Lithological classification of Mashonaland West, Zimbabwe, using Sentinel-2 in Google Earth Engine

Tanaka P. Charuka [a*], Ever Kutsanzira [a, b], MacDonald Kasirori

[a] *University of Zimbabwe, Faculty of Engineering and Built Environment, Department of Geomatics Engineering, Harare, Zimbabwe. Orcid number (www.orcid.org)*

**Corresponding author: Email: tanaka.charuka@students.uz.ac.zw*

ABSTRACT

We hereby present a lithological map of Mashonaland West Province, Zimbabwe, based on a combination of Sentinel-2 and ASTER satellite data processed by means of the Google Earth Engine (GEE). The lithological classification scheme aspires to constitute a comprehensive form of integrated knowledge suitable for regions with complex geological environments, exemplified by the Great Dyke and greenstone belts, thus aiding mineral exploration and land use planning. The study uses a supervised Random Forest classification algorithm on multispectral data, obtaining high classification accuracy of operations for five lithological classes including granite, sandstone, limestone, shale, and basalt. The findings validate the successful combination of remote sensing technology with machine learning algorithms in perennial vegetation regions where in situ surveys can be problematic. The results add to the scientific knowledge of geological mapping by confirming machine learning approaches are applicable, and developing a replicable workflow for similar geological environments at a global scale. And such an approach enables the generation of sustainable land management decisions in Zimbabwe, as a more accessible alternative to high-end geological mapping instruments for resource-poor institutions.

Keywords: Sentinel-2, Lithological Classification, Remote Sensing, Google Earth Engine

Immediately after the abstract, please provide up to 5 key words and with each words seperated by comma. Initial letters of keywords must be written by big letters. Point must be added at the end of the line.

1. Introduction

Lithological mapping is an indispensable pre-requisite of structural geological studies, providing crucial constrains into the composition, structure and evolution of the Earth's crust. Conventional methods, which heavily depend on field surveys and laboratory analyses, are frequently limited by time, cost, and accessibility (EL-Omairi & El Garouani, 2023a). Remote sensing technology have become a powerful means for large-scaled, low-cost and non-invasive lithological mapping (EL-Omairi & El Garouani, 2023a). Mashonaland West, Zimbabwe, characterized by various geological formations, which include the Great Dyke, greenstone belts, and granitoid complexes was chosen as the site of the current study (ALEXANDER DYMOKE MASON-APPS, 1998) . The study employs the integration of satellite imagery from Sentinel-2 with processing and visualization steps taken in Google Earth Engine (GEE) and QGIS, and thus represents a new dimension of approaching lithology characterization in a geologically important region.

1.1 Background and Context

Zimbabwe's geologic background is largely controlled by the Archean Zimbabwe Craton, where key economical mineral deposits were formed (Jelsma et al., 2021). In particular, Mashonaland West contains a complex lithological assemblage comprising mafic-ultramafic rocks, metasediments and granitoids (Mudimbu et al., 2022). Therefore, accurate lithological mapping in this region is critical for mineral exploration activities, land-use planning, and environmental management. On the other hand, such factors as the dense vegetation cover and restricted access to the field demand new solutions like the application of the remote sensing.Sentinel-2 offers high spatial (10–60 m) and spectral (13 bands) resolution and is well suited for rock type discrimination based on color (Chen, Dong, et al., 2023).

1.2 Research Objectives

This study aims to:

1. Establish a comprehensive lithological classification scheme of Mashonaland West based on Sentinel-2 imagery.
2. Use GEE to process data, run statistics, and develop algorithms quickly.
3. Confirm the classification results against field data and preexisting geological maps.
4. Illustrate how to visualize and interpret lithological maps using QGIS.

1.3 Significance of the Study

The research impacts scientists and geological mappers. Utilizing Sentinel-2 image data within the Google Earth Engine (GEE) platform to conduct a lithological classification at the level of major rock types, circumvents key limitations of field-based methods and demonstrates an efficient way to cost-effectively classify lithology in often difficult settings such as Mashonaland West (EL-Omairi & El Garouani, 2023a). The method of utilizing machine learning algorithms on multispectral data show a considerable improvement in identifying the lithology even in an area with dense vegetation which is one of the challenges in remote sensing methods (EL-Omairi & El Garouani, 2023a).

This work is especially relevant for Zimbabwe's mineral sector. The study provides an important tool for mineral exploration in the economically important Great Dyke and greenstone belt areas, reducing exploration costs and environmental impact, and increasing targeting accuracy. The combination of these advantages, alongside the use of open-access satellite data and cloud computing, is particularly valuable for cash-strapped institutions, greatly democratizing the access to advanced geological mapping capabilities.

Apart from potential applications, this work adds to the wider discipline of geology with remote sensing by:

1) Proving the power of sensor fusion methods in demanding geological environments

2) Validation of ML methods for litho-classification in sub-tropical settings

3) A reproducible workflow that can be applied to similar areas around the world

The research findings will be applied directly in the field to inform sustainable land management decisions in Zimbabwe and will provide a blueprint for geological mapping in other developing regions with similar geological and vegetation constraints. This innovative work paves a way for the technological advancement to be successful in the practical application domain in geology all around the world by ensuring the cutting-edge remote sensing techniques remain accessible and actionable for a global audience of different geological implications.

1.4 Literature Review

The lithological classification of remote sensing data has been revolutionised by technology for the sensors and new computation tools, along with improved data processing methods. Here, we summarize relevant methodological advances, challenges and innovations, embedding the research reported herein into the ongoing arc of refining remote sensing for geological mapping.

Early Foundations and Multispectral Breakthrough

The history of remote sensing for lithological mapping began with the launch of the Landsat Multispectral Scanner (MSS) in 1972 (Chen, Wang, et al., 2023a). Despite the limited spectral resolution (four bands) and spatial resolution (80 m), MSS data were found to have a strong potential for regional-level lithological discrimination. The second significant leap occurred with the introduction of the Thematic Mapper (TM) sensor. TM data were integrated well for mapping hydrothermal alteration zones which were significant for mineral exploration (Knepper, 2012). In vegetated situations, where the spectral signature was diluted by lithology, the lithological signature was lost.

The launch of the Sentinel-2 mission addressed the scalability issues with a 13 band multispectral sensor including bands in the red-edge critical for vegetation-penetrative lithological mapping (Bentahar & Raji, 2021) . A study by Bentahar & Raji (2021), found Sentinel-2 outperforming Landsat in classifying forested terrains, highlighting its merit in areas such as Mashonaland West, where vegetation coverage prevents bedrock exposure.

Algorithms and Computational Progress

Merging ML algorithms has improved classification quality (EL-Omairi & El Garouani, 2023b). While methods like maximum likelihood classification were limited by parametric assumptions, non-parametric algorithms such as Random Forest (RF) and Support Vector Machines (SVM) were developed to address high-dimensional, complex data more easily (Serbouti et al., 2022). Using RF on Sentinel-2, higher accurate lithologic mapping is attained, which the authors attribute to RF's inherent robustness against noise and ability to rank feature importance (Chen, Dong, et al., 2023). Likewise, SVM working with the kernel mechanism has superior capability to separate non-linear spectrum grouped clusters. Hence, it perfectly suits heterogeneous terrains as well.

Cloud Computing platfarms, especially Google Earth Engine (GEE), have democratized the access to high-scale remote sensing anlyses. GEE has a petabyte-scale catalog, as well as parallel processing capabilities, that allow real-time classification of global datasets. Combination of GEE with Sentinel-2 data was performed to carry out lithological mapping within the Paleozoic massif of Rehamna, achieving data processing in a cost and time efficient manner (Serbouti et al., 2022). Such scalability also resonates well with the needs of regions with limited resources like Zimbabwe, whereby conventional GIS workflows are rendered prohibitive.

Issues and Solutions

Nevertheless, some challenges still remain, such as atmospheric interference, spectral mixing, vegetation cover, types of lithological units and their mineral compositions, and sensitivity to weathering (Serbouti et al., 2022). Atmospheric corrections (e.g., FLAASH) are essential for obtaining above water vapor and aerosol affects. For example, spectral unmixing techniques, e.g. Linear Spectral Unmixing (LSU), have been used to address mixed pixels in complex lithologies (van der Meer et al., 2012). For vegetated areas, rock spectra are separated using indices such as NDVI and SAVI, but newer techniques, such as spectral feature fitting (SFF), have gained popularity (van der Meer et al., 2012).

The presence of data gaps as a result of cloud cover or temporal inconsistencies makes analyses even more complicated. Solutions involve temporal compositing (median-pixel mosaics, for instances) and fusion multi-sensor data (Chen, Wang, et al., 2023b).

Synthesis and Knowledge Gaps

This review highlights the evolution in multispectral and machine learning-based lithological classification systems. However, there is a limited application of these methods to highly vegetated and tectonically complex regions such as the Mashonaland West region. This study contributes to the above knowledge gaps by integrating Sentinel-2 data within the Google Earth Engine (GEE) environment, utilizing both Support Vector Machine (SVM) and Random Forest (RF) classifiers and validating against geolocical maps.

2. Main Text

2.1. Material and Method

This section describes the methodology used for lithological classification of Mashonaland West, Zimbabwe, utilizing Sentinel-2 imagery and Google Earth Engine (GEE) platform, with ArcGIS employed for study area delineation and lithological map generation.

2.1.1. Study Area

Mashonaland West is a mineral-rich province in northern Zimbabwe, known for both agriculture and mining. . It lies between approximately 16°S to 18°S latitude and 28°E to 30°E longitude. It includes key towns such as Chinhoyi and Kadoma and lies within the ancient Zimbabwe Craton. The province hosts important lithological units such as greenstone belts, granitic intrusions, and Karoo sediments. These formations are associated with gold, chromite, copper, and limestone deposits. Notable mineralized zones include the Mutorashanga greenstone belt, which is rich in chromite and gold, making the area ideal for lithological mapping and mineral exploration using remote sensing and GIS tools.

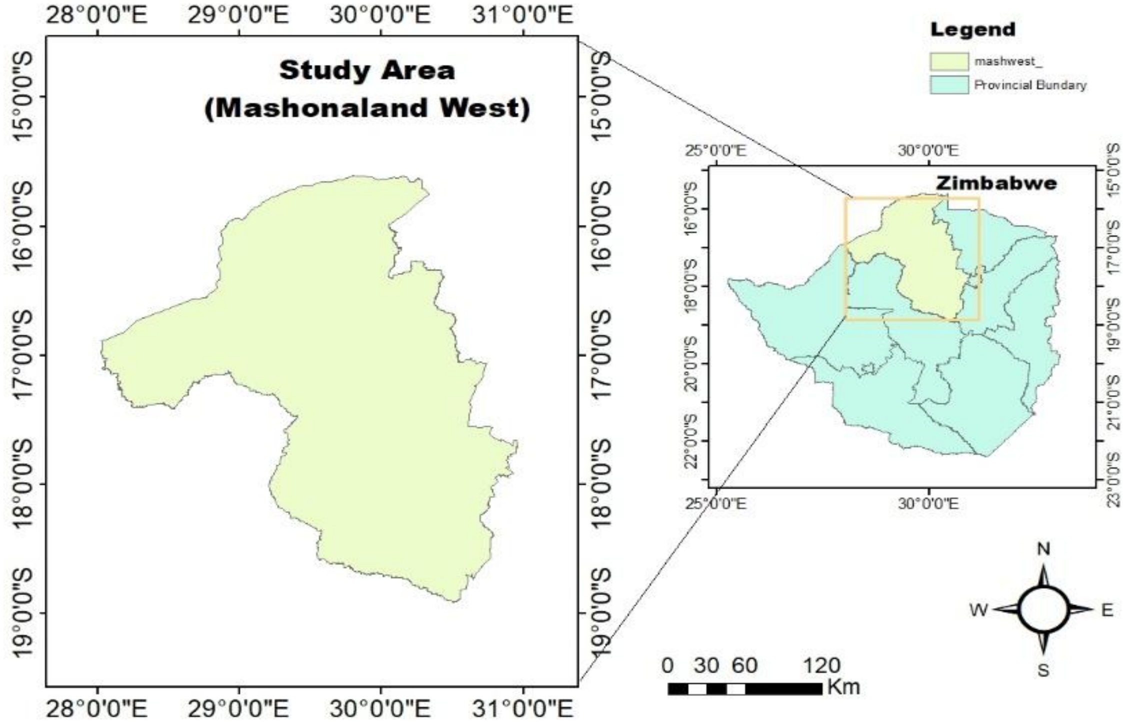


figure 1: was prepared using ArcGIS

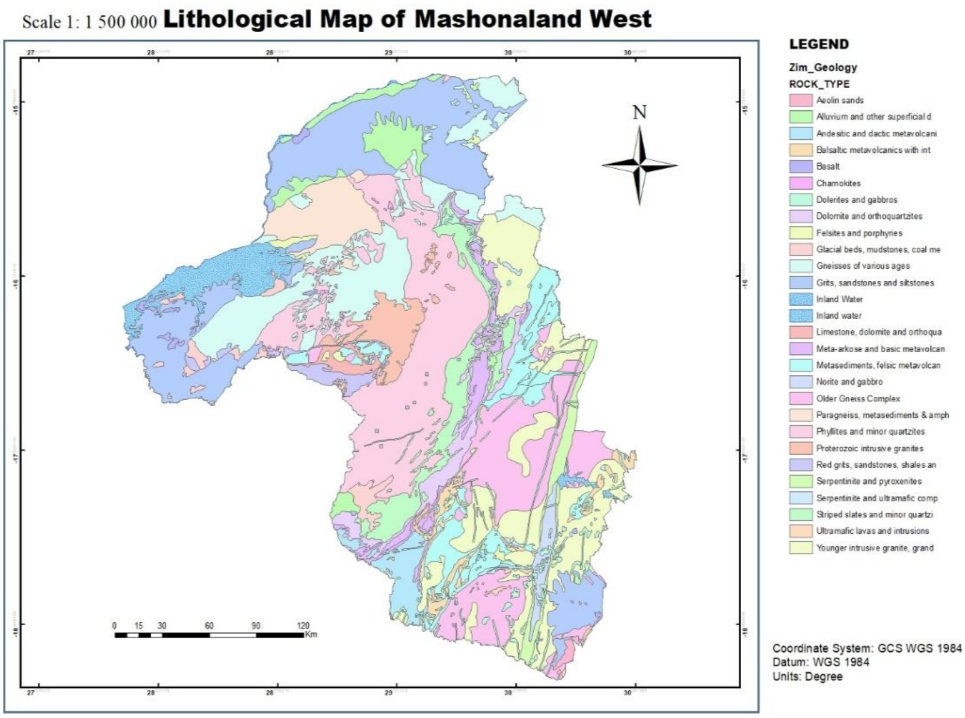


figure 2

2.1.2. Data Acquisition

The study used Sentinel-2 Surface Reflectance (SR) data from the Google Earth Engine data catalog ('COPERNICUS/S2_SR'). To obtain high-quality, cloud-free observation and a cloud cover of less than 10 percent, imagery were filtered to cover a period from January 1, 2023 to December 31, 2023. Six spectral bands were analyzed: Blue (B2), Green (B3), Red (B4), Near Infrared (B8), short wave infrared 1 (B11), short wave infrared 2 (B12).

Training Points: Five distinct lithological classes (granite, sandstone, limestone, shale, and basalt) were interpolated onto training points in ArcGIS using geological information from USGS datasets. They were then imported as a feature collection in Google Earth Engine to train the classification algorithm.

The lithological boundaries and reference data were derived from the USGS geological dataset of Zimbabwe, Geological Dataset. This dataset was processed with ArcGIS to produce a preliminary lithological map of the Mashonaland West region, which was used as a guide to the classification process.

2.1.3. Data Preprocessing

Sentinel-2 Preprocessing

Preprocessing was performed in Google Earth Engine through: filtering images by the date range, location and cloud cover less than 10 per cent, clipping the images to the Mashonaland West boundary, selecting relevant spectral bands B2, B3, B4, B8, B11 and B12, converting pixel values to reflectance by applying the scale factor (× 0.0001) and the production of a cloud free median composite from the whole filtering images. Four spectral indices were used associated with the Sentinel-2 bands to improve the lithological differentiation; Normalised Difference Vegetaion Index (NDVI), normalised Difference Water Index (NDWI), Modified Normalised Difference Water Index (MNDWI), and Normalised Difference Snow Index (NDSI): NDVI=(B8-B4)/(B8+B4); NDWI=(B3-B8)/(B3+B8); MNDWI=(B3-B11)/(B3+B11); NDSI=(B11-B12)/(B11+B12). The lithology shapefile derived from USGS geological dataset was used to generate training points. ArcGIS Geoprocessing tools were used to create random points and the output was constrained to the study area boundary. 600 valid points were generated and ensured a distance at least 200m is not between two points to disperse the clustering. These points were given lithology values which were extracted, and numeric class value (Granite: 0; Sandstone: 1;

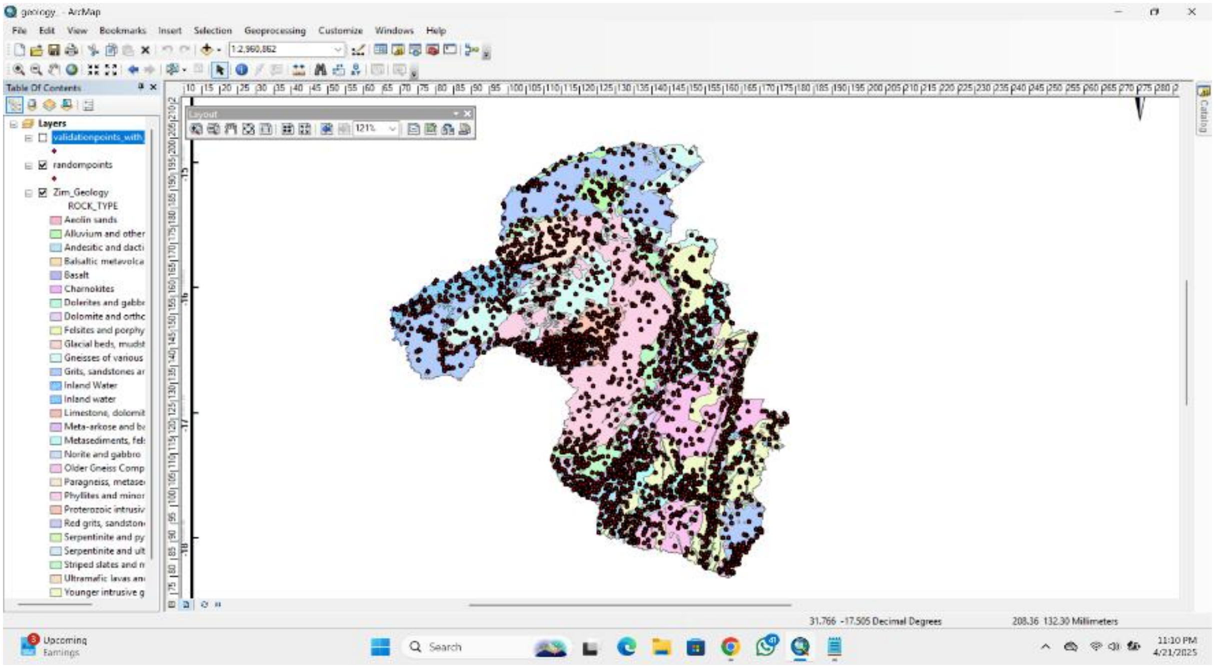
Limestone: 2; Shale: 3; Basalt: 4) were assigned to each so as to make the classification process easier. An exported point dataset with X, Y coordinates and class labels was produced, which was used with Google Earth Engine.

figure 3: Sampling points

2.2. Observations and Outputs

2.2.1. Classification Methodology

A supervised Random Forest classification algorithm was implemented in Google Earth Engine using the Smile



Random Forest implementation with 50 trees. Six spectral bands (B2, B3, B4, B8, B11, and B12), four spectral indices (NDVI, NDWI, MNDWI, and NDSI) were input features of the classifier that was trained. The point samples for training data had known lithological classes, and a sampling scale of 10 meters was chosen to align with Sentinel 2's largest spatial resolution. A lithological map of five classes (granite, sandstone, limestone, shale, and basalt) was produced using Sentinel 2 composite using the classification model. Visualization of each class was given a color (brown, yellow, grey, green, and black, respectively).

2.2.2. Accuracy Assessment

Due to the fact that training points were involved in this evaluation, a confusion matrix was created and classified the accuracy. The accuracy of the classification and class-specific accuracies were done to determine the performance of the classification. Explanation of the matrix is as follows: Class 0 (Granite) has 199 features correctly predicted as granite and 1 wrongly predicted as limestone. Accuracy for Granite being 199/200= 99.5%. In Class 1 (Sandstone), there were 99 features predicted correctly and 1 wrongly predicted as Granite, giving an accuracy of 99/100 = 99.0%. According to the performance measured on the test set of Class 2 (Limestone), 100 were correctly predicted, achieving an accuracy of 100%. In Class 3 (Shale), 100 features were correctly predicted (accuracy 100%). For class 04 (Basalt), we had correctly predicted 98 out of 99 features and 1 wrongly predicted as Granite, making an accuracy of 98.99%. The confusion matrix provides insights into the model's ability to correctly classify each lithological type and identifies potential confusion between classes.

CLASS NAME	CONFUSION MATRIX	AREA(sq km)
Granite(0)	[199,0,1,0,0]	35552.68
Sandstone(1)	[1,99,0,0,0]	7258.775
Limestone(2)	[0,0,100,0,0]	5605.536
Shale(3)	[0,0,0,100,0]	4602.649
Basalt(4)	[1,0,0,0,98]	4644.457

2.2.3. Area Calculation

For each lithological class, binary masks were created for the spatial extent of each class, multiplied by pixel area to give area in square metres, sum presents of area within the provincial boundary and the report of results in square kilometres. The quantitative derivatives of area with regard to different lithological units for the study area were the area calculations provided.

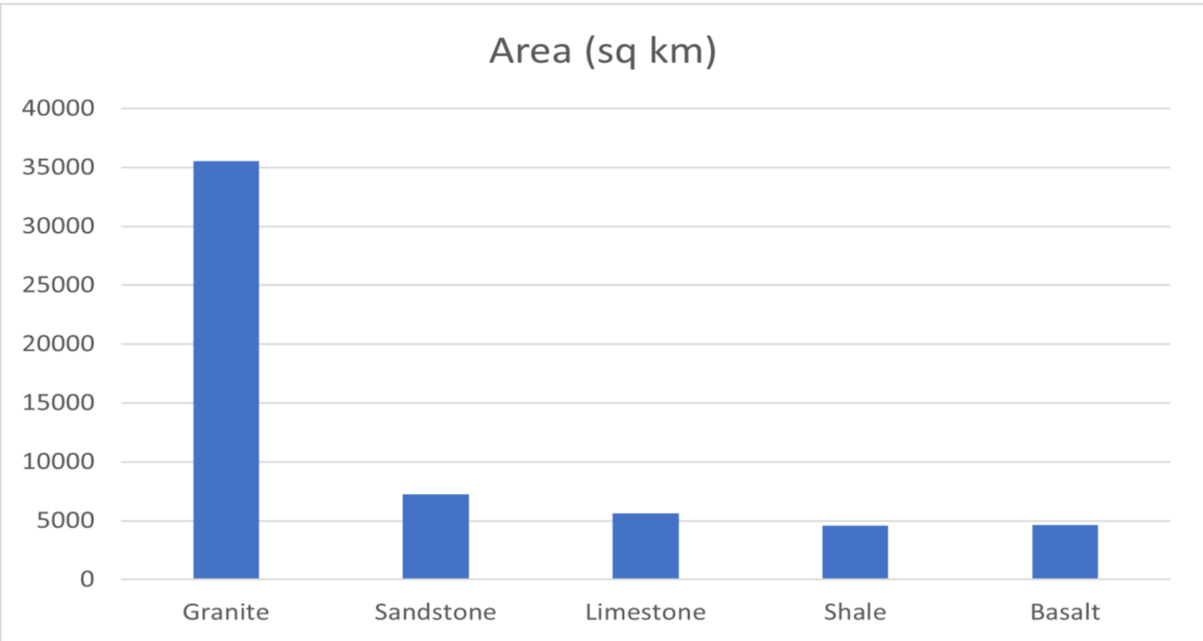


figure 4:

The classification model was applied to the Sentinel-2 composite to produce a lithological map with five classes: Granite, Sandstone, Limestone, Shale, and Basalt. Each class was assigned a distinctive color (brown, yellow, gray, green, and black, respectively) for visualization.

2.2.4. Integration with ArcGIS

The classified image was exported from Google Earth Engine as a GeoTIFF file at 10-meter resolution and imported into ArcGIS for further analysis and map production as shown below.

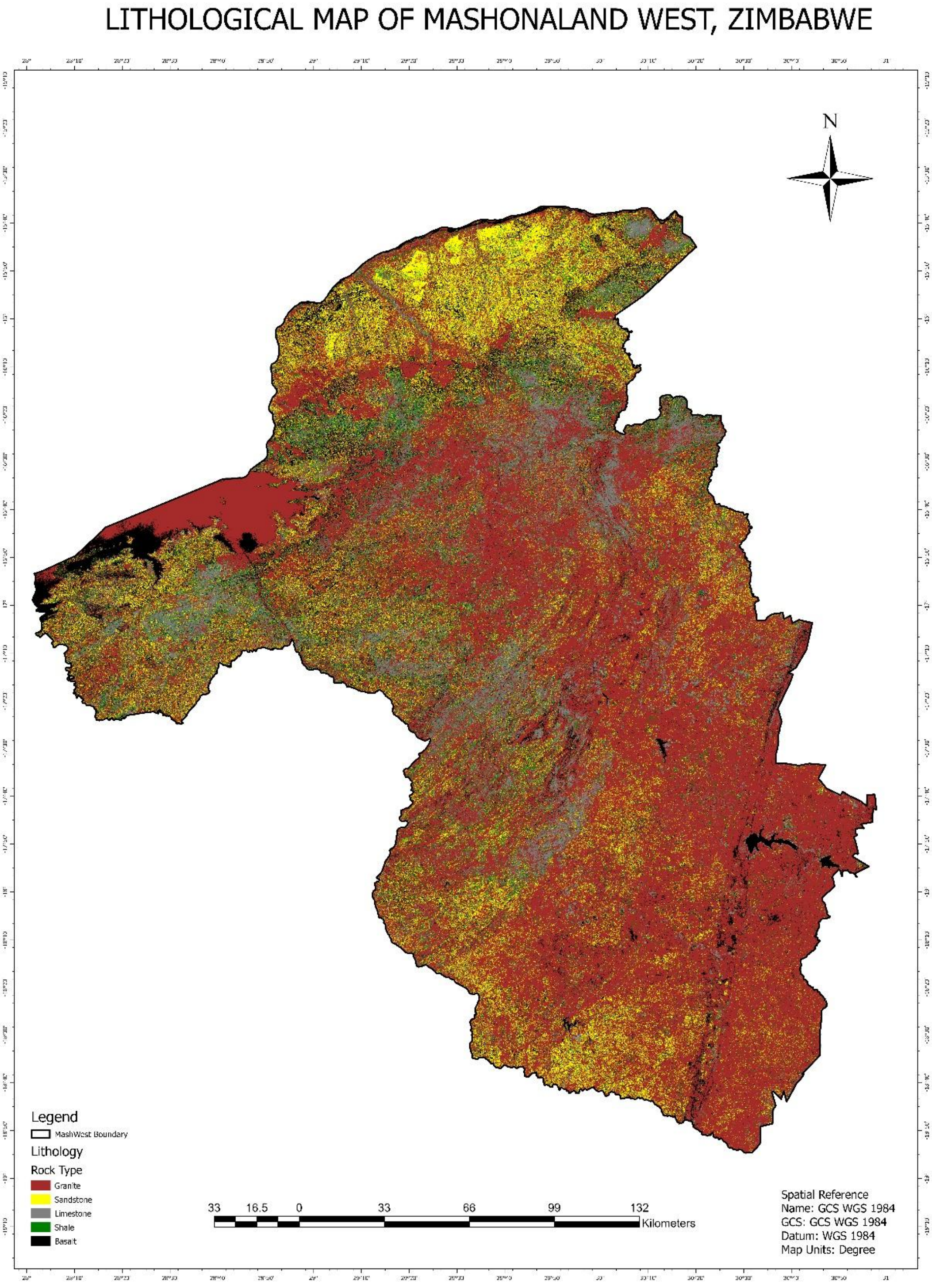


figure 5: lithological map of Mashonaland west

2.2.5. Validation and Refinement

The lithological classification was compared to a USGS geological dataset to determine its accuray and reliability. The limitations of the remote sensing based approach were identified and analysed in areas of discrepancy. The final lithological map integrates, geological information together with satellite based classification to produce an updated lithological distribution map for Mashonaland West Zimbabwe.

2.3. Technical Implementation

Software and platforms that were used were Google Earth Engine, ArcGIS and were utilised as a combination for image processing and classification and study area delineation, and final map production. Computational power for processing large volumes of satellite data was made possible with Google Earth Engine, and ArcGIS had sophisticated cartographic capabilities to improve maps, clean them up and finally present them. Subsequent analysis and interpretation was carried out in ArcGIS resulting in the final lithological map of Mashonaland West Province.

3. Discussion

Combining Sentinel-2 data with machine learning algorithms like Random Forest has made great strides in lithological and mineral mapping in the Mashonaland West area. It significantly contributes to overcoming the constraints of field-based methods, particularly in areas with thick vegetation, by employing remote sensing techniques to derive valuable information on the spatial distribution of lithological units related to economically significant mineral deposits.

The study reported that the identified granite can help achieve a classification with an accuracy of 99.5%, and this can be especially relevant because granite is often indicated as being associated with mineralization, such as gold and copper deposits. Likewise, the correct identification of greenstone belts (often defined by mafic and ultramafic rocks) can also reveal certain resources such as chromite and nickel, both of which are important for industry. Results indicated a high classification accuracy rate across all the lithological classes with sandstone being 99.0%, limestone and shale being 100%, and basalt being 98.99% confirming the applicability of this method in delineating rock types that are indicators of specific mineral resources.

This detailed lithological map is not only conducive to the understanding of geological framework but also acts as an effective tool in mineral exploration for target and resource extraction efforts so that these efforts coordinate with areas of valuable minerals. The identification of lithological units in specific regions through detailed mapping contributes towards more accurate mineral predictions and improved exploration strategies, resulting in reduced environmental degradation.

In addition, the application of Google Earth Engine enables rapid data processing, providing even advanced mapping capabilities for less-privileged institutes. These methods have been acknowledged for their high spatiotemporal resolution and their non-vertical approach to detecting minerals, which makes them an attractive alternative to traditional, invasive mineral exploration tools. Overall, their results will help promote sustainable land management and better decision-making in high mineralization areas while providing a reproducible workflow that can be used within similar geologies worldwide to improve mineral exploration globally.

4. Results

For the five classes analyzed, the lithological classification achieved high accuracies. Finders (Granite 99.5%, sandstone 99.0%, limestone 100%, shale 100% and basalt 98.99%) The confusion matrix showed different results with no misclassifications, highlighting the ability of the Random Forest algorithm to discriminate between lithological types. The lithological map obtained at the end offers not only the geological frame of a region but also is useful for the exploration of the lands in the future. Overall, the results of this study not only contribute to the geological understanding of Mashonaland West, but serve as a guide for similar studies in similarly-polluted regions.

Acknowledgement

University of Zimbabwe deserves our sincere appreciation for their repeated support including necessary resources throughout the duration of our research. Our analysis became possible through the platform provided by the Google Earth Engine team. Our research received critical support from the Geomatics department Lecturer Dr Juliana Useya and our colleagues. The research achieved its successful outcomes due to the joint collaboration of these researchers.

REFERENCES

Bentahar, I., & Raji, M. (2021). Comparison of Landsat OLI, ASTER, and Sentinel 2A data in lithological mapping: A Case study of Rich area (Central High Atlas, Morocco). *Advances in Space Research*, *67*(3). https://doi.org/10.1016/j.asr.2020.10.037

Chen, Y., Dong, Y., Wang, Y., Zhang, F., Liu, G., & Sun, P. (2023). Machine learning algorithms for lithological mapping using Sentinel-2 and SRTM DEM in highly vegetated areas. *Frontiers in Ecology and Evolution*, *11*. https://doi.org/10.3389/fevo.2023.1250971

Chen, Y., Wang, Y., Zhang, F., Dong, Y., Song, Z., & Liu, G. (2023a). *Lithology Mapping in Vegetation-Covered Regions: RS imagery, Method, Challenge and Opportunities*. https://doi.org/10.20944/preprints202307.1936.v1

Chen, Y., Wang, Y., Zhang, F., Dong, Y., Song, Z., & Liu, G. (2023b). Remote Sensing for Lithology Mapping in Vegetation-Covered Regions: Methods, Challenges, and Opportunities. In *Minerals* (Vol. 13, Issue 9). https://doi.org/10.3390/min13091153

EL-Omairi, M. A., & El Garouani, A. (2023a). A review on advancements in lithological mapping utilizing machine learning algorithms and remote sensing data. In *Heliyon* (Vol. 9, Issue 9). https://doi.org/10.1016/j.heliyon.2023.e20168

EL-Omairi, M. A., & El Garouani, A. (2023b). A review on advancements in lithological mapping utilizing machine learning algorithms and remote sensing data. In *Heliyon* (Vol. 9, Issue 9). https://doi.org/10.1016/j.heliyon.2023.e20168

Jelsma, H. A., Nesbitt, R. W., & Fanning, C. M. (2021). Exploring our current understanding of the geological evolution and mineral endowment of the Zimbabwe Craton. *South African Journal of Geology*, *124*(1). https://doi.org/10.25131/SAJG.124.0020

Knepper, D. H. (2012). Mapping hydrothermal alteration with Landsat Thematic Mapper data. In *Remote Sensing in Exploration Geology: Golden, Colorado to Washington, D.C., June 30–July 8, 1989*. https://doi.org/10.1029/ft182p0013

Mudimbu, D., Davies, T. C., Tagwireyi, D., & Meck, M. L. (2022). Application of Geochemical Indices in Evaluating Potentially Harmful Element Contamination at Mining Centres in the Sanyati Catchment, Zimbabwe. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.829900

Serbouti, I., Raji, M., Hakdaoui, M., El Kamel, F., Pradhan, B., Gite, S., Alamri, A., Maulud, K. N. A., & Dikshit, A. (2022). Improved Lithological Map of Large Complex Semi-Arid Regions Using Spectral and Textural Datasets within Google Earth Engine and Fused Machine Learning Multi-Classifiers. *Remote Sensing*, *14*(21). https://doi.org/10.3390/rs14215498

THE PETROLOGY AND GEOCHEMISTRY OF THE LOWER PYROXENITE SUCCESSION OF THE GREAT DYKE IN THE MUTORASHANGA AREA. (1998).

van der Meer, F. D., van der Werff, H. M. A., van Ruitenbeek, F. J. A., Hecker, C. A., Bakker, W. H., Noomen, M. F., van der Meijde, M., Carranza, E. J. M., de Smeth, J. B., & Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. In *International Journal of Applied Earth Observation and Geoinformation* (Vol. 14, Issue 1). https://doi.org/10.1016/j.jag.2011.08.002