	0	0	0	10	0	5	4	1	194	63.40
bare land	6	**133**	23	2	5	1	0	0	8	1	2	0	3	184	72.28
cultivated area	25	23	**327**	0	4	0	0	0	8	0	1	1	0	389	84.06
enclosure	0	1	1	**38**	0	1	0	0	1	1	0	0	0	43	88.37
grassland	1	3	1	0	**21**	3	0	0	0	0	1	1	1	32	65.63
metalliferous surface	0	4	0	0	2	**40**	0	0	0	0	0	0	1	47	85.11
non metalliferous surface	0	0	0	0	0	3	**38**	0	0	0	0	0	0	41	92.68
pit structure	0	0	0	0	1	1	0	**45**	0	0	0	0	0	47	95.74
rock outcrop	15	3	0	7	3	0	0	0	**99**	0	5	4	3	139	71.22
terraces	0	4	0	5	2	0	1	0	0	**75**	0	0	2	89	84.27
tarred road	4	0	1	0	1	0	0	0	2	0	**13**	5	1	27	48.15
water body	0	0	0	0	1	0	0	0	1	0	5	**253**	3	263	96.20
wooded vegetation	2	2	0	0	11	0	0	0	2	1	0	4	**97**	119	81.51
Total	176	199	369	56	56	49	39	45	131	78	32	272	112		
PA	69.89	66.83	88.62	67.86	37.50	81.63	97.44	100.00	75.57	96.15	40.63	93.01	86.61		
OA	80.67%														
Kappa	0.78														

Table 5. Confusion matrix showing overall classification accuracy and kappa for discriminating 12 LULC c lasses using the SVM algorithm

	built up area	bare land	cultivated fields	enclosure	grassland	metalliferous surface	non metalliferous surface	pit structure	rock outcrop	terraces	tarred road	water body	wooded vegetation	Total	UA
builtup area	**136**	16	27	4	10	0	0	0	12	0	4	8	1	218	62.39
bareland	7	**130**	23	3	4	1	0	0	9	1	2	1	0	181	71.82
cultivated fields	23	31	**316**	0	2	0	0	0	7	0	2	1	0	382	82.72
enclosure	0	0	0	**39**	0	1	0	0	2	5	0	0	0	47	82.98
grassland	1	7	1	0	**25**	6	0	0	5	0	0	2	5	52	48.08
metalliferous surface	0	2	0	0	2	**38**	0	0	0	0	0	0	0	42	90.48
non metalliferous surface	0	1	0	1	0	3	**37**	0	0	0	0	0	0	42	88.10
pit structure	0	0	0	0	0	0	0	**45**	0	0	0	0	1	46	97.83
rock outcrop	4	6	1	5	2	0	0	0	**89**	0	4	2	3	116	76.72
terraces	0	3	0	3	0	0	2	0	0	**71**	0	0	1	80	88.75
tarred road	2	1	0	0	1	0	0	0	1	0	**13**	2	0	20	65.00
water body	2	0	1	0	2	0	0	0	4	0	5	**255**	1	270	94.44
wooded vegetation	1	2	0	1	8	0	0	0	2	1	2	1	**100**	118	84.75
Total	176	199	369	56	56	49	39	45	131	78	32	272	112		
PA	77.27	65.33	85.64	69.64	44.64	77.55	0.95	100.00	67.94	91.03	40.63	93.75	89.29		
OA	80.17%														
Kappa	0.7724														

4.3.2 Trial 2

In the second trial the default RF and SVM parameters were still kept for consistency. Some classes were aggregated: ‘enclosures’, ‘pit strictures and ‘terraces’ were aggregated as class ‘stone-wall.’ The rest of the classes were kept as in ‘standalone’ and this yielded a total of 11 classes for the second trial and they were then classified using RF and SVM. The overall accuracy for RF dropped slightly to 80.25% from 80.67% in the first trial and a Kappa coefficient of 0.77 was obtained (Table 6). Categories ‘metalliferous-surfaces’, ‘non-metalliferous surface’, and ‘stone-walls’ achieved 78.43%, 90.00%, and 85.35% respectively for the UA, while PA was pegged at 81.63%, 92.31%, and 93.89% respectively (Table 6). On the other hand, the same 11 classes SVM achieved an Overall Accuracy of 80.62% and a Kappa coefficient of 0.78 (Table7). Categories, metalliferous-surfaces, non-metalliferous surface and stone-walls achieved UA of 75.47%, 78.57% and 91.57% respectively as well as a PA of 81.63%, 84.62% and 90.56% (Table 7).

In the second set of classifications, the RF classifier confused class, metalliferous surface with bare land, grassland non-metalliferous surfaces, and rock outcrop. (Table 6). The second set of classifications did not improve RFs’ ability to discriminate class ‘metalliferous surfaces’ from other classes. On the other hand, for the same class the SVM algorithm confused it with bare land, grassland, non-metalliferous surfaces, and stone walls (Table 7) which meant that reclassifying did not improve the results. RF confused stonewall with builtup areas, cultivated areas, bareland, grassland, metalliferous surfaces, rock outcrop water bodies and wooded vegetation (Table 6) while SVM confused the same class with the classes, bareland, grassland, metalliferous surface, non-metalliferous surface, rock outcrop, and wooded vegetation (Table 7). Having said the above, grouping the classes had a bearing in the overall performance of the two classifiers. RF’s performance overall accuracy dropped after grouping the classes while the overall accuracy for SVM improved significantly due after the aggregation of the classes.

Table 6. Confusion matrix showing overall classification accuracy and kappa for discriminating 11 LULC classes using the RF algorithm

	built up area	bare land	cultivated fields	grass land	metalliferous surface	non metalliferous surface	rock outcrop	stone walls	tarred road	water body	wooded vegetation	Total	UA
built up area	118	28	14	2	0	0	10	6	2	4	1	185	63.78
bare land	6	124	24	3	0	0	7	0	2	0	3	169	73.37
cultivated land	27	29	325	3	0	0	7	0	1	1	0	393	82.70
grassland	1	1	1	25	2	0	0	0	0	0	1	31	80.65
metalliferous surface	0	4	0	3	40	3	1	0	0	0	0	51	78.43
non metalliferous surface	0	0	0	0	4	36	0	0	0	0	0	40	90.00
rock outcrop	14	3	2	2	0	0	94	5	6	4	3	133	70.68
stone walls	5	6	2	6	3	0	1	169	0	1	5	198	85.35
tarred road	4	0	1	2	0	0	2	0	15	3	1	28	53.57
water body	0	0	0	2	0	0	7	0	5	254	2	270	94.07
wooded vegetation	1	4	0	8	0	0	2	0	1	5	96	117	82.05
Total	176	199	369	56	49	39	131	180	32	272	112		
PA	67.05	62.31	88.08	44.64	81.63	92.31	71.76	93.89	46.88	93.38	85.71		
OA	80.25												
Kappa	0.77												

Table 7. Confusion matrix showing overall classification accuracy and kappa for discriminating 11 LULC classes using the SVM algorithm.

	builtup area	bare land	cultivated fields	grass lands	metalliferous surface	non metalliferous surface	rock outcrop	stone walls	terraces	water body	wooded vegetation	Total	UA
builtup area	**124**	14	19	2	0	0	8	6	6	3	1	183	67.76
bare lands	8	**123**	34	5	0	0	4	3	1	1	0	179	68.72
cultivated fields	29	40	**310**	3	0	0	5	1	1	1	0	390	79.49
grasslands	6	11	3	**31**	1	0	1	1	0	1	1	56	55.36
metalliferous surface	0	2	0	6	**40**	3	0	2	0	0	0	53	75.47
non metalliferous surface	0	3	0	0	4	**33**	0	1	0	0	1	42	78.57
rock outcrop	5	2	3	2	0	0	**101**	3	1	2	2	121	83.47
stone walls	0	3	0	2	4	3	1	**163**	0	0	2	178	91.57
tarred road	2	1	0	0	0	0	2	0	**19**	5	0	29	65.52
water body	1	0	0	0	0	0	4	0	2	**254**	1	262	96.95
wooded vegetation	1	0	0	5	0	0	5	0	2	5	**104**	122	85.25
Total	176	199	369	56	49	39	131	180	32	272	112		
PA	70.45	61.81	84.01	55.36	81.63	84.62	77.10	90.56	59.38	93.38	92.86		
OA	80.62%												
Kappa	0.78												

4.2.3 MacNemar's test

Having highlighted the differences in classification accuracies, the MacNemar's test was done on the LULC classes to assess the performance of the prediction accuracies of the two individual classifiers. The confidence level (*p*-value) was set at 5% and on running the test, '*z*-values' of This 1.692, and 2.2 were computed for the two classification trials respectively (Tables 8.1and 8.2)

In the first trial (Table 8.1), the figure was less than 1.96 thus it could be concluded that there was no significant difference between the two advanced classifiers in discriminating the LULC classes under investigation using the individual classes. Moreover, the MacNemars' test result entail that RF and SVM performed similarly and thus could both be used for the same job as the results in this case were not significantly different. RF or SVM can be used and possible bring up similar results in this case. However, on aggregating (grouping the classes) that is, in the second trial (Table 8.2), the figure was greater than 1.96 thus it could be concluded that there was a statistically significant difference between the two advanced classifiers in discriminating the LULC classes under investigation after the grouping classes. This entails that grouping of archaeological classes has a bearing in the performance of the two algorithms in discriminating LULC.

Table 8.1 MacNemar's test results for trial 1

Classifier		RF		
SVM		Correctly classified	Misclassified	**Total**
	Correctly classified	1474	23	**1497**
	Misclassified	36	87	**123**
	Total	**1510**	**110**	
	Z value	**1.692**		

Table 8.2. MacNemar's test results for trial 2

Classifier		RF		
SVM		Correctly classified	Misclassified	**Total**
	Correctly classified	1441	39	**1480**
	Misclassified	62	78	**140**
	Total	**1503**	**117**	
	Z value	**2.289**		

The above results demonstrate that in this study, the grouping of classes had a bearing in the overall performance of the algorithms. There was, statistically a significant difference in the perfomance of RF and SVM after grouping the classes. The above test also highlights that non grouping of classes in the first set of classifications allowed the two algorithms to perform similarly. Thus, their ability to predict was similar before aggregating the classes but however, after grouping the classes their ability to predict changed. In this case SVM performed better than RF after terraces, pit structures and enclosures were aggregated as one class.

5 Discussion

Remote sensing techniques have proven to be efficient, low cost and applicable on large expanses of archaeological survey regions (Parcak 2009; Chen *et al*. 2017). As such the prime objective of this research was to ascertain whether advanced classification algorithms could be used to prospect archaeological sites in the Ziwa landscape, using high resolution Sentinel-2 imagery. The different classifications performed and predicted differently before and after aggregation of the classes.

The importance of each Sentinel-2 band in discriminating, archaeological sites (metalliferous surfaces, pit structures, terraces, enclosures) from non-sites and other LULC classes was assessed using the Random Forest algorithm (Figure 4). Bands in the infrared and visible spectra, were the most crucial in detecting archaeological sites. In both RF classification trials, the different bands in the multispectral sensor contributed differently to the detection of the land cover classes in question (Figure 3) which was expected due to the unique/similar, chemical, and physical makeup as well as mineral composition of the different classes (Rossel *et al*. 2006; Soriano-Disla *et al*. 2014). The model (Figure 4) presented the above-mentioned bands, as the bands that were crucial in the detection and identification of features. However, a combination of all the Sentinel-2 bands contributed immensely to the total spectral signature presented by the different features. Thus, in future research, although there will be major bands contributing to the discrimination of archaeological features, all the bands within the Sentinel-2 multispectral sensor must be considered.

Both classified images, independent of the method used for classification, show a higher intensity of archaeological features in the North- Eastern part of the map. Although the single archaeological features were not successfully discriminated in both trials, their aggregated footprint is clear. As such, the predictive ability of the classifiers to identify an archaeological landscape of interest, as opposed to single features, is noted. This is very useful for archaeologists as it can enable those with interest to direct their work within the predicted area and can then envisage the archaeological landscape of interest, to be confined to the North Eastern part of Ziwa as predicted by the classification map (Figures 5 and 6). This information is crucial as it defines and confines the area of ground surveying to a particular area, thus saving time, resources and efforts. In the first trial, the metalliferous surface was characterized by infertility and absence of material protection by crops/vegetation (Kritzinger 2010). In other words, these areas were bare or had little to no vegetation cover. These areas are thought to have been runoff mine waste which contained heavy metals which drained out of the tanks and mine workings residue collecting and concentrating at the bottom of the slope (Thornton & Chipangura 2017). This land cover class was described in this manner as informed by the hypothesis about its existence as an archaeological mining site which was made by recent studies (Kritzinger 2017).The presence of heavy metals on the soil made the surface dark and greyish in colour which was an obvious anomaly as other surrounding soils were brighter in colour.

In detecting the metalliferous surfaces, the SWIR (1565-1655nm), Blue (458-523nm) and Red Edge 2 (733-748nm) bands were the most important bands (Figure 4). This could be due to the sensitivity of the red edge band to heavy metal concentrations on surface soils as was the case in Clevers & Kooistra (2003) as well as in Mohamed *et al*. (2016). The visible and infrared spectra have been critically used for the detection of heavy metals in soils in particular, that is chromium, manganese and copper (Choe *et al*. 2008; Mohamed *et al*. 2016; Wang *et al*. 2018). These results are in line with the findings of Thabeng *et al*. (2019), were vitrified dung had a whitish colour and thus became brighter than the surroundings and was detected by bands within the visible spectrum. This particular aspect of this research confirms other scholar's findings and hence may be used by archaeologists in prospecting presence of archaeological mining areas with heavy metal residues. Furthermore, these bands are important regardless of whether one is looking for metalliferous surfaces or vitrified dung therefore, it may be a widely applicable application for archaeologists, even those not looking for mining sites.

The terraces are made of piled rock and covering every sloping surface and these are thought to have been used for trapping surface gold in runoff from the mountains (Kritzinger 2017). On the other hand, earlier scholars described terraces as loose, dry-stone 'walls' that were thought to have been used as fields for crops (Soper 1996; Kritzinger 2008). The Red Edge 2 (733-748nm), SWIR (1565-1655nm) and Red (650-680nm) bands contributed immensely to the discrimination of terraces from other LULC classes within the site (Figure 4).

Enclosures were discriminated from other LULC classes through contribution made by the SWIR (1565-1655nm), Red Edge 1 (698-713nm) and SWIR 2 bands (2100-2280) (Figure 4.).

RF also established that, for the discrimination of pit structures from other classes, the Red Edge 1 and 2 and blue bands contributed the most. Pit structures have been described as 'stone line tanks,' or hydraulic tanks built using bedrock 1.8m -3m deep and 4-9m in diameter by Kritzinger (2010). Kritzinger (2012) argues that pit structures were used to recover heavy metals by allowing water flow through a slot and allow water to drain out through a drain exit living the heavy metal concentrate. Other scholars however, suggested that the pit structures could have been used as cattle pens for miniature cattle (Soper, 1996). The above notions are still an ongoing debate in which there has not been a consensus and more work still needs to go into understanding the heritage site (Chirikure & Rehren 2004). The pit structures can be easily seen on the predictive RF model which performed optimally showing possible areas in which the archaeological feature could be located. Grouping the above classes together as human made structures was not considered but instead they were grouped as stone structures. This term captures their structure and gives the reader a picture of structures made of stone/rock. Whereas the term 'human made structures' might not specify the material used thus leaves room for other assumptions.

Peaks were observed in soil structure with high concentration of metals. The metalliferous surface has been proven to contain a significant amount of metal-based minerals that could be attributed to mine residue which could have been a by-product from archaeological mining activities (Thornton & Chipangura 2017). According to Chipangura (2019), small scale informal artisanal mining has been reported within Mutanda, an area close to the Ziwa site. It is possible that the mining location used in the past is still active today, but this requires further investigation to establish to what extent within Ziwa.

The above three classes, pit structures, terraces and enclosures share similar attributes as they are made from, hard rock derived from granite parent material. The bands that can contribute to their discrimination are red edge, SWIR and red as well as blue.

The above band ranges are covered by the visible near infrared (VNIR) and short wave infrared regions (SWIR). This range of spectra has been used widely in geological mapping (Hewson, 2005; Van Der Werff & Van Der Meer 2016). The various proportions of phyllosilicates, other silicates, carbonates, magnesium hydroxide groups, mafic groups and other mineral groups have led to different SWIR absorption variations (Hewson 2005). This is an interesting find highlighting that the various combinations of minerals within the rock material making the three archaeological features can be used in distinguishing between them. It is possible that the mineral combinations within the rocks making up the pit structures, terraces, and enclosures is different. Therefore, this property may be used to distinguish them. Furthermore, the SWIR and VNIR were used by Chen (2007) in his research where he highlighted that identification of rock minerals was directly associated with SWIR, VNIR regions. Altaweel 2005 & Beck 2007 also agree that the SWIR bands produce high reflectance of many different soils, minerals and this could also have contributed to the sensitivity of the stone walls to this range.

Furthermore, to differentiate and map rocks, mineral indices have been used and these have played a crucial role in understating rocks (Van Dar Meer *et al*. 2012). These indices utillised the various absorption rates of the SWIR and VNIR as well as the thermal bands in distinguishing rock features based on their mineral compositions (Van Dar Meer *et al*. 2012). Thus results in this study confirm the prevalence of the SWIR and VNIR bands in extricating terraces, pit structures and enclosures from other LULC classes.

The bands that contributed to the distinguishing features of the above rock structure are similar to those that were used by Van Der Werff & Van der Meer (2016). Their study used hydroxyl bearing rocks, iron oxides and ferrous iron oxides bearing rocks where the distinctions were exposed using Landsat 8 bands 6/7, 4/2 and 4/6 and Sentinel-2A bands 11/12, 4/2 and 4/11 respectively in a quest to evaluate the potential of using Sentinel-2 (13 bands at 10m) and Landsat 8 (11 bands at 15m) in geological mapping. These bands were namely the SWIR1/SWIR2, Red/Blue and Red/SWIR for Landsat 8 while Sentinel-2A yielded SWIR/SWIR2, Red/Blue and Red/SWIR1 which was consistent with studies done by other scholars (Van Der Werff & Van der Meer 2016). Van Der Werff & Van der Meer (2016) aimed at evaluating Sentinel-2A and Landsat 8 band ratios that were relevant for geological remote sensing ad they concluded that the VNIR and SWIR bands in mapping and discriminating geology. Heir research confirm that it is possible to distinguish rock the pit structures, enclosures and terraces through using the SWIR and VNIR bands which distinguish them through the different mineral compositions that could have been present. Thus, their research confirms what this research showed the importance of the SWIR bands and VNIR bands in discriminating the different rock structures within the site.

Band ratios involving the SWIR and VNIR reveal an interesting finding that can be used by archaeologists to uniquely identify, archaeological mine features made from similar rock material.

This clearly demonstrates that the signature picked up was as a result of the geology of the bed rock used to build the pit structure, enclosures and terraces. Furthermore, the area is dominated by granite which was used for building the pit structures (Soper 2006).

In a second set of classifications, the classes, pit structures, terraces and enclosures were combined as class 'stone walls'. In detecting this new class from other LULC classes, RF determined that bands, Red Edge 1, SWIR 2 and SWIR contributed more (Figure 4). The results were consistent with the results that had been highlighted in the first set of classifications.

Furthermore, RF assigned more of the area to archaeological sites than SVM in the second set of classifications. This was a total of 22.79km2 and 16.8km2 for RF and SVM of the study area respectively (Table. 3.1). This is an over prediction. It would be better for a model to perfom optimally than over predict or under predict. Over prediction might ensure that potential sites would not be lost whereas under predicting might lead to lose of potential sites. Over time archaeological sites/features tend to be eroded/washed away form their original place thus, an over prediction would compensate for this mass movement over time due to natural and anthropogenic activities. Over prediction or under Prediction highlights the precision, bias and accuracy of the model (Atkinson & Foody 2002) in this case an over prediction would increase both the precision and accuracy (Atkinson & Foody 2002). Over prediction also reduces the bias/ uncertainty of a model as a result of human manipulation and statistical issues (Atkinson & Foody 2002).Thus, over prediction presented by RF prevented the missing of potential mining sites by allocating more area to the stone walls compensating for possible displacement due to natural and anthropogenic activities. Over prediction may lead to exaggeration of results thus compromising accuracy of predictions

A major finding in this research is that the archaeological features became prominent and easily predictable after some major classes had been aggregated even without high resolution imagery. When using medium resolution imagery and advanced classifiers, best results are achieved when some archaeological features are aggregated than when they are standalone classes. Since archaeological features do not occur in isolation aggregation of the features is necessary in mapping. That is, it is evident the archaeological footprint became pronounced when signatures of different features were aggregated.

Another interesting finding in this research is that the RF and SVM are equally reliable and their results were not far apart. This observation is in line with the findings of Belgiu & Dragut (2016), Dal Ponte *et al*. (2013), Ghosh *et al*. (2014) and Thabeng *et al*. (2019).The metalliferous surfaces in this second set of classifications were discriminated from the rest of the LULC classes through the contribution made by the SWIR, red edge 1 and blue bands (Figure 4). This shows that the discrimination of metalliferous surfaces, amongst other classes, was effective in a gold ore residue area.

Comparing the two algorithms proved beneficial in discriminating these archaeological mining sites from non-sites. In the first trial RF and SVM achieved 80.67% and 80.17% while in the second trial they achieved 80.25% and 80.62% respectively. A MacNemars' test of significance done on the two sets of prediction accuracies for both classifiers, indicated that the predictions of the two classifies were statistically not significantly different for the first trial but significantly different for the second trial. Thus, aggregating the classes had a bearing in the performance of the algorithms.

The user and producer accuracies achieved by the algorithms in predicting the archaeological sites from non-sites were also different in both sets of classifications. Both the user and producer accuracies were satisfactory in both cases.

As previously expressed, most archaeological features were found to be in the highlands of the north eastern part of Ziwa. Highlands were crucial in early artisanal mining as the slope presented by the highlands presented a pivotal run off platform, which in turn facilitated surface gold to flow downward towards the 'gold trapping ' barriers better known as terraces built to trap gold below the mountain escarpment (Belford no year; Kritzinger 2012; Kritzinger 2017).

The results of this study show that the archaeological features seem to have existed in close proximity to each other (Fig 5 & 6). In other words, they seem to be grouped or clustered in one area and this could be attributed to archaeological association of the archaeological site; however this needs further research as it could also be due to simple convenience of the mine workings. Therefore, predictive modeling using machine learning, even with medium-resolution imagery such as Sentinel-2, can detect archaeological mining landscapes with good accuracy

6 Conclusion

The objective of this research was to conduct a supervised classification on Sentinel-2 imagery to develop a LULC map for the detection of archaeological features within the Ziwa landscape. The algorithms performed consistently and produced a reasonable level of accuracy. According to the classification the archaeological features were confined to the north eastern part of the study area. The features existed in close proximity to each other which could demonstrate a relationship between the features; further research could investigate this. RF forest performed better than SVM and this could have been attributed to its strengths that were highlighted earlier. Ore surfaces and grinding area could not be discriminated as they were small (<1 m in size) and fell way below the sensor's resolution (10 m).

RF and SVM managed to perform a LULC map but could not discriminate convincingly the terraces, pit structures and enclosure apart. The models only outlined the possible area in which these features could be located in the greater part of the archaeological landscape. Spectral signatures for discriminating pit structures, enclosures and terraces from each other could not be established as these features were predominately made of rock, thus the signatures were similar. Therefore, although the models can determine the general area where archaeological mining took place, it remains essential to conduct a foot survey to distinguish these individual features.

This research demonstrates that construction of accurate models for predicting archaeological sites, primarily on the basis of their spectral signatures, comes with challenges. To obtain optimal results, the use of higher resolution sensors is recommended as well as using supplementary platforms such as LiDAR, Radar and Object based image analysis methods. This needs to be supported with effective field surveys so as to conduct ground truthing.

This study has demonstrated the importance, abilities and shortcomings of advanced algorithms in the prospection of archaeological sites in the mining area of Ziwa. Furthermore this study has shown that the detection of archeological features is possible using multispectral sensors if the features are of a certain size, form and material. SVM and RF both show potential in the initial stages of surveying of an archaeological landscape especially in areas where the surface sediment is of interest, such as the case in this mining site. However there still is need to acquire high resolution imagery from commercial satellites and aerial resources in order to improve archaeological surveying. This will ensure higher spatial resolution that will enable robust and more accurate feature discrimination and prospection. Sentinel-2 is FREE and therefore archaeologists and heritage managers whose budgets are limited, are still able to derive good data through using them.

Limitations and Recommendations

There were a host of challenges in conducting this study, and one of them was in carefully discriminating the signature between naturally occurring stones on the site and the stones that were used to build the terraces, pit structures and enclosures. This could be due to the fact that the three features were predominantly made up of the same matte (rock) thus discriminating these classes was a not an easy task as they possessed very similar spectral signatures (Soriano-Disla *et al*. 2014). Aggregating these classes did however improve the performance of the models. Furthermore, these rock structures had different, shapes, sizes and geometry, terraces were long extensions of piled rock that form a line, enclosures were huge circular structures made up of stacked rock with high walls while pit structures were smaller and circular in form. Advanced supervised pixel based classifies (RF and SVM), employed in this research, did not consider the geometry, shape, size and texture of features, rather, the spectral signature only (Weih *et al*. 2010). The archaeological class (enclosures, terraces and pit structures) had similar signatures (Table 3) as they were basically made of the same material (rock) therefore discriminating them proved to be a challenge.

Having realized the above issue, object-based image analysis classification (OBIA) would be proposed to encourage effective feature discrimination, identification and prospection as used by Hay & Castila 2006. This is due to the ability of OBIA to utilize not only the class signature but also the geometry, shape and texture of the materials in question there by giving more robust and accurate classification than pixel-based classification (Sevara *et al*. 2016; Davis 2019). Technical workings of OBIA were highlighted by Sevara *et al*. (2016) and studies such as the one done by Belgiu & Csilik (2018) demonstrate that overall accuracies of OBIA cannot be matched with pixel-based outputs.

The pixel-based approaches has been quite critical in this research and its merits could be used to take advantage of both OBIA and pixel based classifications benefits. The two classifications could be used in conjunction with map mangroves using high resolution imagery as was the case in a study done by Wang *et al*. (2004). Future research could consider combining the two so as to improve the classifications and discrimination of similar archaeological features. Biagetti *et al*. (2017), found that pixel-based methods of image classification are limited as pixels are described as not true objects and overlook the spatial photo interpretive elements such as texture, context and shape. Thus, the motive to combine both OBIA and pixel-based methods allows to benefit from all its positive attributes.

From the foot survey, not only did the pit structures, enclosures and terraces have unique shapes, they also had unique heights and spatial patterns which the advanced pixel-based analysis failed to depict. This was also a challenge as these properties could not be used to clearly discriminate the archaeological features using the advanced classifiers. To address the above challenges, height characteristics of these archaeological features could be used to discriminate them. Light detection and ranging (LiDAR) is recommended in discriminating features, made from similar material, having different heights and covered in thick vegetation cover.

LiDAR technology employs laser technology whereby separate distance is measured producing 3D cloud points of canopy and the target features (Glennie *et al*. 2013). This technology could have been used in this research at a scale which is fine (10cm) to accurately detect negligible elevation differences in huge expanses of land (Chase *et al*. 2017). To sieve out the vegetation canopy filtering algorithms could be used to classify the return (Johnson & Ouimet 2013; Chase *et al*. 2017). Having highlighted this, terraces, pit structures and enclosures can be easily discriminated without confusion regardless of them having the similar spectral signatures due to them having unique, heights and dimensions. This technology was endorsed by Johnson & Ouimet (2014), where stone wall networks, building foundations amongst other archaeological features in an area of dense vegetation were observed. For effective LiDAR technology surveying and identification of archaeological mining remains, Principal component analysis, and multiple hill-shading are essential tools as was demonstrated by Fernandez-Lozano & Guitterez-Alonso (2016) in prospecting Roman Gold district archaeological features. Archaeologists do not usually use LiDAR because it is a relatively expensive technique when compared to other technologies. Unlike pixel-based approaches, this technology generates point clouds and the acquisition, processing, analysis, and interpretation of this type of data is relatively difficult. However, it would be ideal to use LiDAR in this research as the pit structures, terraces and enclosures had unique heights, geometry, dimensions, texture, shapes, and sizes LiDAR effective.

Another approach in discriminating them could be the use of band ratio combinations as was the case in Van Der Werff & Van der Meer (2016) and Hassan & Sadek (2017).

Another major challenge with the study site was that Ziwa has undergone very intense erosion processes which could lead to the models' results being misleading. Rampant erosion in archaeological sites causes sediment to be washed and deposited in other areas (Thabeng *et al*. 2019). In this research and previous studies, it was established that metalliferous materials were spread right through the sites but could not be easily identified as sediment was washed downslope. Previous studies have shown that erosion processes such as fluvial, aeolian, slope as well as anthropogenic activities are constantly acting on exposed archaeological exposed features especially in upland areas such as was the case with Ziwa (Kincey *et al*. 2014).

To assist in robust and efficient surveying of such landscapes archaeological site vulnerability modelling could be done in Ziwa and this would be in line with the research done by O'Rourke (2017) where identification of threatened archaeological remains was done using erosion models and change detection modeling.

Another challenge was the presence of grass and shrubs within the stone structures. These added to the overall signature recorded by the sensor thus contributing to errors although a season with the least vegetation cover was chosen. The spectral signature that was recorded by the Sentinel-2 image could have been misleading due to the presence of vegetation in and around the rock structures.

To address this, the use of Synthetic Aperture Radar (SAR) data is recommended, this technology has the ability to penetrate beyond the vegetation canopy as was demonstrated by Lasaponara & Masini 2013, as well, radar data has the ability to penetrate through the prevailing harsh atmospheric conditions. Although radar data is reported to be a complex data set to work with it presents more accurate data features as the stone walled structures provide much needed scattering due to the roughness presented by the Archeological features. SAR does come in very high resolution and is flexible. Its scanning scheme may be customised to get the best out of the scatter from the target object (Fomaro & Pascazio 2014). On the other hand, SAR data comes with computational complexity and high cost in the, acquisition, processing and analysis of the data type that may present (Fomaro & Pascazio 2014). Ground Penetrating Radar (GPR) would not be ideal in this situation (Keay *et al*. 2014. The structures are exposed thus spaceborne Radar such as SAR would be ideal as the backscatter will outline the archaeological features, geometry, orientation, as well as geophysical properties (Lasaponara & Masini 2013). This was demonstrated by a study done by Chen *et al*. (2015). Multi-polarization, multi-band SAR datasets have the ability to enhance visualization as a Red, Green, and Blue (RGB natural color) image. This is achieved using the multibands thus making the archaeological features in question recognizable by the untrained eye for visualization purposes (Lasaponara & Masini 2013).

Another challenge to the study was in the resolution of the sensor that was chosen. The area was not vast and use of a multispectral sensor onboard a different platform (airplane) would have been more appropriate. Sentinel-2 has 10, 20 and 60 m spatial resolutions, 5 day revisit time, 13 spectral bands and a field of view at 290 km (Van De Meer *et al*. 2014). With the above specifications, archaeological features interpreted by recent archaeologists as slag, furnaces and grinding pits could not be included in the research. These important features played a crucial role in the archaeological mining site as they were the elements and residues in the archaeological mine (Thornton & Chipangura 2017). However due to their sizes they could not be identified as they were very small; their diameters fell way below the 10m spatial resolution of Sentinel-2.

To mitigate this, the use of commercially available high-resolution imagery such as World view 2, SPOT, Pleiades and Quickbird employed by Thabeng *et al*. (2019) is recommended. This has a finer spatial resolution at 0.46cm brought about high accuracies as was in the case of Megarry & Davis (2013) where Worldview2 multispectral imagery was used to identify and catalogue new sites in and around a World Heritage Site in Ireland. Very high-resolution imagery is important and Zerbini & Fradley (2018) highlight that the accurate archaeological assessment of the terrain becomes challenging thus, the need to have high resolution. A spatial resolution of 0.5m would mean that any archaeological feature with a horizontal surface greater than 0.5m, as well, larger features become visible (Zerbini & Fradley 2018). Thus, with high resolution imagery, structures such as, grinding sites and furnaces could be visible and these can be included in future modelling and prospection but this could prove very costly

The satellite imagery that was chosen due to it being freely available was thus course to some extent and future research could look into the use of commercially high-resolution imagery in archaeological mine prospection. Unlike Israel and Palestine where high resolution imagery is not available to the public due to the narrow restrictions in imagery retail and legislation (Zerbini & Fradley 2018), imagery for the Ziwa region is available on retail and should be endorsed in future research.

Having highlighted the above limitations and recommendations, advanced classification algorithms RF and SVM used in this research demonstrated satisfactory results. There was consistency in their depiction of the study site. Both algorithms did not manage to discriminate individual archaeological features but managed to prospect the presence of archaeological site within the greater Ziwa terrain.

This research will direct and define the potential area in Ziwa for future research that would be similar or aligned to this work as the findings in this research could save time, energy and resources. This study also highlights the most important bands in multispectral sensors which could potentially be used to identify future stone wall sites within the region. These bands are covered within the VNIR and SWIR spectrum for the identification of archaeological features predominantly made from rock.

Another important aspect of this research is that it will contribute in the debate on whether the area was an early agricultural site or mining site.

An important issue regarding archaeological remote sensing must be noted: where remote sensing techniques can prove their usefulness for archaeological research, they cannot provide any information about the chronology, cultural pertinence, and function of the buried archaeological remains (Zanni & De Rosa 2019).

This research highlights that, for the discrimination of pit structures, terraces and enclosures the major spectrums ranges that contributed were the Red Edge, Red, Blue and Short Wave Infrared regions (SWIR).

Acknowledgements

Professor Robert Thornton provided funding and direction for this study under his NRF grant No. CPRR 109871 project entitled "Artisanal and Small-Scale Mining in South Africa: Ethnographic and archaeological Perspectives." A special thanks to the National Research Foundation (NRF) for providing the necessary funds under Professor Thornton's direction that enabled me to conduct the Masters research. I thank the University of the Witwatersrand for the opportunity to fulfil my study prospects and for their support in making this research possible.

Moreover, I would like to extent a special gratitude to the National Museums and Monuments of Zimbabwe for affording me the opportunity to conduct this study in their Ziwa archaeological site, an area rich in beauty and archaeological wealth.

My fieldwork in Zimbabwe was made easy through the assistance from Dr. Njabulo Chipangura of the Mutare Museum, Mutare Zimbabwe, and the staff of the Ziwa and Nhahokwe Heritage area including Nqobani, Johannes, Shingi, and Stanley I really appreciate your guidance that made the field work interesting and worthwhile.

To my supervisors, Professor Robert Thornton and Dr Stefania Merlo, your guidance and knowledge has enabled me to be better within this sphere and I am looking forward to more research work with you.

To my family, you were my strength and comfort in tough times you are greatly appreciated.

References

Abate, N., Elfadaly, A., Masini, N. and Lasaponara, R. 2020. Multitemporal 2016-2018 Sentinel-2 Data Enhancement for Landscape archaeology: The Case Study of the Foggia Province, Southern Italy, *Remote Sensing* 12: 1309-1335.

Abburu, S. and Golla, S.B. 2015. Satellite Image Classification Methods and Techniques: A review, *International Journal of Computer Applications* 119: 20-25

Adam, E. and Mutanga, O. 2009. Spectral discrimination of papyrus vegetation (*Cyperuspapyrus L.)* in swamp wetlands using field spectrometry, *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 612-620.

Adam, E, Mutanga, O, Odindi, J. and Abdel-Rahman, E.M. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers, *International Journal of Remote Sensing* 35: 3440-3458.

Agapiou, A. and Sarris, A. 2018. Beyond GIS layering: Challenging the (Re) use and fusion of archaeological prospection data based on Bayesian Neural Networks (BNN), *Remote Sensing* 10: 1762-1784.

Agapiou, A., Alexakis, D.D. & Hadjimitsis, D.G. 2014a. Spectral sensitivity of ALOS, ASTER, IKONOS, LANDSAT and SPOT satellite imagery intended for the detection of archaeological crop marks, *International Journal of Digital Earth* 7: 351–372.

Agapiou, A., Alexakis, D.D., Sarris, A. & Hadjimitsis, D.G. 2014b. Evaluating the potentials of Sentinel-2 for archaeological perspective, *Remote Sensing* 6: 2176–2194.

Altaweel, M. 2005. The use of ASTER satellite imagery in archaeological contexts, *archaeological prospection* 12: 151–166.

Archer, K.J., Kimes, R.V. 2008. Empirical characterisation of random forest variable importance measures, *Computational Statistics and Data Analysis* 52: 2249–2260.

Atkinson, P.M. and Foody, G.M. 2002, *Uncertainty in Remote Sensing and GIS*: Southampton, John Wiley and Sons Ltd.

Baghdadi, N. and Zribi, M. 2016. *Optical Remote Sensing of Land Surface: Techniques and Methods*, Elsevier.

Banning, E.B., Hawkins, A.L. & Stewart, S.T. 2006. Detection functions for archaeological survey*, American Antiquity* 71: 723–742.

Beck, A.R. 2007. Archaeological site detection: The importance of contrast. In *Proceedings of the 2007 Annual Conference of the Remote Sensing and Photogrammetry Society*, Newcastle University, Sept. 11–14: 307–312.

Beck, A., Philip, G., Abdulkarim, M. & Donoghue, D. 2007. Evaluation of Corona and Ikonos high resolution satellite imagery for archaeological prospection in western Syria, *Antiquity* 81: 161–175.

Beck, A.R. 2007. Archaeological site detection: The importance of contrast, In *Proceedings of the 2007 Annual Conference of the Remote Sensing and Photogrammetry Society,* 307-312.

Belgiu, M. & Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions, *Journal of Photogrammetry and Remote Sensing* 114: 24–31.

Belgiu, M. and Csillik, O. 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, *Remote Sensing of Environment 204*: 509-523.

Ben-Hur, A., Weston., J 2010. A user's guide to support vector machines. *In: Data Mining Techniques for the Life Sciences* 223–239.

Biagetti, S., Merlo, S., Adam, E., Lobo, A., Conesa, F.C., Knight, J., Bekrani, H., Crema, E.R., Alcaina-Mateos, J. & Madella., M. 2017. High and medium resolution satellite imagery to evaluate late Holocene human–environment interactions in arid lands: A case study from the Central Sahara, *Remote Sensing* 9: 351- 372

Boser, B.E., Guyon, I. & Vapnik, V.N. 1992. A training algorithm for optimal margin classifiers, Pittsburgh, ACM, *In Proceedings of the FIfth Annual Worlaghop of Computational Learning Theory* 5: 144-152.

Butt, A., Shabbir, R., Ahmad, S.S. and Aziz, N. 2015. Land use change mapping and analysis using Remote Sensing and GIS: a case study of Simly watershed, Islamabad, Pakistan, *The Egyptian Journal of Remote Sensing and Space Science 18*: 251-259.

Burrett, R.S. 2006. "A Ziwa Tradition Site and Associated Lsa Material, Mount Buchwa, Zimbabwe," *The South African Archaeological Bulletin* 61: 102-104.

Breeze, P.S., Drake, N.A., Groucutt, H.S., Parton, A., Jennings, R.P., White, T.S., Clark-Balzan, L., Shipton, C., Scerri, E.M., Stimpson, C.M. and Crassard, R. 2015. Remote sensing and GIS techniques for reconstructing Arabian palaeohydrology and identifying archaeological sites, *Quaternary International* 382: 98-119.

Breiman, L. 2001. Random forests, *Machine learning* 45: 5-32.

Campana, S. 2017. Drones in archaeology, State-of-the-art and future perspectives, *Archaeological Prospection 24*: 275-296.

Cavalli, R,M,, Colosi, F., Palombo, A., Pignatti, S. & Poscolieri, M. 2007. Remote hyperspectral imagery as a support to archaeological prospection, *Journal of Cultural Heritage* 8: 272–283.

Cerra, D., Agapiou, A., Cavalli, R.M. & Sarris, A. 2018. An objective assessment of hyperspectral indicators for the detection of buried archaeological relics, *Remote Sensing* 10: 500-525

Chase, A.S., Chase, D.Z. and Chase, A.F. 2017. LiDAR for archaeological research and the study of historical landscapes, In *Sensing the Past* 89-100.

Chen, F., Lasaponara, R. & Masini, N. 2017. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring, *Journal of Cultural Heritage* 23: 5–11.

Chen, M., Su, W., Li, L., Zhang, C., Yue, A. and Li, H 2009. Comparison of pixel-based and object-oriented knowledge-based classification methods using SPOT5 imagery, *WSEAS Transactions on Information Science and Applications 3*: 477-489.

Chen, X., Warner, T.A. and Campagna, D.J. 2007. Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada, *Remote Sensing of Environment 110*: 344-356.

Chen, F., Lasaponara, R. and Masini, N. 2017. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring, *Journal of Cultural Heritage 23*: 5-11.

Chirikure, S. and Rehren, T. 2004. Ores, furnaces, slags, and prehistoric societies: aspects of iron working in the Nyanga Agricultural Complex, ad 1300–1900, *African Archaeological Review 21*: 135-152.

Chipangura, N., Chiripanhura, P. & Nyamagodo, S. 2017. Policy Formulation and Collaborative Management: The Case of Ziwa Site, Eastern Zimbabwe, *Museum International 69*: 118-127.

Chipangura, N. 2019. Artisanal and small-scale mining (Chikorokoza) of Gold in the Eastern Highlands of Zimbabwe: Archaeological, Ethnographic and Historical Characteristics, Unpublished PhD dissertation, Anthropology, South Africa, University of the Witwatersrand.

Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B. and Kim, K.W. 2008. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain, *Remote Sensing of Environment 112*: 3222-3233

Chutia, D., Bhattacharyya, D.K., Sarma, K.K., Kalita, R. and Sudhakar, S. 2016. Hyperspectral remote sensing classifications: a perspective survey, *Transactions in GIS 20*: 463-490

Clark, V.R., Timberlake, J.R., Hyde, M.A., Mapaura, A., Palgrave, M.C., Wursten, B.T., Ballings, P., Burrows, J.E., Linder, H.P., McGregor, G.K. and Chapano, C. 2017. A first comprehensive account of floristic diversity and endemism on the Nyanga Massif, Manica Highlands (Zimbabwe-Mozambique), *Kirkia*, 19: 1-53.

Cihlar J. 2000. Land cover mapping of large areas from satellites: status and research priorities, *International Journal of Remote Sensing* 21: 1093–1114.

Congalton, R.G. and Green, K. 1999. *Assessing the accuracy of remotely sensed data: principles and applications*, Florida: Lewis Publishers

Caciagli, C. 2001, On precision in architecture, *Nexus Network Journal 3*: 11-15.

Davis, D.S. 2019. Object-based image analysis: a review of developments and future directions of automated feature detection in landscape archaeology, *Archaeological Prospection 26*: 155-163.

Davis, D.S., Douglass, K. 2020. Aerial and Spaceborne Remote Sensing in African archaeology: A Review of Current Research and Potential Future Avenues, *African Archaeological Review* 37:9–24.

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., Naesset, E. 2013. Tree species classification in boreal forests with hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing* 51: 2632–2645.

De Laet, V., Paulissen, E., & Waelkens, M. 2007. Methods for the extraction of archaeological features from very high- resolution Ikonos-2 remote sensing imagery, Hisar (Southwest Turkey), *Journal of Archaeological Science* 34: 830–841.

De Laet, V., Paulissen, E., Meuleman, K, & Waelkens, M. 2009. Effects of image characteristics on the identification and extraction of archaeological features from Ikonos-2 and Quickbird-2 imagery: Case study Sagalassos (Southwest Turkey), *International Journal of Remote Sensing* 30: 5655–5668.

de Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K., Skidmore, A 2006. Comparing accuracy assessments to infer superiority of image classification methods, *International Journal of Remote Sensing* 27: 223–232.

Denbow, J.R. 1979. Cenchrus ciliaris: An ecological indicator of Iron Age middens using aerial photography in eastern Botswana, *South African Journal of Science* 75: 405–408.

de Villiers, H. 1969. "Skeletal Remains from Inyanga, Rhodesia," *The South African Archaeological Bulletin* 24, no. Quaternary Studies in South Africa (95/96): 173.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. and Meygret, A 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services, *Remote sensing of Environment 120*: 25-36.

Du, P., Samat, A., Waske, B., Liu, S., Li, Z. 2015. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features, *ISPRS Journal of Photogrammetry and Remote Sensing* 105: 38–53

Eisenbeiss H., Lambers K., Sauerbier M., Li Z. 2005a. Photogrammetric documentation of an archaeological site (Palpa, Peru) using an Autonomous Model Helicopter, *In Proceedings of the CIPA 2005 XX International Symposium*, Torino, Italy, Camera and Imaging Products Association, 238-243.

Forte, M., Campana, S. and Liuzza, C. 2010. *Space, Time, Place*, London: Archaeopress.

Fornaro, G. and Pascazio, V. 2014. SAR interferometry and tomography: Theory and applications, *Academic Press Library in Signal Processing* 2: 1043-1117.

Fuldain González, J.J. and Varón Hernández, F.R. 2019. NDVI Identification and Survey of a Roman Road in the Northern Spanish Province of Álava, *Remote Sensing 11*: 725-744.

Gatti, A. and Bertolini, A. 2013. *Sentinel-2 products specification document*, Available online (accessed August, 2020) https://earth, esa. int/documents/247904/685211/Sentinel-2+ Products+ Specification+ Document.

Giardino, M.J. 2011. A history of NASA remote sensing contributions to archaeology, *Journal of Archaeological Science 38*: 2003-2009.

Ghosh, A., Fassnacht, F.E., Joshi, P., Koch, B. 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales, *International Journal of Applied Earth Observation and Geoinformation* 26: 49–63.

Glennie C.L., Carter W.E., Shrestha R.L., Dietrich, W.E. 2013. Geodetic imaging with airborne LiDAR: The earth's surface revealed, Rep Prog Phys 76:086801.

Govender, M., Chetty, K. and Bulcock, H. 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies, *Water SA 33*: 145-151.

Gualtieri, J.A. and Cromp, R.F. 1999. Support vector machines for hyperspectral remote sensing classification, International Society for Optics and Photonics In *27th AIPR Workshop: Advances in Computer-Assisted Recognition* 3584: 221-233.

Haddad, N. 2007. Towards creating a dialogue between the specialized technician and non-technician users of the 3D laser scanner, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36*: C53.

Haddad, N.A. 2011. From ground surveying to 3D laser scanner: A review of techniques used for spatial documentation of historic sites, *Journal of King Saud University-Engineering Sciences* 23: 109-118.

Hay, G.J. and Castilla, G. 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT), In *Proceedings of 1st International Conference OBIA*: 4-5.

Hesse, R. 2010. LiDAR-derived Local Relief Models - a new tool for archaeological prospection, *Archaeological Prospection 17*: 67-72.

Hewson, R.D., Cudahy, T.J., Mizuhiko, S., Ueda, K. and Mauger, A.J. 2005. Seamless geological map generation using ASTER in the Broken Hill-Curnamona province of Australia, *Remote Sensing of Environment 99*: 159-172.

Hornik, K., Meyer, D. and Karatzoglou, A. 2006. Support vector machines in R, *Journal of Statistical Software 15*: 1-28.

Hubbard, P. 2007. "Review of R. Soper, The Terrace Builders of Nyanga," *Journal of African Archaeology* 5: 343-344.

Huffman, T.N. 1986. Archaeological evidence and conventional Africa, explanations of Southern Bantu settlement patterns, *Journal of the International African Institute* 56: 280-298.

Hunt, T. and Karim S. 2014. "Inter-analyst variability in identification and classification of pre-colonial stone walled structures using GoogleEarth, southern Gauteng, South Africa (Field & technical report)," *The South African Archaeological Bulletin* 69:87-95

Immitzer, M., Vuolo, F. and Atzberger, C. 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe, *Remote Sensing 8:1*66-192

Jimu, L. and Ngoroyemoto, N. 2011. Habitat characteristics and threat factors of the rare and endangered Prunus africana (Hook. f.) Kalkman in Nyanga National Park, Zimbabwe, *International Journal of Biodiversity and Conservation 3*: 230-237.

Joseph, G. and Jeganathan, C. 2018. *Fundamentals of Remote Sensing,* 3rd edn, Telangana, India: Universities Press.

Keay, S.J., Parcak, S.H. & Strutt, K.D. 2014. High resolution space and ground-based remote sensing and implications for landscape archaeology: The case from Portus, Italy*, Journal of Archaeological Science* 52: 277–292.

Keeney, J. and Hickey, R. 2015. Using satellite image analysis for locating prehistoric archaeological sites in Alaska's Central Brooks Range, *Journal of Archaeological Science: Reports 3*: 80-89.

Kerle, N., Templfi, K., Janssen, L.L. and Huurneman, G.C. 2009. *Principles of remote sensing – An introductory textbook*, The Netherlands: The International Institute for Geo-Information Science and Earth Observation.

Klehm, C., Barnes, A., Follett, F., Simon, K., Kiahtipes, C. & Mothulatshipi, S. 2019. Toward archaeological predictive modeling in the Bosutswe region of Botswana: Utilizing multispectral satellite imagery to conceptualize ancient land-scapes, *Journal of Anthropological Archaeology* 54: 68–83.

Klehm, C. and Gokee, C. 2020. Geospatial Analysis in African archaeology: Current Theories, Topics, and Methods, *African Archaeological Review* 37:1–7.

Kincey, M., Batty, L., Chapman, H., Gearey, B., Ainsworth, S. and Challis, K. 2014. Assessing the changing condition of industrial archaeological remains on Alston Moor, UK, using multisensor remote sensing, *Journal of Archaeological Science 45*: 36-51.

Kritzinger, A. 2012a. 'Laboratory analysis reveals direct evidence of precolonial gold recovery in the archaeology of Zimbabwe's Eastern Highlands', In Proceedings of *9th International Mining History Congress,* Johannesburg 17-21 April.

Kritzinger, A. 2012. 'Laboratory analysis reveals direct evidence of precolonial gold recovery in the archaeology of Zimbabwe's Eastern Highlands', In Proceedings of *9th International Mining History Congress,* Johannesburg 17-21 April.

Kritzinger, A. 2017. *Gold Mining Landscapes of Nyanga: Discovering Zimbabwe's Hidden Heritage*, Zimbabwe: Time Horizons.

Kritzinger, A 2008. Gold not grain: precolonial harvest in the terraced hills of Zimbabwe 's Eastern Highlands, National Museums and Monuments of Zimbabwe *Cookeia* series 13: 1-44.

Kulkarni, V.Y. and Sinha, P.K. 2013. Random forest classifiers: a survey and future research directions, *International Journal of Advanced Computation* 36: 1144-53.

Kumar Mishra, R. and Zhang, Y. 2012. A review of optical imagery and airborne LiDAR data registration methods, *The Open Remote Sensing Journal 5:54-63*.

Laponara, R., Elfadaly, A., & Attia, W. 2016. Low cost space technologies for operational change detection monitoring around the archaeological area of Esna-Egypt, In O. Gervasi, B. Murgante, S Misra, AMA. C. Rocha, CM. Torre, D. Taniar, B O. Apduhan, E. Stankova, & S. Wang (Eds.), In proceedings of the *International Conference on Computational Science and Its Applications*, China 4-7 July 2016

Lasaponara, R. & Masini, N. 2013. Satellite Synthetic Aperture Radar in archaeology and cultural landscape: An overview, *Archaeological Prospection* 20: 71–78.

Lasaponara, R. and Masini, N. 2012. Image enhancement, feature extraction and geospatial analysis in an archaeological perspective, *Satellite remote sensing*. Netherlands: Springer.

Lasaponara, R. and Masini, N. 2006. Identification of archaeological buried remains based on the normalized difference vegetation index (NDVI) from quickbird satellite data, *IEEE Geoscience and Remote Sensing Letters 3*: 25-328.

Leisz, S.J. 2013. An overview of the application of remote sensing to archaeology during the twentieth century, In D. C. Comer & M. J. Harrower (Eds.), *Mapping archaeological landscapes from space*, New York: Springer.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E. and Gascon, F. 2016. Sentinel-2 Sen2Cor: L2A processor for users, In proceedings of the *Living Planet Symposium* 2016, viewed on 22 September 2020, <http://esamultimedia.esa.int/multimedia/publications/SP-740/SP-740_toc.pdf>

Lu, D. and Weng, Q. 2007. A survey of image classification methods and techniques for improving classification performance, *International journal of Remote sensing 28*: 823-870.

Luo, L., Wang, X., Guo, H., Lasaponara, R., Shi, P., Bachagha, N, Li, L., Yao, Y., Masini, N., Chen, F. and Ji, W. 2018. Google Earth as a powerful tool for archaeological and cultural heritage applications: a review, *Remote Sensing 10:*1558-1590.

Luo, L., Wang, X, Guo, H., Lasaponara, R., Zong, X., Masini, N., Wang, G., Shi, P., Khatteli, H., Chen, F. and Tariq, S. 2019. Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017), *Remote sensing of environment 232*:111280-111313.

MacQuilkan, P. and Sadr, K. 2010. Remote sensing of archaeological ruins: a comparison, *Position IT Nov/De*c 2010: 50-51.

McConnell W., Moran E.F. 2001. Meeting in the Middle: The Challenge of Meso-Level Integration, *Journal of Land Use Policy* 305: 1-3.

Mather, P. and Tso, B. 2016. *Classification methods for remotely sensed data*, USA: CRC press.

Mathur, A. and Foody, G. M. 2008. Multiclass and Binary SVM Classification: Implications for Training and Classification Users, *IEEE Geoscience and Remote Sensing Letters* 5: 241–245.

Mason, R.J. 1968. Transvaal and Natal Iron Age settlement revealed by aerial photography and excavation, *African Studies* 27: 167-180.

Masini, N. and Lasaponara, R. 2006. Satellite-based recognition of landscape archaeological features related to ancient human transformation, *Journal of Geophysics and Engineering 3*: 230-235.

Megarry, W. and Davis, S. 2013. Beyond the Bend: remotely sensed data and archaeological site prospection in the Boyne Valley, Ireland, In *Mapping Archaeological Landscapes from Space*, New York: Springer.

Mohamed, E.S., Ali, A.M., El Shirbeny, M.A., El Razek, A.A.A. and Savin, I.Y. 2016. Near infrared spectroscopy techniques for soil contamination assessment in the Nile Delta, *Eurasian Soil Science 49*: 632-639.

Mountrakis G., Im, J. and Ogole, C. 2011. Support vector machines in remote sensing: A review, ISPRS *Journal of Photogrammetry and Remote Sensing* 66: 247–59.

Mupira, P. 2003. 'The Case of Nyanga Cultural Landscape, N. E. Zimbabwe', A paper presented at *ICOMOS 14th Genx`eral Assembly: Place – memory – meaning: preserving intangible values in monuments and sites*, Victoria Falls, Zimbabwe 27-31 October 2003, viewed on 10 July 2018 < http://www.international.icomos.org/victoriafalls2003>

Mupira, P. 2011. 'An investigation of structural stability and causes of deterioration in rough dry stone monuments at Ziwa, NE Zimbabwe'*—A study conducted under the Global Heritage Preservation Fellowship Program, Final Report, February 2011,* National Museums and Monuments of Zimbabwe, Mutare Museum, P.O. Box 920, Mutare, Zimbabwe; and The Global Heritage Fund ,Harare.

Mutanga, O.E., Adam, E. and Cho M.A. 2012. High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random Forest Regression Algorithm, *International Journal of Applied Earth Observation and Geoinformation* 18: 399–340

Opitz, R & Herrmann, J. 2018. Recent trends and long-standing problems in archaeological remote sensing, *Journal of Computer Applications in Archaeology* 1: 19–41.

O'Rourke, M.J. 2017. Archaeological site vulnerability modelling: the influence of high impact storm events on models of shoreline erosion in the western Canadian Arctic, *Open Archaeology*, *3*: 1-16.

Pal, M and Mather, P. 2005. Support vector machines for classification in remote sensing, *International Journal of Remote Sensing* 26: 1007–1011.

Parcak, S.H. 2007. Satellite remote sensing methods for monitoring archaeological tells in the Middle East*, Journal of Field Archaeology* 32: 65–81.

Parcak, S.H. 2009. *Satellite Remote Sensing for Archaeology*, New York: Routledge.

Patias, P. 2006. 'Cultural heritage documentation', *International Summer School Digital Recording and 3D Modeling, Aghios Nikolaos, Creta, Grecia, April* 24-29 2006.

Petropoulos, G.P., Kalaitzidis, C. and Vadrevu, K.P. 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery, *Computers & Geosciences 41*: 99-107.

Pugin, J.M. 2016. *Locating the rock art of the Maloti-Drakensberg: Identifying areas of higher likelihood using remote sensing,* (MSc dissertation), Johannesburg, University of the Witwatersrand.

Reid, S.H. 2016. Satellite remote sensing of archaeological vegetation signatures in coastal West Africa, *African Archaeological Review* 33: 163–182.

Remondino, F. and Rizzi, A. 2010. Reality-based 3D documentation of natural and cultural heritage sites—techniques, problems, and examples, *Applied Geomatics 2*: 85-100.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification, *ISPRS Journal of Photogrammetry and Remote Sensing 67*: 93-104.

Rossel, R.V., Minasny, B., Roudier, P. and Mcbratney, A.B. 2006. Colour space models for soil science, *Geoderma 133*: 320-337.

Sahebjalal, E. and Dashtekian, K. 2013. Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods, *African Journal of Agricultural Research 8*: 4614-4622.

Sadr, K. & Rodier, X. 2012. Google Earth, GIS and stone- walled structures in southern Gauteng, South Africa, *Journal of Archaeological Science* 39: 1034–1042.

Sadr, K. 2016. A comparison of accuracy and precision in remote sensing stone-walled structures with Google Earth, high resolution aerial photography and LiDAR; a case study from the South African Iron Age, *Journal of Archaeological Prospection* 23: 95-104

Sadr, K. 2019. The Late Iron Age Type N stonewalled structures on the highveld of South Africa, Azania*: Archaeological Research in Africa* 54: 245-270.

Sadr, K. 2019. Kweneng: A newly discovered pre-colonial capital near Johannesburg, *Journal of African Archaeology* 17: 1-22.

Sadr, K. 2019. Kweneng: How to lose a precolonial city (Field & technical report), *The South African Archaeological Bulletin* 74: 56-62.

Saturno, W., Sever, T.L., Irwin, D.E., Howell, BF. and Garrison, TG 2006. Putting us on the map: remote sensing investigation of the ancient Maya landscape, In *Remote Sensing in Archaeology*, New York: Springer.

Sauerbier, M. and Eisenbeiss, H. 2010. UAVs for the documentation of archaeological excavations, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*: 526-531.

Saumagne, C. 1952. La photographie aérienne au service de l'archéologie en Tunisie, *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 96: 287–301.

Schaich, H., Bieling, C. and Plieninger, T. 2010. Linking ecosystem services with cultural landscape research, *Gaia-Ecological Perspectives for Science and Society 19:* 269-277.

Schmid, T., Koch, M., DiBlasi, M. & Hagos, M. 2008. Spatial and spectral analysis of soil surface properties for an archaeological area in Aksum, Ethiopia: Applying high and medium resolution data, *Catena* 75: 93–101.

Seddon, J.D. 1968. An aerial survey of settlement and living patterns in the Transvaal Iron Age: Preliminary report, *African Studies* 27: 189–194.

Shumba, M. 2003. 'Ziwa National Monument: its Significance to the Local People Hangs Delicately in the Balance', Papers from the *14th General Assembly and Scientific Symposium of ICOMOS*, Victoria Falls, October 2003, Available from:<http://www.international.icomos.org/victoriafalls2003>,Accessed 20 July 2018.

Sivitskis, A.J., Lehner, J.W., Harrower, M.J., Dumitru, I.A., Paulsen, P.E., Nathan, S., Viete, D.R., Al-Jabri, S., Helwing, B., Wiig, F. and Moraetis, D. 2019. Detecting and Mapping Slag Heaps at Ancient Copper Production Sites in Oman, *Remote Sensing 11*:3014-3037.

Soper, R. 1996. The Nyanga terrace complex of eastern Zimbabwe: new investigations, *Azania: Archaeological Research in Africa* 31: 1-35.

Soper, R.C. 2002. *Nyanga: Ancient Fields, Settlements and Agricultural History in Zimbabwe*, British Institute in Eastern Africa.

Soper, R. 2006. *The Terrace Builders of Nyanga*, Harare: Weaver Press.

Sooväli, H., Palang, H., Kaur, E., Peil, T. and Vermandere, I. 2003a. Combining approaches in landscape research. The case of Saaremmaa, Estonia. In: Palang, H& Fry, G. (eds). *Landscape interfaces, Cultural Heritage in Changing Landscapes*, London: Kluwer Academic Publishers.

Sooväli, H., Palang, H., Kaur, E, Peil, T. and Vermandere, I. 2003. Combining approaches in landscape research, In *Landscape Interfaces*, Dordrecht: Springer.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M. and McLaughlin, M.J. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties, *Applied Spectroscopy Reviews 49*: 139-186.

Strassburger, M. 2015. Pre-colonial gold mining in Nyanga, Zimbabwe, in Jacquo Silvertant, ed, *Mining archaeology Perspectives, Conflicts, Challenges,* Netherlands: Silvertant Erfgoedprojecten.

Strobl, C., Boulesteix, A.L., Zeileis, A, Hothorn, T. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution, *BMC Bioinformatics* 8:25-45.

Summers, R.F.H. and Inyanga Research Fund 1958. *Inyanga: Prehistoric Settlements in Southern Rhodesia,* Cambridge: Cambridge University Press.

Summers, R. 1966. The Iron Age of Southern Rhodesia, *Current Anthropology* 7: 463-484.

Summers, R. 1969. *Ancient mining in Rhodesia and adjacent areas*, Zimbabwe: Trustees of the National Museums of Rhodesia.

Taylor, K. and Lennon, J. 2011. Cultural landscapes: a bridge between culture and nature, *International Journal of Heritage Studies* 17: 537-554.

Tartaron, T.F. 2003. The archaeological survey: sampling strategies and field methods, *Hesperia Supplements* 32: 23-45.

Thabeng, O.L., Merlo, S. and Adam, E. 2019. High-resolution remote sensing and advanced classification techniques for the prospection of archaeological sites' markers: The case of dung deposits in the Shashi-Limpopo Confluence area (southern Africa). *Journal of Archaeological Science* 102: 48-60.

Thabeng, O.L., Merlo, S. & Adam, E. 2020. From the Bottom Up: Assessing the Spectral Ability of Common Multispectral Sensors to Detect Surface Archaeological Deposits Using Field Spectrometry and Advanced Classifiers in the Shashi-Limpopo Confluence Area, *African Archaeological Review* 37: 25–49.

Thornton J. R. & Chipangura. N. (no year). A late Iron Age metallurgical waste dumping site at Ziwa Zimbabwe Discovery and Analysis using remote sensing GoogleEarth microscopy and XRF analysis available at< https://www.academia.edu /33925343> (accessed 02/02/2018).

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B. and Nõges, T. 2016. First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery, *Remote Sensing* 8: 1-15.

Topouzelis, K., Psyllos, A. 2012. Oil spill feature selection and classification using decision tree forest on SAR image data, *International Society for Photogrammetry and Remote Sensing* 68: 135–143

Tuia, D., Muñoz-Marí, J., Kanevski, M. and Camps-Valls, G. 2011. Structured output SVM for remote sensing image classification, *Journal of Signal Processing Systems 65*: 301-310.

UNESCO 2015, *Database of National Cultural Heritage Laws* Updated, Available from: https://en.unesco.org/news/unesco-database-national-cultural-heritage-laws-updated, accessed` 5 October 2018.

UNESCO 2015. *Ziwa national monument*, Available from:<http://whc.unesco.org/en/tentativelists/903/>, accessed 10 October 2018.

Van der Meer, F.D., Van der Werff, H.M.A. and Van Ruitenbeek, F.J.A 2014. Potential of ESA's Sentinel-2 for geological applications, *Remote Sensing of Environment 148*: 124-133.

Van der Werff, H. and Van der Meer, F. 2016. Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing, *Remote Sensing 8*: 883-898.

Verhoeven, G.J. 2012. Near-infrared aerial crop mark archaeology: from its historical use to current digital implementations, *Journal of Archaeological Method and Theory 19*:132-160.

Wang, F, Gao, J. and Zha, Y. 2018. Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges, *ISPRS Journal of Photogrammetry and Remote Sensing 136*:73-84.

Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W. 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data, *The Crop Journal* 4:212–219.

Weih, R.C. and Riggan, N.D. 2010. Object-based classification vs. pixel-based classification: comparative importance of multi-resolution imagery, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38:C7-C12.

Winton, H. and Horne, P. 2010. National archives for national survey programmes: NMP and the English heritage aerial photograph collection, In *Landscapes through the Lens, Aerial Photographs and Historic Environment*, Pennsylvania: Oxbow Books.

Xie, Y., Sha, Z. and Yu, M. 2008. Remote sensing imagery in vegetation mapping: a review, *Journal of Plant Ecology 1*: 9-23.

Yan, L., Roy, D.P., Zhang, H., Li, J. and Huang, H. 2016. An automated approach for sub-pixel registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery, *Remote Sensing 8*: 520-542.

Zanni, S. & Rosa, A.D. 2019. Remote sensing analyses on Sentinel-2 Images: Looking for Roman Roads in Srem Region (Serbia), *Geosciences* 9: 25-42.

Zerbini, A., & Fradley, M. 2018. Higher resolution satellite imagery of Israel and Palestine: Reassessing the Kyl-Bingaman Amendment, *Journal of Space Policy* 44:14-28.

Zhong, B., Liang, T., Wang, L. and Li, K. 2014. Applications of stochastic models and geostatistical analyses to study sources and spatial patterns of soil heavy metals in a metalliferous industrial district of China, *Science of the Total Environment* 490:422-434