Masters Program in Geospatial Technologies

Multispectral Image Analysis of ASTER and Sentinel-2 Satellite Data for Mineral Exploration Using Machine Learning Algorithms – A Case Study from Adobha, Eritrea

Sirak Tesfamariam Asfaha

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

Multispectral Image Analysis of ASTER and Sentinel-2 Satellite Data for Mineral Exploration Using Machine Learning Algorithms – A Case Study from Adobha, Eritrea

Sirak Tesfamariam Asfaha

Master's Thesis in Geospatial Technologies

Supervisor

Dr. Torsten Prinz

Institut für Geoinformatik, Universität Münster

Münster, Germany

Co-supervisors

Prof. Carlos Granell
Universitat Jaume I
Castellón, Spain

Prof. Marco Painho
NOVA IMS
Lisbon, Portugal

February 2025

Declaration

I, Sirak Tesfamariam Asfaha, confirm that this thesis, titled ***Multispectral Image Analysis of ASTER and Sentinel-2 Satellite Data for Mineral Exploration Using Machine Learning Algorithms – A Case Study from Adobha, Eritrea***, is my original work, completed in accordance with the academic regulations of University of Münster. All sources have been properly cited, and no part of this thesis has been submitted for another qualification. Contributions from other peoples are acknowledged, and all research and analysis were conducted by me unless stated otherwise.

To improve clarity and structure, I used a Language Model (LLM) to help with organization and grammar. However, all research, ideas, and content are my own, and I take full responsibility for my work.

Sirak Tesfamariam Asfaha

20th Feb. 2025

Acknowledgment

I extend my heartfelt gratitude to my supervisor, Dr. Torsten Prinz, for his invaluable guidance, support, and encouragement throughout this research. His expertise and insights were crucial in shaping this study. I am deeply grateful to my co-supervisors, Prof. Painho and Prof. Carlos, for their thorough review of my thesis and their constructive feedback, which significantly enhanced the quality of this work.

Special thanks to Mr. Kifletsion Berhe for providing the essential data that made this research possible. His assistance was vital to the analysis and success of this study. I also thank Mr. Tesfahiwet Yemane for his unwavering support and motivation, which have been a source of strength since my undergraduate years.

To my family, I express my deepest appreciation for their constant support, patience, and belief in me. Their faith in my abilities has been a driving force throughout this journey. Finally, I dedicate this work to the memory of my dear friend, Brhane Gebrekidan, who recently passed away. His friendship, kindness, and inspiration will always be cherished and remembered.

20th Feb. 2025

Sirak Tesfamariam Asfaha

Abstract

Mineral Prospect Mapping (MPM) plays a major role in identifying economical minerals deposits, particularly in arid regions with minimal vegetation cover. However, most MPM methods require extensive geochemical parameters, including multi element analyses and concentration measurements, making them challenging to implement in resource-constrained regions. This study proposes an integrated approach combining multispectral remote sensing with machine learning to predict mineral anomaly zones in Adobha, Eritrea, using limited geochemical parameters. The proposed approach uses Sentinel-2 and ASTER satellite imagery to detect alteration minerals through band ratio analysis and RGB color composition. Principal Component Analysis (PCA) reduces the dimensionality of alteration indices derived from merged RGB composites. Random Forest (RF) and Support Vector Machine (SVM) classifiers are trained using geochemical data (Au, Cu, Pb, Zn) combined with the first principal component of each alteration type to predict anomaly zones across the study area.

The oversampled RF model achieved the highest performance with an F1 score of 0.81 and an AUC-PR score of 0.74, outperforming other models. The predicted anomaly zones showed strong spatial correlation with existing traditional (artisanal) mining activities and geophysical survey results (VTEM), particularly along major fault structures. Iron oxide and phyllic alterations emerged as the most significant predictors, consistent with the region's geological setting. This study demonstrates that effective mineral prospecting is possible using limited geochemical parameters when combined with satellite-derived alteration indices and machine learning. The methodology offers a cost-effective approach for initial mineral exploration in understudied regions, providing a practical framework for large-scale prospecting campaigns.

Keywords: Mineral Prospect mapping (MPM); Random Forest; Support Vector machine; Anomaly Detection, Multispectral data

Contents

List of Figures

List of Tables

1. Introduction

The advancement of Earth observation technologies has expanded the role of geological remote sensing in areas like lithological and mineralogical mapping (Bentahar & Raji, 2021; Van der Meer et al., 2014). The availability of non-commercial satellite data makes remote sensing technology as a tool in mineral explorations (Chen et al., 2022). This allows mineral prospect mapping (MPM) to be optimal and cost effective, in a region of arid climate and minimal vegetation cover. Detecting of base and precious metals directly through remote sensing techniques is challenging but various hydrothermal alterations minerals associated with ore deposits can be identified (Testa et al., 2018). Hydrothermal Alterations are formed by the interaction of wall rock and hydrothermal fluid on the process of uprising through structures, such as fault and fractures, caused by pressure difference (Shirmard et al., 2020). The presence of different hydrothermal alteration features on the surface is the key indicators of underground deposit or outcrops (Shirmard et al., 2020), these surface features could be connected to different types of ore deposits, which provides as a starting point towards the MPM (Hunt, 1976). The mineral assemblage of a hydrothermal alteration have a spectral characteristics, in electromagnetic spectra, that can be used for detecting and discriminating between different alteration types using remote sensing (Khaleghi et al., 2020).

ASTER and Sentinel are among the commonly used satellite images in mineral exploration. The key features of both satellite image are summarized on Table 1. Sentinel-2 is a multispectral satellite sensor that offers high-resolution imaging with 13 spectral bands. These bands cover the Visible, Near Infrared (VNIR), and Shortwave Infrared (SWIR) wavelength regions. It provides spatial resolutions of 10meters for bands B2, B3, B4, and B8; 20 meters for bands B5, B6, B7, B8a, B11, and B12; and 60 meters for atmospheric correction bands B1, B9, and B10. It has a swath width of 290 km and covers latitudes between -56° and +84°. At the equator, Sentinel-2

completes a minimum revisit time of 10 days (Van der Werff & Van der Mer, 2015) . For environmental monitoring the 10m and 20m bands of sentinels-2 are used whereas the 60m band are mainly applied to cloud detection and atmospheric correction (Louis, 2016).

According to Mielke et al. (2014), their research demonstrated that Sentinel-2A MSI effectively detects gossan. Additionally, Van der Meer et al. (2014), highlighted the capability of Sentinel-2A MSI for mapping iron-oxide mineralogy, showcasing its performance in geological remote sensing. Iron minerals such as Hematite, goethite and jarosite show a descriptive spectral feature near 0.43um, 0.6um, 0.85um and 0.93um, which are nearest to band1, band2, band8/8A and band 9 of sentinel-2, due to iron-bearing constituent mineral's electronic transitions (Van der Meer et al., 2014).

ASTER's 14 spectral bands in the VNIR, SWIR, and TIR regions enable the detection of rock-forming and alteration minerals, such as hydrous and hydroxyl-bearing sulfates, carbonates, and silicates (Hajaj et al., 2024; Khaleghi et al., 2020). ASTER's spectral bands cover a range from 0.52 to 2.43 μm, within which carbonate, hydroxyl-bearing, and iron oxide/hydroxide minerals can be identified using the VNIR and SWIR bands. The bands in the VNIR region shows absorption signatures linked to specific metals and mineral groups, such as Fe^{3+}, Fe^{2+}, and hydroxyl-bearing groups in the SWIR region. Particularly, the SWIR bands have a unique potential to distinguish mineral types like carbonates, clays, phyllosilicates, and sulfates (Testa et al., 2018).

Therefore, ASTER enables the characterization of key minerals that indicate specific alteration types, making it a powerful tool for mapping and identifying hydrothermal alteration zones (Pour and Hashim, 2012; Testa et al., 2018).

Table 1. Comparison of ASTER and Sentinel-2 satellite sensor Features (Source: Duda et al., 2015; Louis, 2016)

Feature	ASTER	Sentinel-2
Launched Date	December 1999	June 2015
Sensor Type	Multispectral and Thermal infrared	Multispectral
Spectral Bands	14 bands - VNIR: 3 bands - SWIR: 6 bands - TIR: 5 bands	13 bands - 4 visible and NIR bands - 6 red-edge and SWIR bands - 3 atmospheric correction bands
Spectral Resolution	VNIR: 15m SWIR: 30m TIR: 90m	10m (4 bands) 20m (6 bands) 60m (3 bands)
Spectral Range	VNIR: 0.52 - 0.86 µm SWIR: 1.60 - 2.43 µm TIR: 8.125 - 11.65 µm	0.443 - 2.190 µm
Temporal Resolution	16 days	5 days (with both Sentinel-2 and 2B)
Radiometric Resolution	8 bit (TIR), 10 bit (VNIR)	12 bit
Orbit Altitude	705 km	786 km
Swath Width	60 km	290 km

1.1 Mineral Prospect Mapping

Mineral prospect mapping (MPM) is the process of identifying potential areas for exploring undiscovered economically valuable mineral deposits (Josso et al., 2023; Lachaud et al., 2023). This process involves grouping each pixel (grid cell) of the study area as either prospective or non-prospective. MPM is a complex task in mineral exploration due to the sophisticated nature of mineral formation and the dynamic geological processes involved (Zerai et al., 2023).

Advancements in machine learning, integrated with geochemical and structural analyses, have shown promising potential in mapping and modeling non-linear relationships between mineral sites and their corresponding evidential features (Lachaud et al., 2023; Zerai et al., 2023). MPM approaches can be broadly categorized into two types: knowledge-driven and data-driven models. The knowledge-driven approach relies on experts' understanding of ore deposits to define model parameters, such as assigning weights and relationships between geological, geophysical, and

geochemical data layers. These weights provide subjective understanding into how spatial characteristics are associated with mineral deposits. In contrast, the data-driven approach primarily relies on analyzing and interpreting data to identify patterns and build predictive models. Unlike the knowledge-driven approach, which depends on expert judgment and prior knowledge, data-driven methods focus on uncovering relationships directly from the data (Josso et al., 2023; Lachaud et al., 2023).

Recently, data-driven methods, Random Forest and Support Vector Machine, have become powerful techniques in different geological applications, showing considerable potential in MPM (Lachaud et al., 2023).

A random forest (RF) is an ensemble learning technique that generates multiple decision trees during training. Boosting and bagging are commonly used ensemble learning methods for developing regression or classification trees. Around 2/3 of the dataset, known as in-bag samples, are used for training, while the remaining 1/3, referred to as out-of-bag (OOB) samples, server as a validation set for assessing model performance as shown in Figure 1. The OOB error is employed to estimate prediction errors and assess variable importance measures. To generate the forest, two parameters must be specified: the number of trees (Ntree) and the maximum number of variables considered at each split (Mtry). According to Belgiu & Drăguţ. (2016), empirical research showed that classification accuracy is affected by changes to Mtry than by changes to Ntree parameters. These characteristics make the RF algorithm computationally efficient and robust against overfitting.

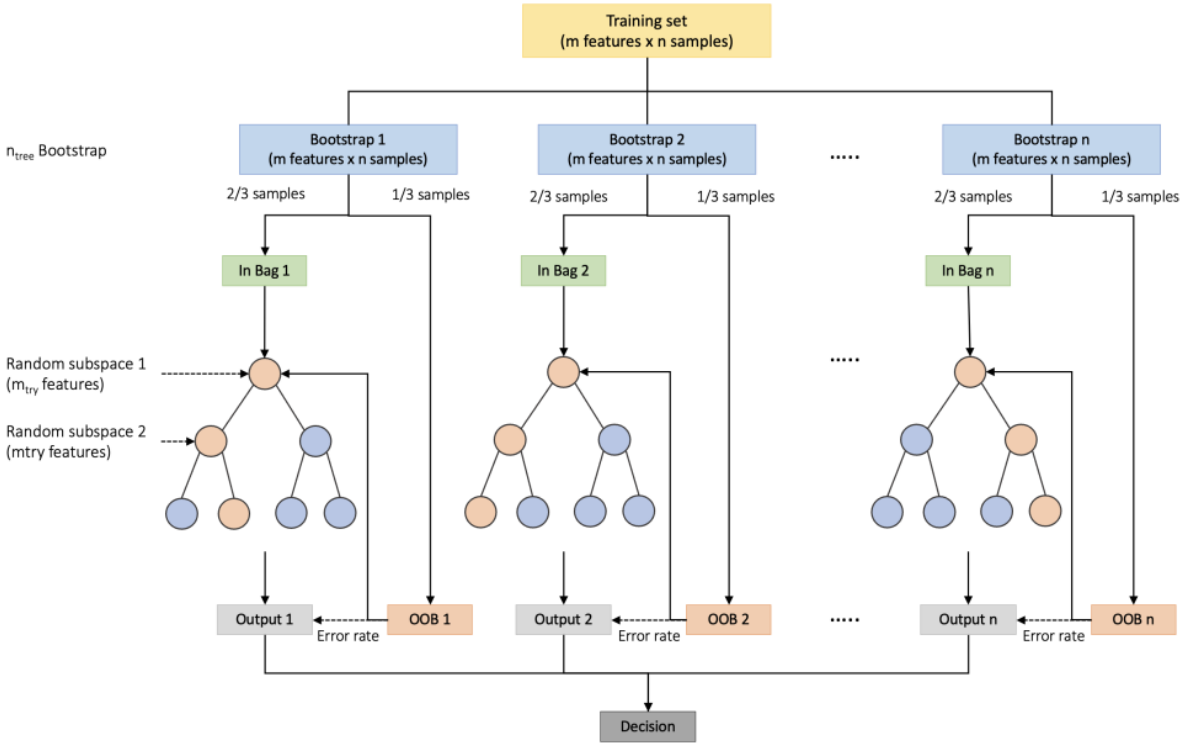


Figure 1. Random Forest Training Process Using Bootstrap Aggregation and Out-of-Bag (OOB) Error Estimation. (source: Lachaud et al., 2023)

Support Vector Machine (SVM) is a non-parametric supervised learning algorithm used for classification and regression tasks. The core principle of the SVM algorithm is to iteratively find an optimal hyperplane that separates data points of different classes into their respective categories using training data. The optimal hyperplane, also known as the maximum margin, represents the decision boundary that minimizes misclassifications during the training phase, as shown in Figure 2. The selection of a kernel function affects SVM performance. Radial Basis Function (RBF) and polynomial kernels are among the commonly used kernels in remote sensing applications (Mountrakis et al., 2011). The RBF kernel is effective in capturing complex, non-linear relationships within unevenly distributed data, while the polynomial kernel is suitable for tasks where the relationship between features is less complex (Rochim et al., n.d.). SVM's performance also depends on proper parameter tuning, particularly the penalty parameter (C) and kernel parameters such as gamma for the RBF kernel. Improper tuning can lead to either overfitting or

underfitting, affecting the model's generalizability. Moreover, SVM can be computationally intensive for large datasets, as it requires solving a convex quadratic optimization problem. However, the advancement in computational techniques such as cloud computing has reduced this limitation. One notable advantage of SVM is its ability to efficiently handle limited training samples.

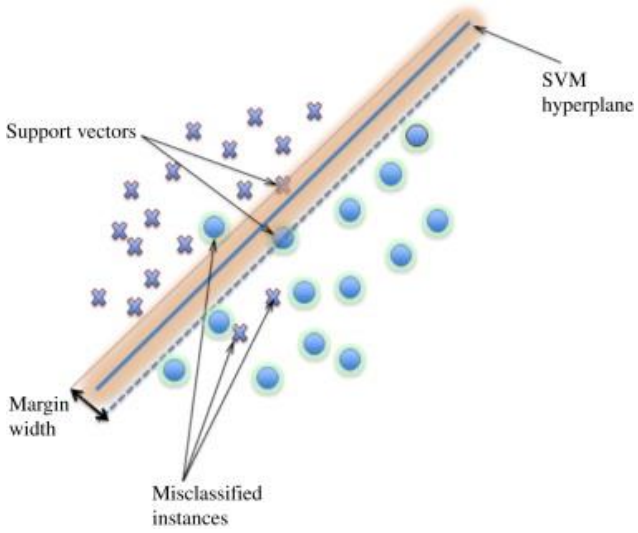


Figure 2. Support Vector Machine (SVM) Decision Boundary with Hyperplane, Margin, Support Vectors, and Misclassified Instances. (source: Mountrakis et al., 2011)

1.2 Research Gap

Many existing studies integrate satellite imagery with machine learning for mineral exploration. To achieve good results, they often rely on training datasets that contain extensive geochemical parameters. These extensive datasets typically include a wide range of elements, isotopes, and mineralogical data. However, acquiring such comprehensive geochemical data can be challenging, and costly, especially in resource-constrained regions. Factors such as limited access to laboratory facilities make it difficult to analyze for all geochemical features.

This work addresses research gap by using limited geochemical parameters to predict mineral anomalies using machine learning techniques (RF and SVM). This approach is valuable in regions

where detailed geochemical data is unavailable or difficult to obtain, as it reduces dependency on extensive datasets while still achieving reliable results.

1.3 Research Questions, Aims and Objectives

This research aims to establish a methodology that integrates image enhancement techniques, including band ratio (BR) and Principal Component Analysis (PCA), with machine learning algorithms such as RF and SVM to distinguish between anomalous and non-anomalous areas. The study investigates how ASTER, Sentinel-2, and machine learning can improve traditional mineral prospecting mapping, notably in capturing ore-forming complexities, and how this approach can enhance the coverage of large-scale mineral exploration in understudied regions. The primary objectives include identifying and mapping hydrothermal alteration zones using image enhancement techniques from ASTER and Sentinel-2 data, training and evaluating RF and SVM using limited geochemical parameters combined with satellite-derived alteration indices for predicting mineral anomalies, validating the predictive models against ground truth data to assess accuracy in delineating anomaly zones, and evaluating the performance of RF and SVM classifiers in handling imbalanced datasets and predicting anomalies using imbalanced data. Through these objectives, the study seeks to advance mineral exploration methodologies in areas with limited laboratory resources for detailed geochemical analysis.

2. Literature review

In remote sensing and geological investigations, Image enhancement techniques enable identification and mapping of surface features. Principal Component Analysis (PCA) is widely used for dimensionality reduction and enhancing spectral contrast in multispectral data. For example, it used to identify hydrothermal alteration zones by isolating features while suppressing noise (Crowley et al., 1989; Tangestani et al., 2005). However, it has a major drawback, the

resulting color composites can be challenging to interpret, and valuable information may be lost in the unused components (Kwarteng and Chavez, 1989a). Selective PCA (SPCA) further enhances this approach by focusing on specific spectral bands, thus highlighting unique geological features like argillic and phyllic zones (Chavez and Kwarteng, 1989; Abrams et al., 1983; Tangestani et al., 2005).

Hajaj et al. (2024) demonstrated the application of ASTER data for detecting alteration minerals and silicification in the Kerdous inlier, Morocco. Using Band Ratio (BR), SPCA, and Constrained Energy Minimization (CEM), the study effectively mapped hydrothermal alteration zones, identifying minerals such as kaolinite, illite, and hematite. The integration of alteration minerals with silicification through a weighted overlay approach provided an understanding for mineral prospectivity mapping.

Zerai et al. (2023) focused on the integrating of ASTER data with soil geochemical analysis to map hydrothermal alteration zones in southwestern Eritrea. Employing One-Class Support Vector Machines (OCSVM) and Principal Component Analysis (PCA), the study identified anomalies associated with Au–Cu–Zn mineralization, showing the potential of machine learning techniques in delineating prospective zones.

Similarly, Abdelkareem et al. (2024) utilized Sentinel-2 and ASTER datasets to explore hydrothermal minerals in the Arabian-Nubian Shield. By applying band ratio techniques and GIS-based overlay analyses, the study highlighted the spatial distribution of high-grade hydrothermal zones. The results underscored the value of integrating remote sensing and field validation for efficient mineral exploration.

Machine learning approaches such as RF and SVM are applied in mineral prospectivity mapping and anomaly detection. RF is known for its robustness, ability to handle high-dimensional data, and insensitivity to overfitting, making it suitable for spatial predictive tasks such as mapping Fe-Mn crusts and epithermal gold deposits (Belgiu & Drăguţ, 2016; Josso et al., 2023; Lachaud et al., 2023). while, SVM excels in cases with limited training data and high-dimensional inputs due to its non-parametric nature and efficient kernel-based approach, as demonstrated in mineral exploration studies including those on skarn and porphyry-epithermal deposits (Mountrakis et al., 2011, Gao et al., 2016; Xiang et al., 2020).

Several studies have highlighted the effectiveness of these algorithms in mineral prediction. For example, Josso et al. (2023) applied RF to predict Fe-Mn crust occurrences in the world's oceans, achieving a prediction accuracy of 87.2% for crust locations and 98.2% for non-crust locations, validated through out-of-bag errors and confusion matrices. Similarly, Xiang et al. (2020) used RF for mapping mineral prospectivity in the Yangtze River Valley metallogenic belt, showing its ability to handle multivariate datasets and spatial complexity. On the other hand, SVM used for mineral prospectivity mapping in limited data scenarios, such as in the identification of epithermal Au deposits where it demonstrated high classification accuracy with different kernel functions (Lachaud et al., 2023).

2.1 Location

The study area lies within the Northern Red Sea Administration Zone in the northern part of Eritrea, as shown in Figure 3, covering a total area of 2,100 km². Access within the area is possible via bush tracks, watercourses, and cross-country routes. However, some areas in the central part of the study region are challenging to reach. Rainfall occurs between mid-July and September, and the climate is categorized as arid to semi-arid. At an altitude of approximately 1,000 meters, the

mean temperature is around 28°C. Vegetation is sparse, primarily concentrated in the larger valleys.

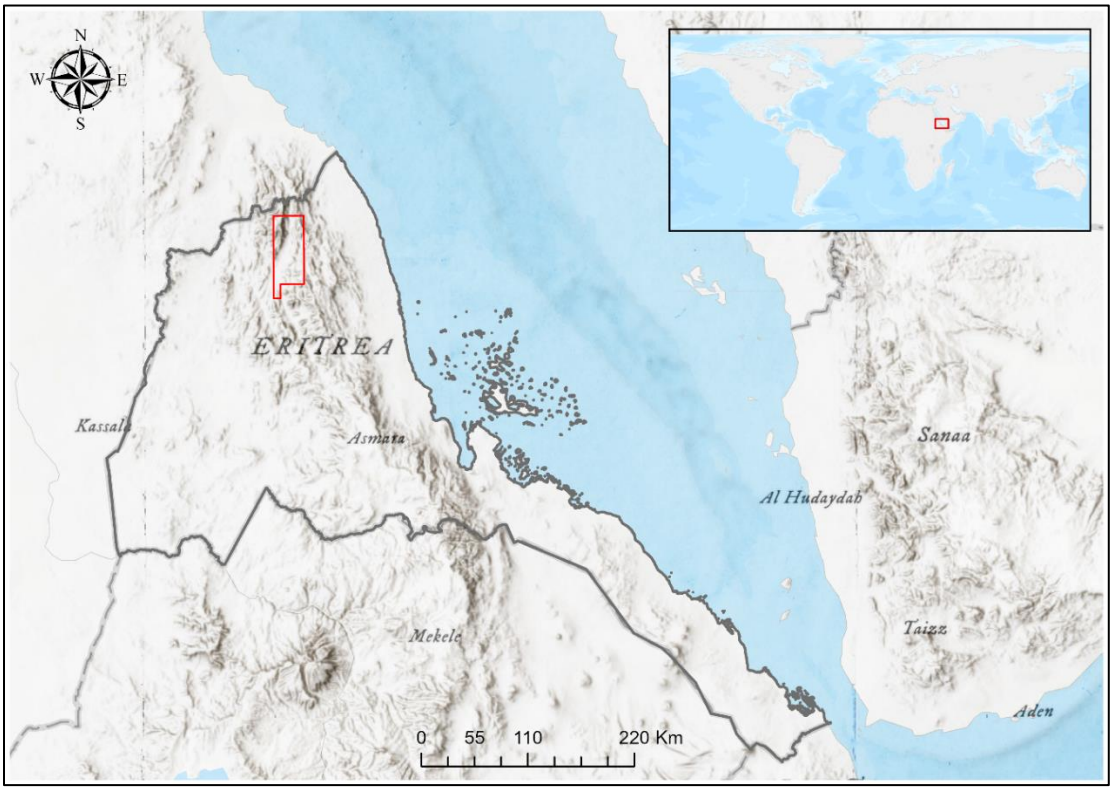


Figure 3. Map showing the location of the study area, shown in red, in the northern region of Eritrea (depicted in gray). The map highlights the country's topography and its geographical context, with neighboring regions labeled. An inset map in the upper right corner provides a global reference, indicating the study area's position within the broader international landscape.

2.2 Regional Geology

Eritrea is part of the southern Arabian-Nubian Shield (ANS), resides in the northern part of Neoprotrozic East African Orogen (Drury & De Souza Filho, 1998). The ANS extends across Egypt, Sudan, Eritrea, Ethiopia, Israel, Jordan, Sudia Arabia and Yemen along both sides of the Red Sea (Barrie et al., 2007). Geologically, the ANS is made of granitoid-greenstone belt terranes and mid-crustal gneissic terranes and it is unconformably covered by a Mesozoic to Cenozoic rocks (Drury & Berhe, 1993; Johnson et al., 2011; Teklay, 1997). The shield formed when the East and west Goundwana collided during the closure of the Mozambique Ocean during Neoproterozoic Pan-African orogenic cycle (stern, 2008, Zhao, et al., 2019).

Regionally, the geology of the area is divided into four terranes trending N-S direction from west to east, Figure 4a. Based on their stratigraphic units and structural characteristics they are subdivided as follows: the Barka terrane, is characterized by its high grade metamorphic rocks ranging from amphibolite to granulite facies, and also include metasediments like Fe-rich pelites, carbonates, rare quartzites, and undeformed felsic dykes, the Hagar terrane predominantly mafic and felsic metavolcanic rocks pre-tectonic intrusive and different schists, the Adobha Abi terrane is a sedimentary basin, composed of highly deformed metasedimentary rocks such as metapelites (mud-rich rocks), impure carbonates, and distinctive pink carbonates, along with layers of deformed sandstone and conglomerate. The Nakfa terrane is the largest terrane of four and contain granite-greenstone belts and syn- to post-tectonic granitoid rocks (Andersson et al., 2006; S. M. Berhe, 1990; Drury & De Souza Filho, 1998; Teklay, 1997).

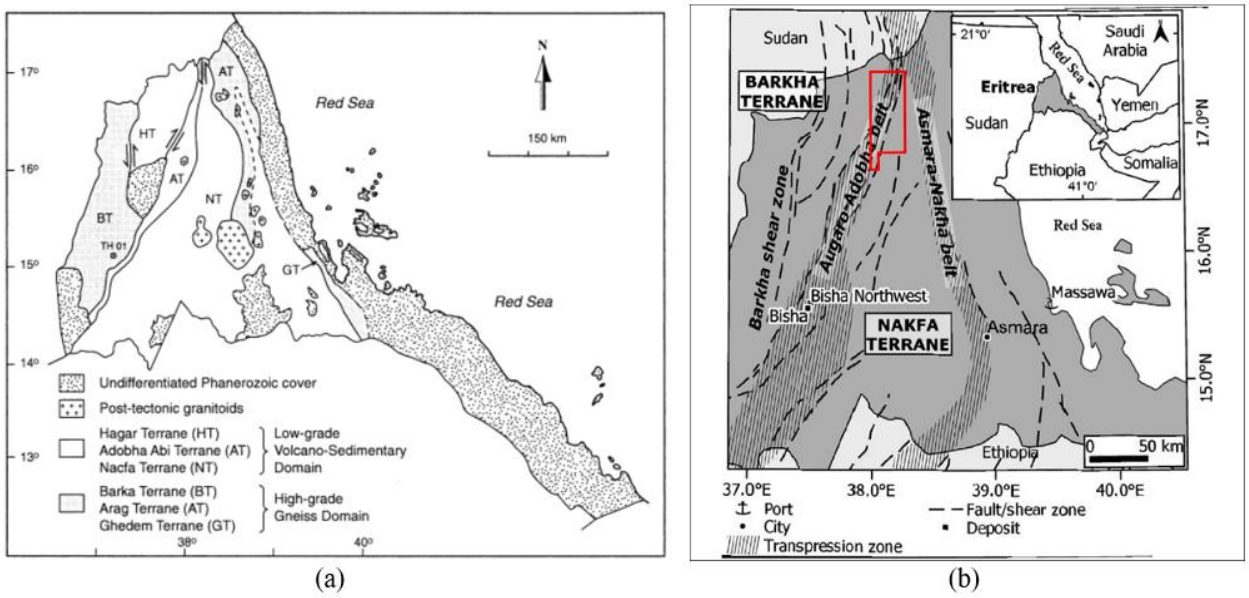


Figure 4. Regional terranes (a) highlighting the geology of the country and tectonic structures (b) emphasizing the regional fault zones, shear zones, and the location of the study area (red boundary). This figure provides an integrated view of the geological and structural framework (modified from Hood et al., 2019; Andersson et al., 2006)

The project area's geology dates back to the Upper Proterozoic era, characterized by low-grade metamorphism, ranging from the pumpellyite-actinolite facies to the greenschist facies

(Woldehaimanot & Behrmann, 1995). This area is predominantly associated with metavolcanic and metasedimentary rocks.

On the eastern side, a coarse-fragmental (clastic) unit dominates, comprising metasedimentary layers of meta-greywacke, metasandstone, metaconglomerate, metapelite, felsic rocks, and fragmental tuffs. In the north-central and eastern parts, a prominent phyllite assemblage underlies a diverse sequence of metasedimentary rocks, including pelitic shale, graphitic shale, siltstone, and fragmental tuff. The central region mainly consists of fine-grained quartz-chlorite schist, calc-schist, quartz-feldspar-sericite schist, phyllite and/or slate, dolomite, graphitic shale (black shale), and metavolcanic rocks, including mafic and felsic tuffs, the geological map of the study area is presented in Figure 5. In the northwest corner, there is a thick sequence of metabasalt, characterized by massive, dark green, fine- to medium-grained rocks, with minor occurrences of porphyritic and amygdaloidal metabasalts. A large exposure of foliated granite, known as Hasta Granite, outcrops in the southeastern part of the study area, distinguished by coarse grains of quartz, plagioclase, K-feldspar, and biotite.

The area is structurally controlled by two prominent NNE-trending faults, known as the Badin and Ela Babu faults, which define a graben called the Adobha Abiy terrane (Drury & Berhe, 1991). The graben is approximately 15 km wide in the south and narrows to 1.5 km in the north. Coarse-grained sandstones and conglomerates fill the graben in the south, transitioning northward into finer, thinly bedded siltstone and shale. The Ela Babu fault bifurcates in the central part of the study area, and at its northern tip, it marks a contact zone between basalt and shale. This contact hosts gold (Au) geochemical anomalies identified by (Chalice Gold Mines, 2011) and extends into the northwest corner of the Gerasi South Exploration License area. Sericite and clay alteration are observed in most granitic rocks, appearing as irregular to structurally controlled zones. Quartz

veining is well developed throughout the granitic rocks, though no systematic orientation was noted. Many quartz veins contain voids, suggesting the removal of carbonate and sulfide minerals due to weathering.

In the southern part of the study area, near the Elababu fault zone, promising shear-hosted gold mineralization. This gold mineralization occurs within quartz–sulfide veins and is associated with sericite-silica-pyrite wall rock alteration along the vein margins. The width of these alteration zones is broadly positively correlated with the size of the quartz veins. The gold is concentrated in sheared quartz veins, where sulfides are preferentially located in the selvage of the quartz veins and veinlets.

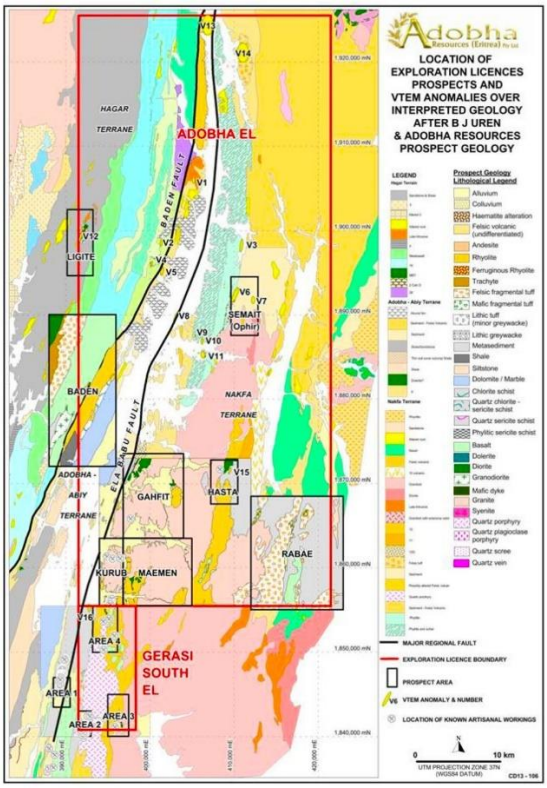


Figure 5. Geological map showing exploration licenses, prospects, VTEM anomalies, traditional mining activities, and faults in the Adobha Exploration License (EL) areas (source: kifletsion Berhe & Chisholm, 2013).

3. Materials and Methodology

3.1 Data Acquisition

The main datasets, ASTER and Sentinel-2 satellite images were used for the MPM. Before proceeding to analysis, both satellite images were reprojected to UTM Zone 37 North with the WGS-84 datum. ASTER optical and thermal sensor imagery, captured by the TERRA satellite, was obtained from the Earth Engine Data Catalog using Google Earth Engine. The downloaded images were in Level-1T, radiometrically calibrated and geometrically corrected pre-processed data (Duda et al., 2015). Six spatially overlapping scenes, obtained between January 01 to January 30, 2007, which is a dry season, with cloud cover less than 5%. The overlapping images were then combined into a single image by mosaic compose, a google earth engine function used to produce a spatially continuous image (Gorelick et al., 2017). The mosaicked satellite image was then clipped according to the study area. However, the ASTER image does not fully cover the whole study area as in the Figure 6. ASTER spectral bands have different resolution, to make the all bands consistent in their resolution, the VNIR band were resampled to SWIR band spectral resolution which is 30 meters.

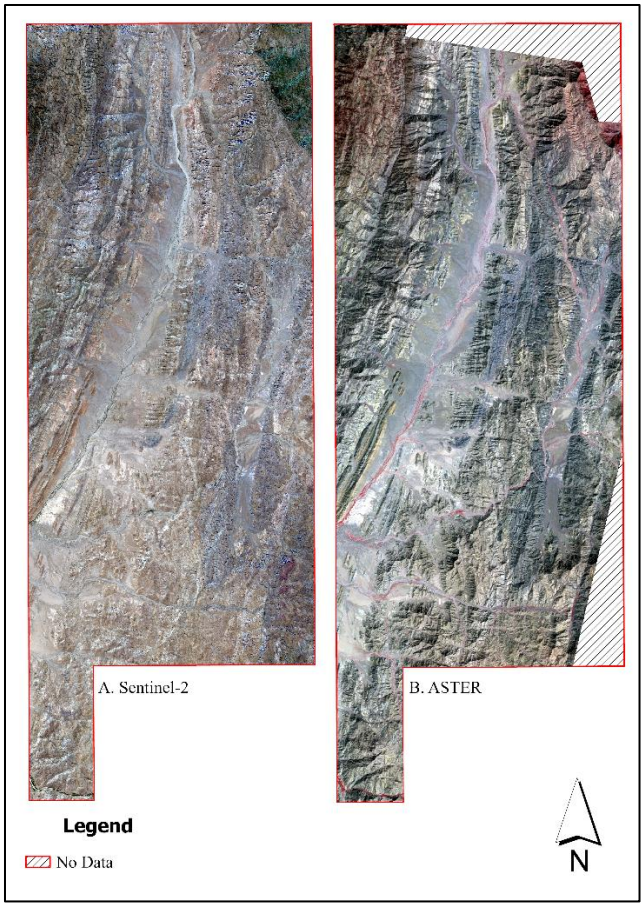


Figure 6. Both satellite images are displayed in True color composition, highlighting the extent of coverage within the study area. Sentinel-2 and ASTER are represented using B4,B2,B3 and B3N,B2,B1 respectively.

Sentinel-2 Level-2A is a high-resolution satellite dataset designed to provide atmospherically corrected surface reflectance values, making it highly suitable for detailed land surface analysis (Louis, 2016). For this study, the images were gathered from December 1 to December 30, 2023 with a cloud cover threshold of less than 5%.

A total of 14 overlapping Sentinel-2 scenes were selected to cover the study area. These scenes were processed using the mosaic composition method, a technique that merges individual images to eliminate overlaps. This approach not only ensures spatial continuity but also preserves the integrity of spectral information across the entire dataset.

The combination of ASTER and Sentinel-2 data offers a synergistic approach to mineral exploration, utilizing their complementary strengths. ASTER's SWIR and TIR bands excel in identifying hydrothermal alteration minerals like clays, carbonates, and sulfates (Shirmard et al., 2020), while Sentinel-2's high-resolution VNIR and SWIR bands are used for detecting iron oxides, hydroxides, and gossan formations (Werff & Meer, 2015). By combining ASTER's spectral precision with Sentinel-2's spatial detail, studies can obtain a more in-depth and accurate mapping of hydrothermal alteration zones. This combined approach is advantageous in arid regions with sparse vegetation, where surface features are more exposed, improve the efficiency of mineral prospect mapping.

The resulting mosaicked image was clipped to match the extent of the ASTER scene previously downloaded. This step was crucial for reducing unnecessary data, which weren't covered by ASTER and ensuring that all subsequent analyses were focused within the defined study boundaries. To maintain consistency in spatial resolution across datasets, the 10-meter VNIR (Visible and Near Infrared) bands from Sentinel-2 were resampled to a 15-meter resolution to match the SWIR (Short-Wave Infrared) bands. This step makes uniform pixel dimensions, enabling accurate comparative and combined analysis between VNIR and SWIR data.

The geochemistry data were collected by Adobha Resources (Eritrea) Pty Ltd in 2011. Systematic soil and rock chip samples, with a total of 1,789, were collected covering the area of interest in Figure 7. The rock chip samples were crushed and pulverized at the Adi-guidad laboratory, while the soil samples were shipped to and analyzed at the Genalysis Australia laboratory. The missing values and values below the detection limit were removed from the original datasets. The resulting dataset includes GPS locations, Au, Cu, Zn and Pb concentrations as well as rock descriptions.

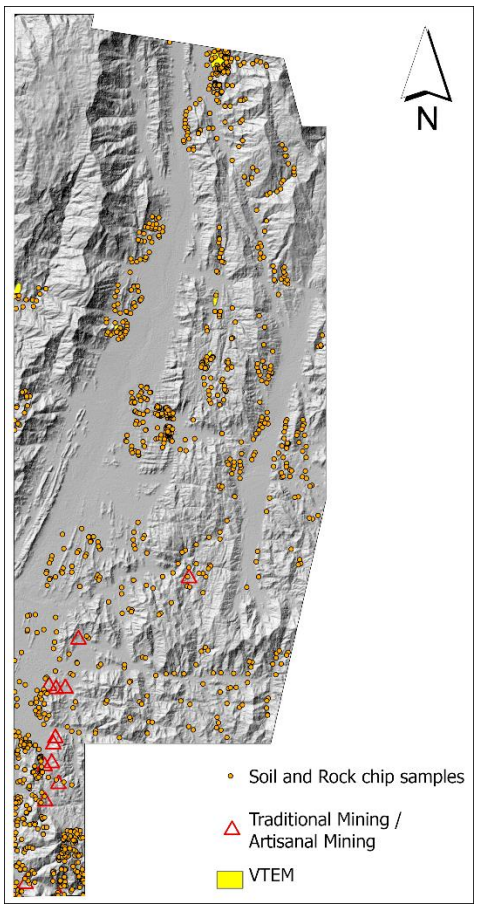


Figure 7. Locations of 1789 soil and rock chip samples, known traditional mining and VTEM survey

3.2 Methodology

Figure 8 shows the methodology used to predict areas of anomalies in Au, Cu, Zn, and Pb within the study area. The methodology is divided into three sections: data collection and preprocessing, pixel enhancement, and anomaly prediction. It begins with the collection and preprocessing of Sentinel-2, ASTER, and geochemical data using Google Earth Engine and Python. Initial analysis of raw data revealed notable noise and topographic effects that obscured mineralogical signatures, making preprocessing vital to enhance data quality and ensure compatibility between difference data source. During the pixel enhancement stage, techniques such as band ratios, RGB band composition, and Principal Component Analysis (PCA) were employed. These methods were prioritized because unprocessed spectral bands lacked sufficient contrast to clearly identify

alteration zones associated with mineralization. Band ratios were used to minimize topographic effects and enhance specific mineral signatures, while RGB band compositions helped visualize spatial distribution patterns of alteration minerals. PCA was applied to reduce data dimensionality and eliminate redundant information, as preliminary analysis showed that the original dataset contained highly correlated bands, which could lead to overfitting in predictive models.

The processed data was then used to train two predictive models: Random Forest (RF) and Support Vector Machine (SVM). Both models received PCA-transformed bands and geochemical data as input. Preliminary tests with simpler models (e.g., logistic regression) showed poor performance in handling the high-dimensional and imbalanced dataset, and large portion of research in MPM literature centers to RF and SVM (Sheykhmousa et al., 2020), which led to the selection of RF and SVM. These models were chosen because they are robust to high-dimensional data and proven in similar remote sensing applications (Belgiu & Drăguţ, 2016; Mountrakis et al., 2011). To optimize model performance, K-fold cross-validation and grid search were applied. K-fold cross-validation was used to split the dataset into multiple folds, ensuring that the model generalizes well and is not biased toward a particular class. Grid search was employed to systematically identify the optimal hyperparameters for both models, enhancing classification accuracy. These optimization techniques were necessary to ensure that the models were not overfitting or underfitting the data, which was a concern during preliminary tests.

The RF model produced a predicted map, which was then compared with the SVM-predicted results. Finally, the predicted maps were validated against an existing geological map, including locations of traditional mining activities and VETM (Versatile Time Domain Electromagnetic, a geophysical survey) anomalies. This validation step was critical because preliminary results indicated that without proper validation, the predicted anomaly zones could not be confidently

linked to real-world mineralization. Comparing the model outputs with known mining sites and VETM anomalies confirmed that the models were able to identify areas of interest that aligned with existing geological knowledge.

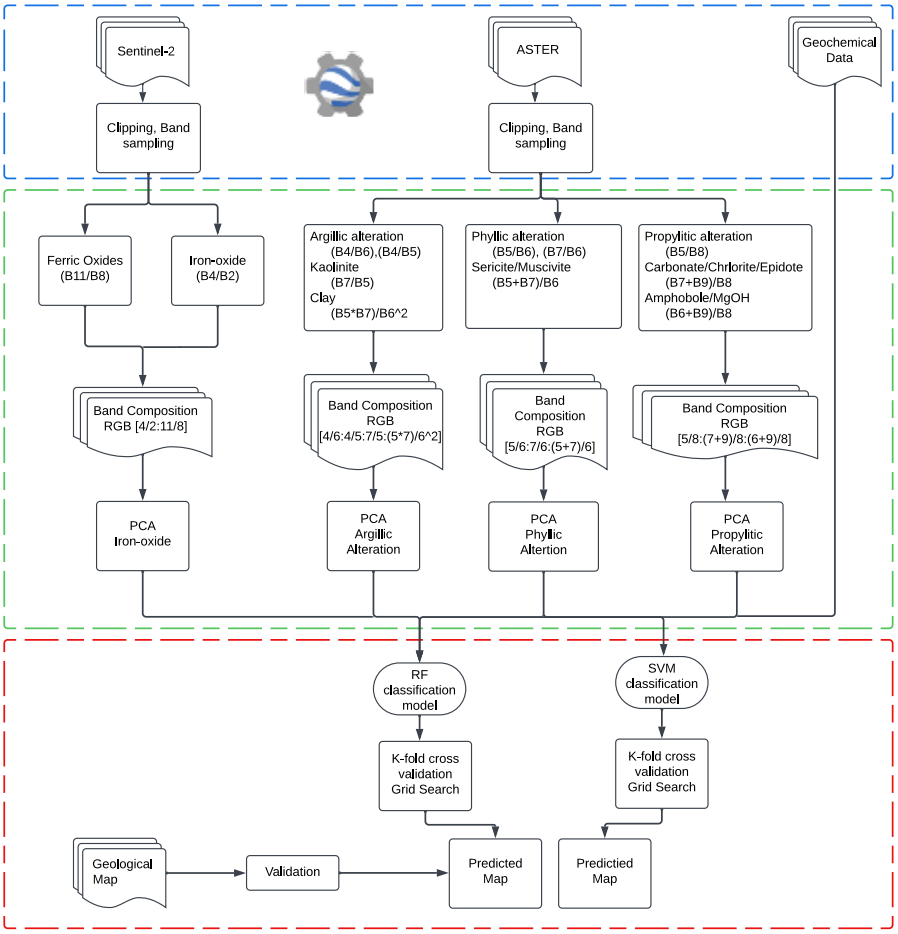


Figure 8. Workflow Diagram for Mineral Prospect Mapping: The blue dotted line represents data collection, the green dotted line indicates pixel enhancement, and the red dotted line signifies model training and anomaly prediction

3.2.1 Band ratio, RGB band composition and Principal Component Analysis (PCA)

Band ratio is an image enhancement technique mainly used to highlight features and minimize topographic effects that are not visible in the unprocessed bands. This technique involves dividing the maximum reflectance by the maximum absorption of the same pixel (Hajaj et al., 2024). It serve as indicators for mineral assemblages or particular mineral groups (Werff & Meer, 2015).

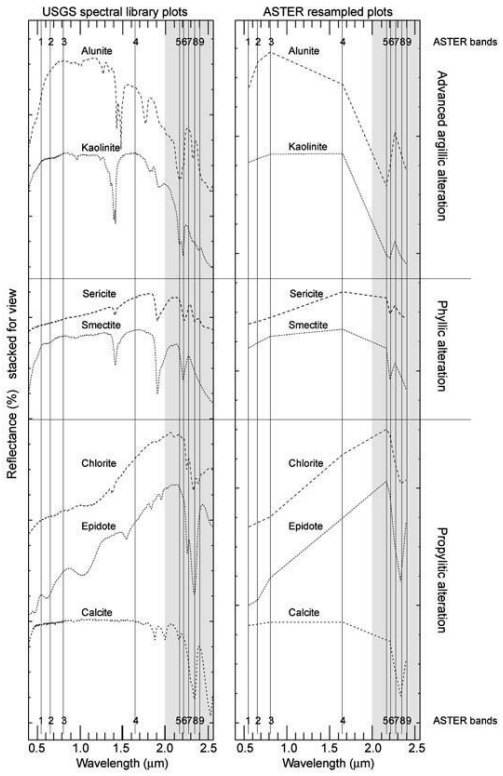
Sentinel-2 VNIR spectral band were utilized to map iron-oxide/hydroxide minerals, as it capture Fe^{3+}/Fe^{2+} information within the spectral wavelength in a region between 0.45-1.2um (Chen et al., 2022). In this study, as shown in Figure 9, band ratio of 4 and 2 were used to identify iron-oxide/hydroxide minerals (Pour et al., 2019; Van der Meer et al., 2014) while band ratio of 8 and 11 was employed to detect Ferric oxides minerals, as there is high reflectance in band 4 and band 11 and high absorption in band 2 and band 8. These band ratios were chosen because they bring out the absorption and reflectance features of target minerals, which are not easily visible in raw data.

ASTER VNIR and SWIR bands were selected to highlight zones of alterations, including argillic, phyllic and propylitic alterations. Different endmember minerals associated with alteration zone were used: minerals like alunite, kaolinite and clay for argillic alteration; sericite, muscovite and sericite-muscovite-illite-smectite for phyllic alterations and epidote-chlorite-carbonate, Amphibole MgOH, and epidote-chlorite-amphibole for propylitic alteration (Van der Meer et al., 2014; Yajima et al., 2007; Zerai et al., 2023). These bands were chosen because they are known to capture the spectral signatures of alteration minerals, which are key indicators of mineralization.

Six bands were selected for band ratio analysis based on their spectral signature of the endmember minerals, Figure 8. These include Band 4, 5, 6, and 7 for argillic alteration; Band 5, 6 and 7 for phyllic alteration; and Band 5, 6, 7, 8 and 9 for prophyllic alteration (Yajima et al., 2007). The individual band index of each alteration type was assigned to primary Red, Green and Blue (RGB) colors. The RGB band composite was used to identifying and visualize the mineral distribution patterns (Hajaj et al., 2024). This approach distinguishes between altered and non- altered area, which was not possible using single band image The preparation of both alteration index maps and

RGB band composites was performed in QGIS using raster calculator and merge tools, respectively.

Principal component analysis (PCA) is a dimensional reduction method used to eliminate redundant information. It transforms multidimensional datasets so that the first principal component (PC) or axis captures the greatest variance. These PCs are uncorrelated linear combinations (eigenvector loadings) of the original data, with each successive PC accounting for progressively less variance (Loughlin, 1991). To calculate the PCA, the mineral indices corresponding to each alteration types were first combined as RGB bands and taken as input using Principal components in the Spatial Analyst extension of ArcGIS Pro. The output image with its associated statistical report were exported as *.img* and *.txt* respectively.



Feature	ASTER	Sentinel-2
Iron		
Ferric Iron, Fe3+	2/1	4/3
Ferrous Iron, Fe2+	5/3 + 1/2	12/8 + 3/4
Laterite	4/5	11/12[a]
Gossan	4/2	11/4
Ferrous silicates (Biotite, chloride, amphibole)	5/4	12/11[a]
Ferric oxides	4/3	11/8
Carbonates/Mafic minerals		
Carbonate/Chlorite/Epidote	(7 + 9)/8	–
Epidote/Chlorite/Amphibole	(6 + 9)/(7 + 8)	–
Amphibole/MgOH	(6 + 9)/8	–
Amphibole	6/8	–
Dolomite	(6 + 8)/7	–
Silicates		
Sericite/Muscovite/Illite/Smectite	(5 + 7)/6	–
Alunite, Kaolinite, Pyrophyllite	(4 + 6)/5	–
Phengitic	5/6	–
Muscovite	7/6	–
Kaolinite	7/5	–
Clay	$(5 \times 7)/6^2$	–
Alteration	4/5	11/12[a]
Host rock	5/6	–
Other		
Vegetation	3/2	8/4
NDVI	(3 − 2)/(3 + 2)	(8 − 4)/(8 + 4)

Feature	ASTER	Landsat 5 TM	Landsat 8 OLI	Sentinel-2A MSI
TM Ratios				
Hydroxyl bearing alteration	4/[5,6,7]	**5/7**	6/7	11/12
All iron oxides	–	**3/1**	4/2	4/2
Ferrous iron oxides	2/4	**3/5**	4/6	4/11

Figure 9. USGS Spectral Reflectance and ASTER Spectral Signatures of Minerals (Left) (adapted from Yajima et al., 2007), and spectral band combinations of ASTER and Sentinel-2 (Right) (source: Hajaj et al., 2024; Werff & Meer, 2015)

3.3 Training Test and validation data preparation

The geochemical data, consisting of rock chip and soil sample, contains four elements: Au, Cu, Pb and Zn. Anomalous regions of the study area are determined by threshold values where Au ≥ 10 ppb, Zn ≥ 200 ppm and Cu and Pb ≥ 100 ppm (kifletsion Berhe & Chisholm, 2013). Using QGIS raster calculator, point samples exceeding these thresholds for all four elements were assigned to 1 (anomalies) while those below the thresholds were assigned to 0 (non- anomalies) as shown in Table 2. This step creates a binary classification problem that fits for machine learning.

Table 2. Geochemical Data showing the sample location (Easting, Northing) and concentrations of Au, Cu, Pb, and Zn, along with Anomaly Classification (Class 1 = Anomalies, Class 0 = Non-anomalies)

Easting	Northing	Au_ppb	Cu_ppm	Pb_ppm	Zn_ppm	classvalue
394864.43	1858224.83	41	20	10	66	1
394438.95	1858357	7	18	10	63	0

The first principal component (PC) for each alteration, calculated using PCA, was used as a parameter. The pixel values corresponding to these parameters were assigned to geochemistry point samples at their respective locations, as illustrated in Table 3, using a point sampling tool QGIS plugins. This integration aligns the spectral and geochemical data, allowing the models to learn from both datasets.

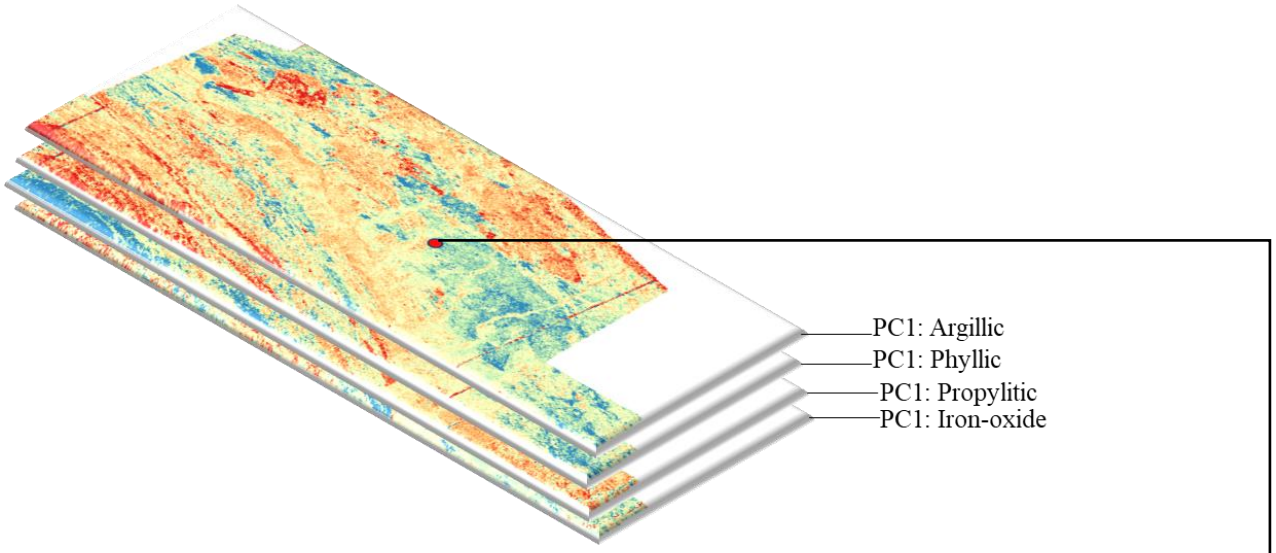


Table 3. Integration of Principal Component Analysis (PCA) Layers for Hydrothermal Alteration Maps with Geochemical Data

Easting	Northing	Au_ppb	Cu_ppm	Pb_ppm	Zn_ppm	Phyllic_PC	Iron-oxide	Argillic_P	Propylitic	classvalue
394864.43	1858224.83	41	20	10	66	2.69009	2.38696	1.86079	3.12537	1
394438.95	1858357	7	18	10	63	2.60145	2.07293	1.8853	2.89364	0

The final dataset was exported as a GeoJSON file to be used for training machine learning prediction models, specifically Random Forest (RF) and Support Vector Machine (SVM), in Google Colab. The dataset was split into 70% for training, 15% for testing, and 15% for validation. The dataset has imbalanced dataset, the approximate ratio of major classes to minor class is 5:1. This tells us for every 5 samples in class 0, there is 1 sample in class 1. This imbalance created a major challenge, as preliminary attempts to train models on this dataset resulted in poor performance, with models biased toward the majority class (non-anomalies). To address data imbalance, Figure 10, oversampling and under-sampling techniques were applied using SMOTE (Synthetic Minority Oversampling Technique) and Random Undersampling, respectively, based on the original data. The SMOTE oversampling technique does not alter the geographical distribution of the data but increases the training sample size by generating synthetic data points

within the feature space of the training set (Prado et al., 2020). This resampling technique help the models to learn from both classes and improve the detection of rare yet important anomalies.

Figure 10. Class Distribution in the Original Dataset, Showing Imbalance Between Class 0 and Class 1. There are 1194 non-anomalies and 237 anomaly data

3.4 Model Training-Random Forest and Support Vector Machine

Both Random Forest and SVM are non-parametric supervised machine learning algorithms which are commonly used in anomaly detection (Josso et al., 2023; Lachaud et al., 2023). These machine learning models were selected based on their stability and effectiveness in high dimensional and small spatial datasets of remote sensing, and they outperform other machine learning classifiers (Josso et al., 2023; Mountrakis et al., 2011). These two predictors were used to distinguish input data (geochemical data along the first PC of alterations) into binary or two classes: anomalies (1) and non-anomalies (0).

The datasets were split into training, test and validation, as well as oversampling was applied to the minority class (i.e., class 1) and under sampling was applied to the majority class (i.e., class 0) before training the mode. The entire process was conducted in Google Colab which provides good computational resources to train the model efficiently.

To optimize model performance, grid search and K-fold cross-validation were applied. The k-fold involves splitting the training dataset into multiple folds. For example, Figure 11, if the number of folds is 10, the model is trained on nine folds while the remaining fold is reserved for testing in each iteration. This ensures the model generalizes well. In this study the model was trained using 10-fold cross-validation. Grid search is a systematic approach to finding the best combination of hyperparameters for a model. The grid search used in Random Forest algorithm are: *n_estimators*, *criterion, max_depth, min_samples_leaf, min_samples_leaf, max_features,* and *class_weight* whereas for SVM *C, gamma* and *kernel* are used. The range of value assigned to each parameter are giving in the Table 4. Best parameters of each classifier were then chosen based on F1 score of the positive classes (class = 1 or anomalies).

Table 4. Hyperparameter Ranges Used for Random Forest and Support Vector Machine Classifiers. The table lists the specific parameter values and ranges applied for each classifier

Random Forest	
Parameters	**Value**
n_estimators	50, 100, 200, 400, 500
criterion	gini, entropy, log_loss
max_depth	5, 10, 15, 20, 25, None
min_samples_split	5, 10, 15, 20, 25, 30
min_samples_leaf	5, 10, 20
max_features	0.5, 1, 1.5, 2, sqrt

Support Vector Machine	
Parameters	**Value**
C	0.1 - 50.1: interval 0.1
gamma	0.1 - 1.01: interval 0.01
kernel	rbf, linear, poly

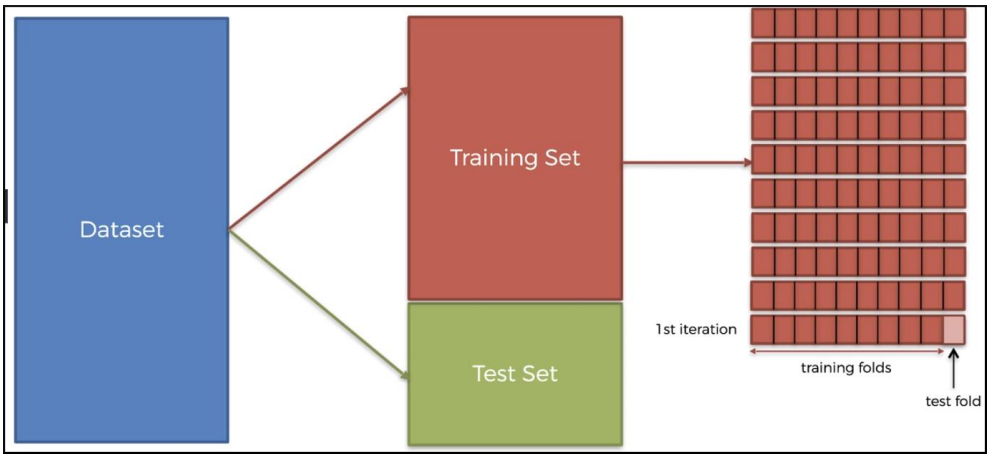


Figure 11. K-Fold Cross-Validation Process for Model Training and Testing (source: Zhongyang Wang, n.d.)

The predicted maps generated from RF and SVM were exported as GeoTIFF and visually validated against existing traditional mining (artisanal works) and high anomalies identified in the VETM geophysical survey to determine whether both benchmarks fall within the predicted anomaly zones.

3.5 Model Evaluation

In imbalanced data, accuracy is not the best choice for evaluating a model because accuracy is simply the percentage of correctly classified predictions divided by the total predictions. The formula for accuracy is:

$$Accuracy = \frac{(TP + TN)}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})}$$

True Positive (TP):
- Anomaly detected correctly

True Negative (TN):
- No anomaly correctly classified as no anomaly

False Positive (FP):
- The model predicts an anomaly, but there is no anomaly

False Negative (FN):
- The model predicts no anomaly, but an anomaly is actually present

Based on this formula, the positive (minority) class has small impact compared to the negative (majority) class on the accuracy metric. To address this issue with imbalanced datasets, F1-score

and the Precision-Recall curve are more appropriate metrics, since the main goal of this research is to identify Class 1 (anomalies). The formula for F1-score is as follows:

$$F1 - score = 2\,x\,\frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

The F1-score focuses on the minority class (Class 1, representing the minority) to evaluate the model's ability to correctly identify these classes. On the other hand, the Precision-Recall curve plots Recall (True Positive Ratio) on the x-axis and Precision on the y-axis for different classification thresholds. It illustrates how the balance between these two metrics changes as the threshold for classifying positive (mineralized) instances is adjusted. This curve is essential for visualizing model performance across different thresholds.

The AUC (Area Under the Curve) in the Precision-Recall curve represents the area under the curve and provides a single value ranging from 0 to 1. Ideal or perfect models have an AUC of 1, indicating the model's high precision and recall at all thresholds. An AUC of 0.5 represents random classification, while an AUC of 0 indicates the model fails to correctly identify the positive (minority) class. These metrics provide accurate evaluation of the model's ability to detect anomalies, which is the primary focus of this study.

4. Result and Discussion

4.1 Band Ratio and RGB Color Composition

4.1.1 Iron oxide / hydroxide minerals

Figure 12A-B shows the indices map for iron oxide/ hydroxide minerals from a sentinel-2 data. Band 4/band 2 was used to highlight areas of iron oxide/hydroxide minerals whereas Band 8/Band 11 was used to map ferric oxides. In both the images the red color (high values in the legends - 2.82 and 2.45) corresponds to the normalized or scaled intensities of iron oxide and ferric oxide reflectance calculated using band ratios respectively. The areas enriched in iron oxides could be

attributed to the presence of massive dark meta-basalt rocks found in the region (kifletsion Berhe & Chisholm, 2013). There is a concentrated presence of iron oxide and ferric oxide along the main cross-cutting structure (fault) trending NNE. Generally, both minerals are distributed in a NE direction and scattered throughout the study area, but they are mostly concentrated in the southern part. Using an RGB band composition for these two minerals helps us visualize and analyze their spatial distribution. The false color RGB image is generated using indices of iron oxide/hydroxide and ferric oxide assigned to the red and green channels, respectively. The study area is characterized by a reddish coloration, even visible in satellite imagery, due to the dominant presence of hematite (kifletsion Berhe & Chisholm, 2013). The result on Figure 13A illustrates yellow areas where both iron oxide/hydroxide and ferric oxide are abundant. This region may indicate zone where hematite coexists with other ferric iron oxides. Red area, dominated by high iron oxide/hydroxide and low ferric oxide, shows the presence of hematite, indicating oxidized formation i.e. banded iron formations (BIFs) or regions affected by iron rich hydrothermal activity. Green area, indicate high ferric oxide and low iron oxide/hydroxide, could represent areas limonite or goethite dominate. These might indicate a transitional zones or weathered layers overlying hematite-rich deposits. Orange area shows area with moderate to high in both minerals and may represent a mixed iron oxide phase, including hematite and other ferric oxides, possibly represent areas where transitional alteration zones. Finally, the dark areas represent low in both minerals, are likely indicative of fresh or lithological unites with minimal iron content. Concentrated patches of red colors along linear features represent fault-controlled hematite mineralization or veins, while broader expanses of red and orange corresponds to iron rich stratigraphic layers like oxidized meta sediment units.

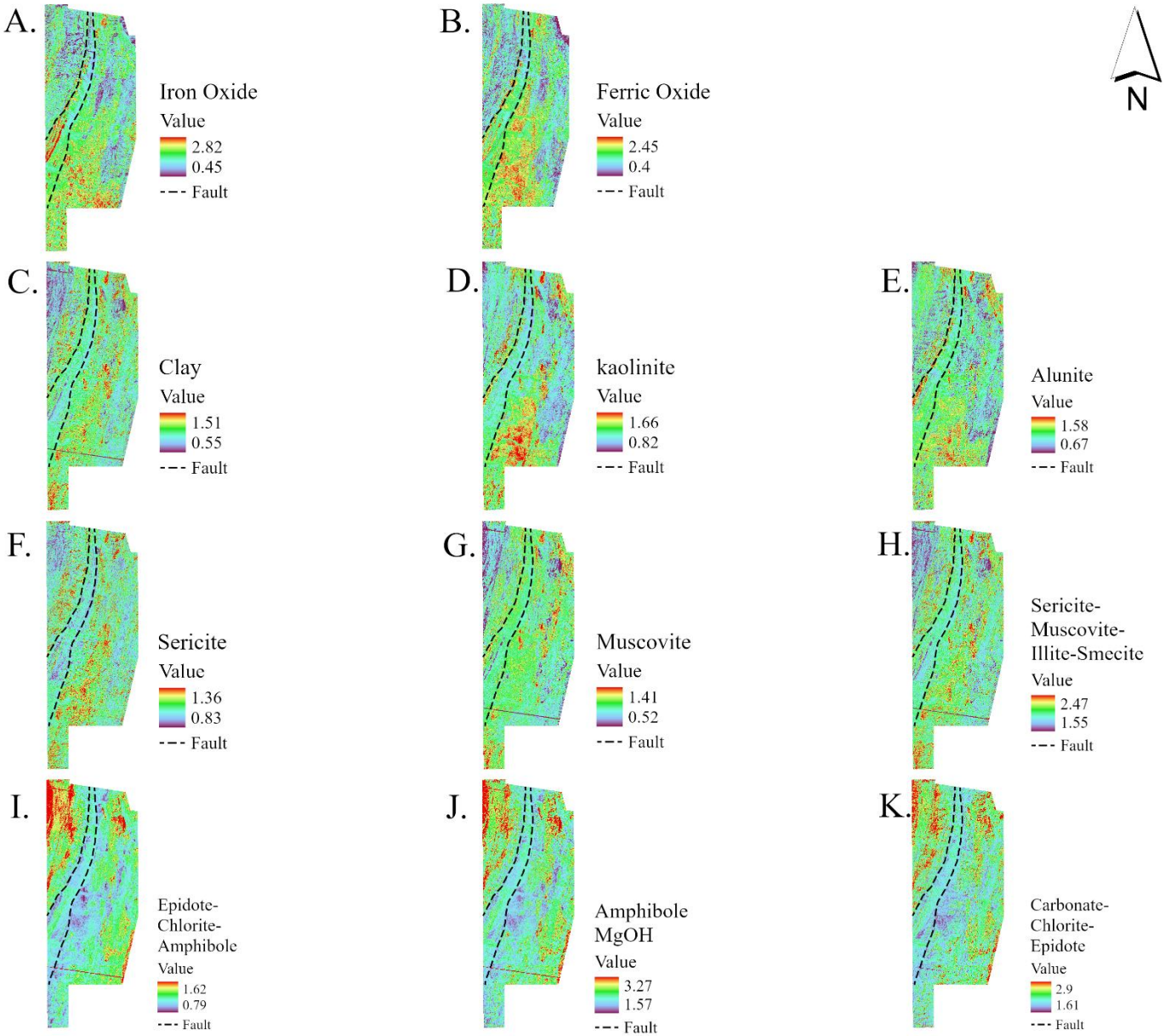


Figure 12. Mineral Index Maps: Maps (A-K) display the spatial distribution of various mineral indices, including Iron Oxide (A), Ferric Oxide (B), Clay (C), Kaolinite (D), Alunite (E), Sericite (F), Muscovite (G), Sericite-Muscovite-Illite-Smecite

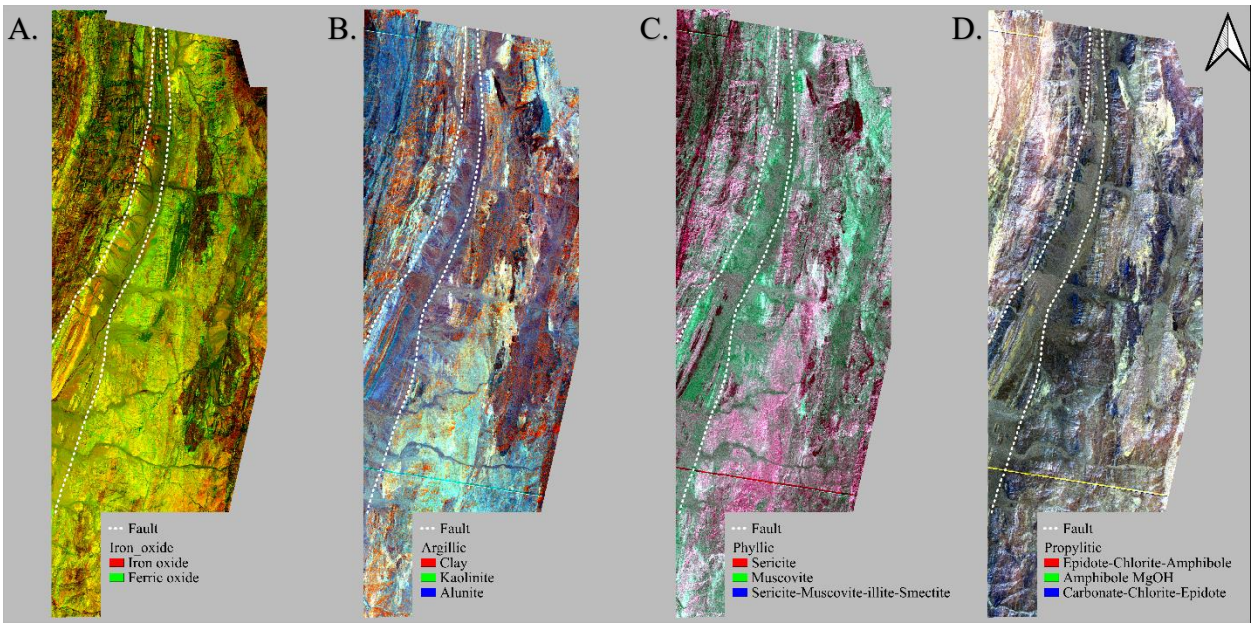


Figure 12. RGB Color Composition Maps for Mineral Alteration Based on Alteration Indices Ratios

4.1.2 Argillic alterations

The endmember minerals for argillic alteration such as kaolinite, alunite and clay show a similar pattern as depicted in Figure 12C-E. The red color (with values of 1.51, 1.66 and 1.58) indicates the presence of the Clay, Kaolinite and Alunite minerals, respectively, that made up the argillic alterations. In the central and southern part of the study area there is an indication of concentrated kaolinite, alunite and clay. Three of the indices show a NE trend pattern, specifically in some portion of kaolinite and alunite indices, go along the major fault NNE the same as iron oxide/hydroxide and ferric oxides indices. In the false color composite as shown in Figure 13B, Red, Green and Blue represent clay, kaolinite and alunite respectively, provide a detailed mineralogical spatial distribution. Cyan represents the presence of both kaolinite and alunite. Magenta area points to overlapping clay and alunite mineralization, while yellow area indicate zones both clay and kaolinite. Regions of white color represent highly altered zones from clay, kaolinite and alunite. This area is associated with advanced argillic alteration signifying a porphyry-epithermal deposits. Advanced argillic alteration is dominated by high temperature clay

minerals like kaolinite and alunite and found in high sulfidation porphyry and epithermal deposit (Shanks III, n.d.). This area falls within or proximal to the traditional mining and VETM anomalies zones. The dark area represents, lack of the end member minerals, unaltered or minimally altered zones.

4.1.3 Phyllic alterations

The constituent of phyllic alterations in this study were sericite, muscovite and sericite-muscovite-illute-smecite minerals. Figure 12F-H, show the distribution of these minerals and it can be seen that their concentrated muscovite minerals in the NW corner of the study area and a thin elongated patched in the fault zone. The patterns of sericite and sericite-muscovite-illute-smecite have homogenous NE trend. In the areas of traditional mining area, there is significant concentration of sericite-muscovite-illite-smecite minerals. Figure 13C, is a false color map where the red channel represent sericite, green channel represents muscovite and blue channel for the combination of sericite, muscovite, illite and smecite. Yellow areas represent sericite and muscovite coexist, indicating transitional zones between phyllic and potassic alteration. Cyan area shows the presence of both of muscovite plus sericite with secondary clay minerals such as illite and smectite. This may reflect argillic alteration, where weathering processes altered muscovite rich rocks and introducing additional clays. Magenta zones indicate areas dominated by sericite with additional secondary clay minerals such as illite and smecite. This region represents intense hydrothermal alterations where sericite is the primary alteration product but the secondary minerals could be from the weathering.

4.1.4 Porphyritic alterations

The porphyritic alterations displayed in Figure 12I-K consist of Epidote-chlorite-Amphibole, MgOH Amphibole and Carbonate-Chlorite-Epidote indices map. Each indices map shows concentrated minerals in the NNW and a thin patch in the ESE indicated as red color. In the false

color map, the red channel represents Epidote/chlorite/Amphibole, the green channel is MgOH Amphibole and the blue channel indicate the Carbonate/Chlorite/Epidote indices. Based on the primary color composition in Figure 13D, the yellow zones show the presence of both epidote-chlorite-amphibole and MgOH amphibole. This zone could indicate a lithological transition in mafic to ultramafic sequences, and higher temperature alteration as MgOH amphiboles are stable in high grade and magnesium rich metamorphic environments (Andersson et al., 2006). Cyan zones highlights where MgOH amphibole is mixed with carbonate alteration and low-grade metamorphic assemblages (chlorite and epidote). This zone could indicate late stage hydrothermal alteration (Li et al., 2022). Magenta zone represent the coexistence of both propylitic alterations (epidote-chlorite-amphibole) and carbonate alteration. This zone is likely to ore hosting since high temperature epidote-chlorite-amphibole assemblages react with carbon rich fluids produces carbonate alteration and carbonate alterations are common in porphyry and skarn system (Sillitoe, 2010).

4.2 Principal component analysis results

The PCA reports generated from band ratio and RGB color composition of Sentinel-2 and ASTER are in the Appendix 1. The first principal component (PC1) was taken chosen as it contains more descriptive information than the subsequent PCs as can be seen from Table 5 below. The PCA results for iron oxide (layer 1) and ferric oxide (layer 2) from Sentinel-2 imagery show a moderate positive correlation 0.63155. This indicates a nearly linear increase between both iron oxide and ferric oxide. The eigenvalue of PC1 is 81.61 % of the total variance, whereas the second component (PC2) holds the remaining 18.39%. This informs that most of the variability in the data can be summarized by PC1, which is influenced by contributions from both iron oxide (0.72649) and ferric oxide (0.68717). PC1 in Figure 14A represents the overall distribution of both minerals within the study area.

Table 5. Principal Component Analysis (PCA) of Sentinel-2 Imagery – Correlation, Eigenvalues, Eigenvectors, and Percentage of Variance Explained for Iron Oxide and Ferric Oxide Distribution

	Correlation			**Eigenvalues, Eigenvectors and % Eigenvalues**			
Layer	1	2	Eigenvector	Layer 1	Layer 2	Eigen Value	% Eigen Values
1	1	0.63155	PC1	0.72649	0.68717	0.04762	81.6068
2	0.63155	1	PC2	0.68717	-0.7265	0.01073	18.3932

The PCA results for argillic alterations such as clay (layer1), kaolinite (layer2) and alunite (layer3) show relationships and variability among these minerals. The covariance matrix indicate that clay and kaolinite revel a strong variance compared to clay and alunite. This is supported by the correlation matrix, clay and kaolinite have a moderate positive correlation (0.51468), kaolinite and alunite have strong positive correlation (0.78948) and clay and alunite are uncorrelated (-0.00363). The eigenvalue, helps us to know how much of the variation in the data each PC explains, shows that the first principal component (PC1) explains 62.9986% of the variations. This means PC1 captures the overall shared variability between the materials, with the strongest contribution from clay (0.62702) and kaolinite (0.67004), while less contribution from alunite (0.39736), Table 6 summarize the PCA results. From a geological point of view, In Figure 14B, PC1 shows the overall amounts and trends of clay and kaolinite with less amount alunite. This could represent intermediate argillic alteration patterns in epithermal or porphyry type deposit.

Table 6. Principal Component Analysis (PCA) Results for Argillic Alterations – Correlation, Eigenvalues, Eigenvectors, and Percentage of Variance Explained for Clay, Kaolinite, and Alunite Distribution

	Correlation				**Eigenvalues, Eigenvectors and % Eigenvalues**				
Layer	1	2	3	Eigenvector	Layer 1	Layer 2	Layer 3	Eigen Value	% Eigen Values
1	1	0.51468	-0.00363	PC1	0.62702	-0.71042	0.31962	0.00602	62.9986
2	0.51468	1	0.78948	PC2	0.67004	0.28254	-0.68646	0.00337	35.31
3	-0.00363	0.78948	1	PC3	0.39736	0.64458	0.65316	0.00016	1.6914

The PCA's report, described in Table 7, for phyllic alteration of sericite (layer1), muscovite (layer2) and sericite-muscovite-illite-semctite (layer3) minerals, shows that there is a first ranked correlation between muscovite and the composite layer 3 (0.86378) then between sericite and

composite layer 3 (0.83919) and last is the correlation between sericite and muscovite (0.45084). This means the composite layer 3 is closely tied to both sericite and muscovite because it contains characteristics of both minerals. The eigenvalue of PC1 explains 88.8573% of the overall variation in the dataset, as presented in Figure 14C. The eigenvectors for PC1 shows the composite layer 3 has the largest contribution (0.81601), while sericite (0.38348) and muscovite (0.43253) have moderate influence. In relation to mineral deposit, the PC1 is characterize phyllic alteration zone, which is commonly linked with porphyry copper deposit, where minerals like sericite and muscovite are dominant.

Table 7. Principal Component Analysis (PCA) Results: Correlation, Eigenvalue, and Eigenvector Analysis of Sericite, Muscovite, and sericite-muscovite-illite-semctite in Phyllic Alteration

	Correlation				**Eigenvalues, Eigenvectors and % Eigenvalues**				
Layer	1	2	3	Eigenvector	Layer 1	Layer 2	Layer 3	Eigen Value	% Eigen Values
1	1	0.45084	0.83919	PC1	0.38348	0.72084	-0.57735	0.00603	88.8573
2	0.45084	1	0.86378	PC2	0.43253	-0.69252	-0.57735	0.00076	11.1427
3	0.83919	0.86378	1	PC3	0.81601	0.02832	0.57735	0	0

The correlation matrix, generated from PCA epidote-chlorite-amphibole (layer1), MgOH-amphibole (layer2) and carbonate-chlorite-epidote (layer3), shows a strong positive correlation between layer 1 and layer 2 (0.94510), another positive correlation between layer 2 and layer 3 (0.88675) and a moderate correlation between layer 1 and layer 3 (0.76354). These results revel that layer 2 (MgOH-amphibole) is closely related to both layers due to shared mineral features. In the eigenvalue results, presented in Table 8, the PC1 explains 93.3001% of the total variation, which means it captures all the important pattern in the data. And the eigenvectors for PC1 show that layer 2 (0.75232) and layer 3 (0.57292) contribute the most to this component whereas layer 1 (0.32311) has small contribution. Geologically, the PC1, Figure 14D, indicate a propylitic alteration. This type of alteration is commonly found in porphyry copper-gold deposits and is identified by minerals like epidote, chlorite, amphiboles and carbonates.

Table 8. Correlation, Eigenvalue and Eigenvector analysis of PCA results for Epidote-Chlorite-Amphibole, MgOH-Amphibole, and Carbonate-Chlorite-Epidote in Propylitic Alteration

	Correlation				Eigenvalues, Eigenvectors and % Eigenvalues					
Layer	1	2	3	Eigenvector	Layer 1	Layer 2	Layer 3	Eigen Value	% Eigen Values
1	1	0.9451	0.76354	PC1	0.32311	-0.46265	0.82557	0.01842	93.3001
2	0.9451	1	0.88675	PC2	0.75323	-0.4024	-0.52031	0.00119	6.0456
3	0.76354	0.88675	1	PC3	0.57292	0.78996	0.21846	0.00013	0.6542

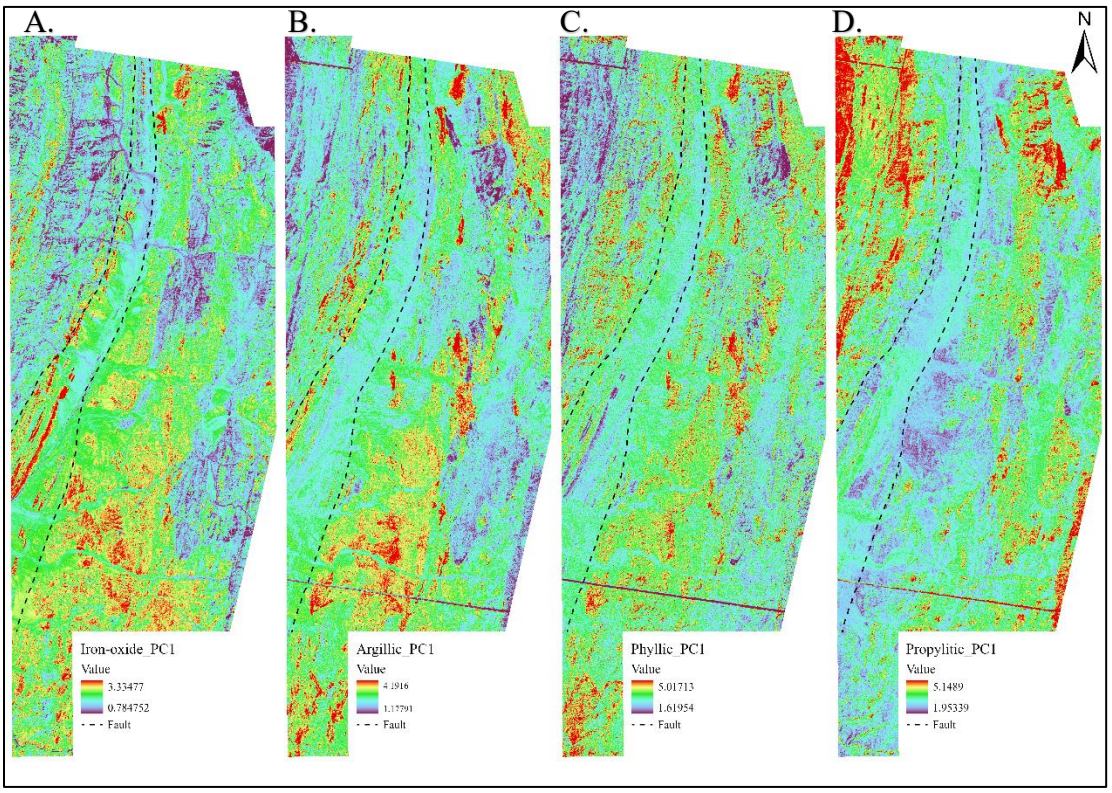


Figure 13. First Principal Component (PC1) Maps for Different Alteration Types, Showing Spatial Distribution of Iron-Oxide, Argillic, Phyllic, and Propylitic Alterations Along Fault Zones.

4.3 Random Forest and Support Vector Machine

The prediction of an anomaly map was generated using Random Forest (RF) and Support Vector Machine (SVM) classifiers. Both classifiers were trained using the parameters shown in Table 9. The parameter values were selected based on the F1-score of the minority class (Class 1 or anomalies) during grid search. This was applied to the original model, oversampled model, and under sampled model, with validation performed using 10-fold cross-validation.

The reason for setting the grid search score to the F1-score of the minority class instead of accuracy is that Class 1 (minority class) is of primary interest, as it identifies potential anomaly zones. By focusing on the F1-score, the model is better tuned to balance precision (how many predicted anomaly areas are correct) and recall (how many actual anomaly areas are detected).

This balance is crucial because it reduces errors such as missing anomaly areas (false negatives) and over-predicting anomalies (false positives). By prioritizing the F1-score, the model avoids favoring the majority class (Class 0) and provides more reliable predictions for the minority class across both the original and resampled datasets.

Table 9. Hyperparameters used for Training Random Forest and SVM classifiers under Different Sampling Techniques

	Random Forest		
Parameters	Original	Oversampled	Under sampled
N estimator	100	200	100
Criterion	entropy	entropy	entropy
Max depth	15	15	15
Min samples leaf	5	5	5
Min samples split	10	10	5
Max features	sqrt	sqrt	sqrt

	SVM		
Parameters	Original	Oversampled	Under sampled
Kernel	rbf	rbf	rbf
C	9.6	44.5	9.6
Gamma	0.99	0.97	0.99

The performance of the RF classifier trained on the original dataset is 0.75, showing the ability of a RF in handling an imbalanced dataset. After applying oversampling to minority class (anomalies) to balance it with the majority class (non-anomalies), the F1-score increase to 0.81. This indicates that oversampling improves the ability of the model to identify the anomalies areas. The balanced distribution of classes leads to better model performance. However, the F1-score of the

undersampled RF model decreased to 0.79. This decline is likely due to the reduction in dataset size after applying undersampling.

The SVM original model has an F1-score of 0.35, indicating that SVM model struggles with imbalanced dataset and this is possibly due to its sensitivity to class imbalance. F1-score increases to 0.75 in the oversampled SVM mode. This improvement tells that SVM benefits from balanced dataset, as oversampling addresses the class imbalanced issue. The same as RF under sampled model, SVM under sampled model also show a decrease in F1score 0.70, slightly lower than its performance on the oversampled data.

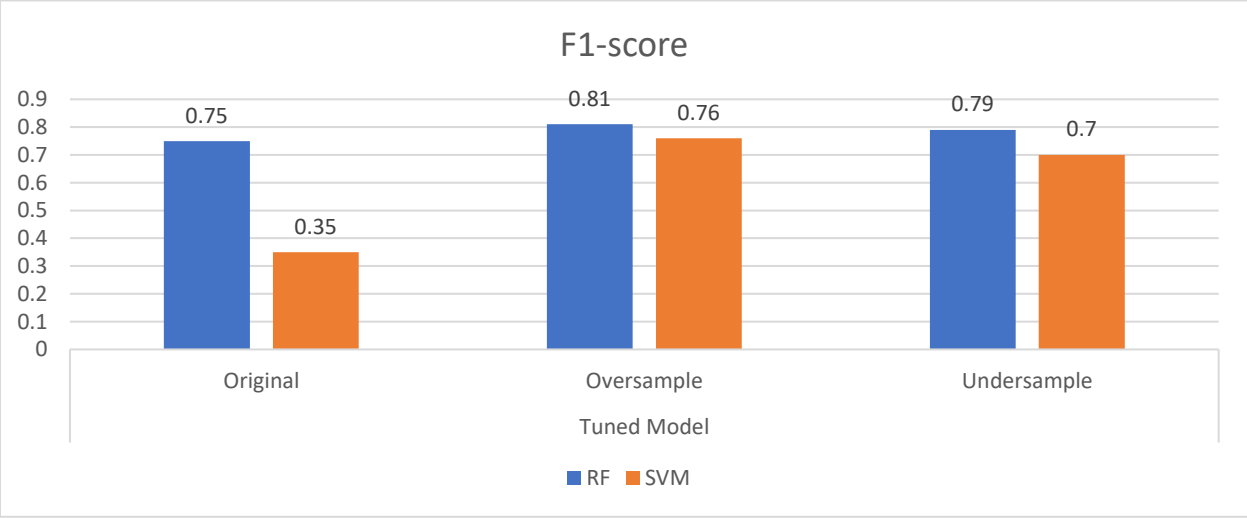
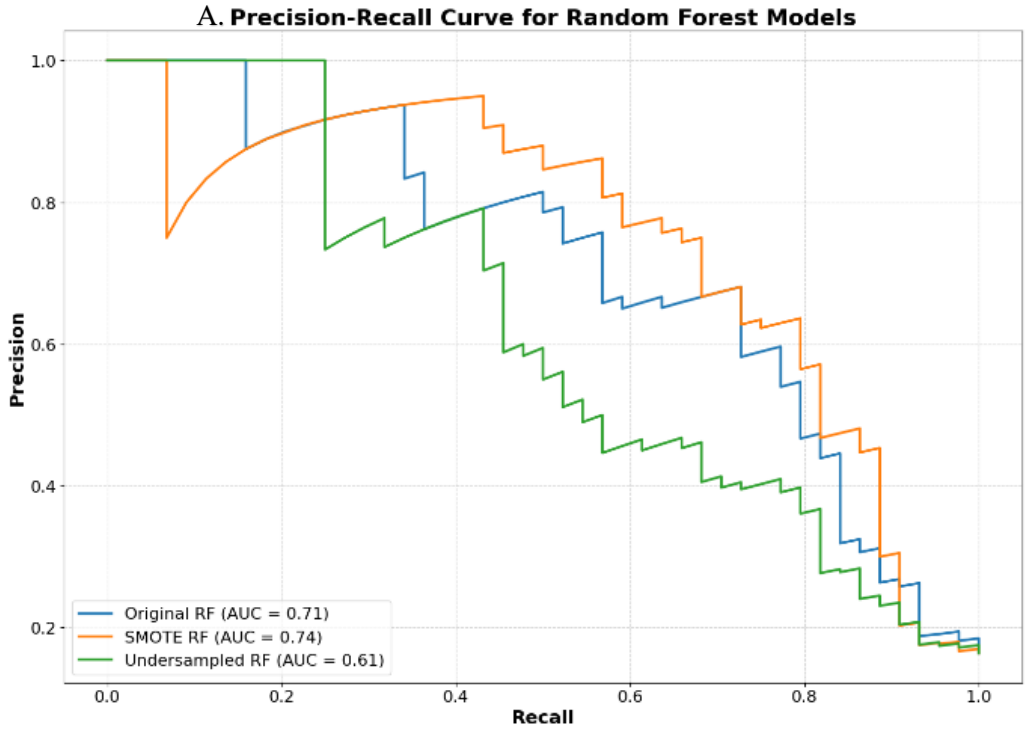


Figure 14. Comparison of F1-Scores for Random Forest (RF) and Support Vector Machine (SVM) Models Across Original, Oversampled, and Undersampled Datasets.

4.4 Evaluation of the Predictive Models' Performance

Based on F1-score in Figure 14, the RF models got first rank compared to SVM models. Specifically, the Oversampled model of RF has the highest F1-score of 79 %. To better understand which model could be used to predict the area, Precision-Recall curve was performed and as in the Figure 16 shows the RF of the oversampled model stood out over the other models.

The PR plot shows how well two models, RF and SVM, conducted under different sampling methods: oversampling, original data and under sampling. RF, in Figure 15.A, with oversampling perform the best, with AUC-PR of 0.74, followed by the original data model at 0.71. However, RF with undersampled data has lowest AUC-PR of 0.61, likely due to small number of minor class samples. For SVM, in Figure 15.B, the best performance was with oversampling, AUC-PR = 0.60. Both original data and under sample are below the average AUC-PR which is 0.5. This could be due to imbalanced class distribution and insufficient data set. The RF's higher AUC specifically with oversampling, make it the preferable choice over other models and this result also algins with the ranking result made based on F1-score.



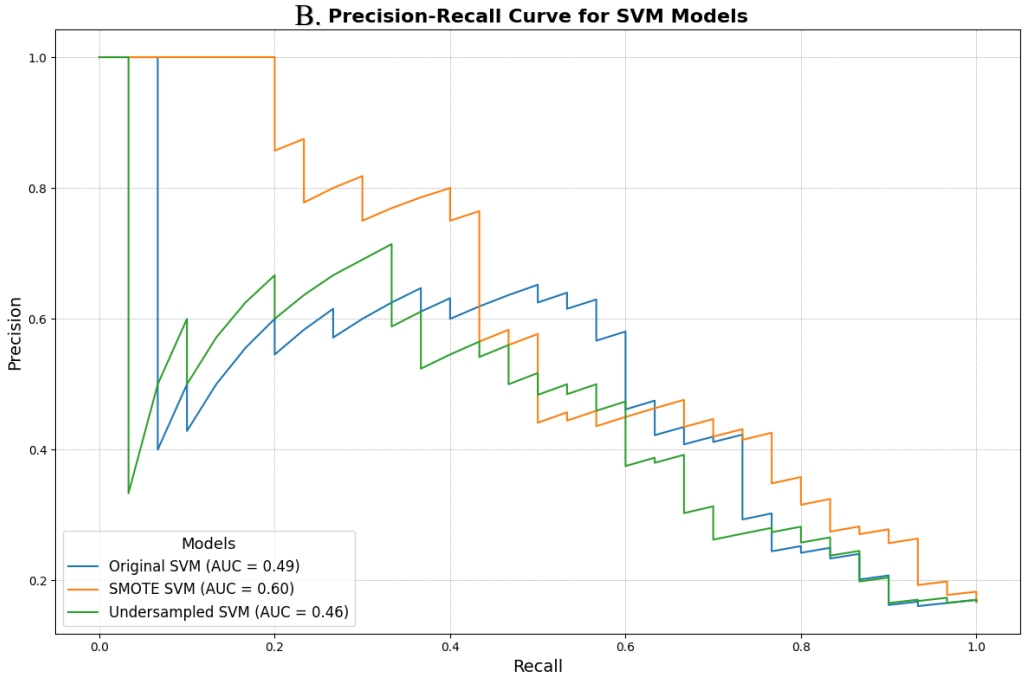


Figure 15. A. Precision-Recall Curves for Random Forest (RF) and B. Support Vector Machine (SVM) Models Under Different Sampling Techniques, Showing AUC Performance Comparisons.

4.5 Feature Importance

RF has the ability to evaluate the importance of inputs predictors by ranking the contribution of each predictor to the model's predictions (Josso et al., 2023). Figure 16 illustrates that Iron_oxide_PCA1 is highest ranked feature, with an importance value of 0.35, making it the most dominant predictor in the model. This result also aligns with the local geology of the study area, where it is visible that the study area is predominantly characterized by the presence of hematite minerals having reddish-brown color. The second most important features are Phyllic_PCA1 with an importance value of near to 0.25. The remaining features, Argillic_PCA1 and Propylitic_PCA1 have similar importance values, each around 0.20. The result shows that Iron-oxide and Phyllic are the primary drivers of the model's performance while Argillic and Propylitic serve as a supportive feature.

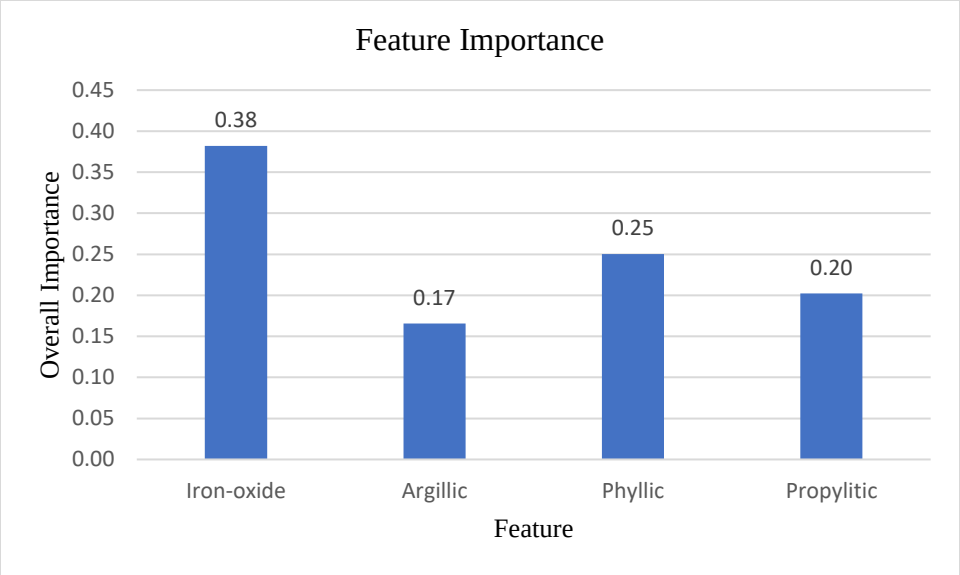


Figure 16. Feature Importance Ranking from Random Forest Model for Mineral Alteration Predictors

The final predicted map, generated by the oversampled RF model for the entire study area is shown in Figure 17. The red color indicates anomalous areas for Au, Cu, Pb, and Zn. A higher concentration of these anomalies is observed in tshe southern part of the study area and along the regional fault. Most traditional mining activities are located within the predicted anomaly zones or within a 100-meter buffer zone around them. Another validation is provided by the VTEM survey, where areas of high anomalies confirmed by the geophysical survey strongly coincide with the predicted zones, particularly in the north and northeast. The anomalies predicted within the fault zones are areas of interest, as hydrothermal deposits often show a positive correlation with structural features.

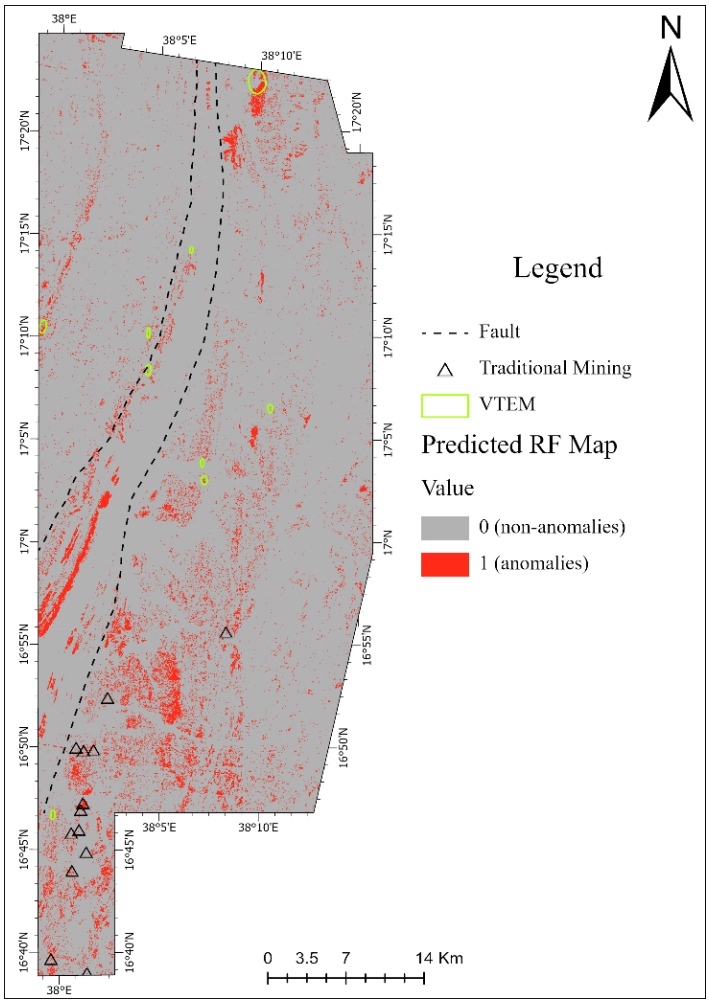


Figure 17. Predicted Anomalies Map Generated by the Oversampled RF Model, Showing Anomalous (Red) and Non-Anomalous (Gray) Areas Along Fault Zones, with Validation from Traditional Mining Sites and VTEM Survey Data

5. Conclusions

Remote sensing and machine learning have created new opportunities for mineral exploration, specifically in regions with limited access to comprehensive geochemical data. This study presents a practical approach to mineral prospect mapping by combining ASTER and Sentinel-2 satellite data with machine learning algorithms in the Adobha region of Eritrea. The goal was to predict mineral anomalies using limited geochemical parameters, addressing a common challenge in resource-constrained environments. The methodology integrated image enhancement techniques with machine learning classifiers. The Random Forest model performed well, achieving an F1

score of 0.81 after handling class imbalance through oversampling. The model's predictions showed strong correlation with known traditional mining sites and geophysical survey results, validating its effectiveness. The approach offers several key advantages. It works well with limited geochemical data, making it cost-effective for initial exploration. The methodology is scalable and can cover large areas. Additionally, the common issue of imbalanced geological datasets was successfully managed through effective sampling techniques. However, some limitations are acknowledged. The model relies on four geochemical parameters (Au, Cu, Zn, and Pb), which may not capture all aspects of mineralization. The study focuses on a single region, limiting the generalizability of the findings. Traditional machine learning algorithms (RF and SVM), while effective, might miss some complex geological patterns. Future work should focus on testing this approach in different geological settings. Incorporating more advanced machine learning techniques (e.g., deep learning) could improve results. Developing user friendly tools would make this methodology more accessible to exploration companies. Adding other data sources, like detailed geophysical surveys, could enhance the model's accuracy. This study demonstrates a practical way to combine remote sensing and machine learning for mineral exploration. Our results show that effective mineral prospecting is possible even with limited data. While there's room for improvement, this work provides a solid foundation for advancing automated mineral exploration techniques.

References

Andersson, U. B., Ghebreab, W., & Teklay, M. (2006). Crustal evolution and metamorphism in east-central Eritrea, south-east Arabian-Nubian Shield. *Journal of African Earth Sciences*, *44*(1), 45–65.

Barrie, C. T., Nielsen, F. W., & Aussant, C. H. (2007). The Bisha volcanic-associated massive sulfide deposit, western Nakfa terrane, Eritrea. *Economic Geology*, *102*(4), 717–738.

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

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), 945–963.

Berhe, kifletsion, & Chisholm, J. (2013). *Adobha Annual Report for the period ending 22 July 2012* (Nos. GIP13-004; p. 87).

Berhe, S. M. (1990). Ophiolites in Northeast and East Africa: Implications for Proterozoic crustal growth. *Journal of the Geological Society*, *147*(1), 41–57.

Chalice Gold Mines. (2011). *Chalice Gold Mines* [Annual Report].

Chen, Q., Xia, J., Zhao, Z., Zhou, J., Zhu, R., Zhang, R., Zhao, X., Chao, J., Zhang, X., & Zhang, G. (2022). Interpretation of hydrothermal alteration and structural framework of the Huize Pb–Zn deposit, SW China, using Sentinel-2, ASTER, and Gaofen-5 satellite data: Implications for Pb–Zn exploration. *Ore Geology Reviews*, *150*, 105154.

Drury, S., & Berhe, S. (1991). *Thematic mapper data for geological reconnaissance of pan-African terranes in Eritrea. 2. The context for water-resources*.

Drury, S., & Berhe, S. (1993). Accretion tectonics in northern Eritrea revealed by remotely sensed imagery. *Geological Magazine*, *130*(2), 177–190.

Drury, S., & De Souza Filho, C. (1998). Neoproterozoic terrane assemblages in Eritrea: Review and prospects. *Journal of African Earth Sciences*, *27*(3–4), 331–348.

Duda, K., Daucsavage, J., Siemonsma, D., Brooks, B., Oleson, R., Meyer, D., & Doescher, C. (2015). Advanced spaceborne thermal emission and reflection radiometer (aster) level 1 precision terrain corrected registered at-sensor radiance product (ast_l1t). *US Geological Survey, USA*.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.06.031

Hajaj, S., El Harti, A., Jellouli, A., Pour, A. B., Himyari, S. M., Hamzaoui, A., & Hashim, M. (2024). ASTER data processing and fusion for alteration minerals and silicification detection: Implications for cupriferous mineralization exploration in the western Anti-Atlas, Morocco. *Artificial Intelligence in Geosciences*, *5*, 100077.

Hunt, G. R. (1976). Visible and near-infrared spectra of minerals and rocks: XI. Sedimentary rocks. *Modern Geology*, *5*, 211–217.

Johnson, P., Andresen, A., Collins, A., Fowler, A., Fritz, H., Ghebreab, W., Kusky, T., & Stern, R. (2011). Late Cryogenian–Ediacaran history of the Arabian–Nubian Shield: A review of depositional, plutonic, structural, and tectonic events in the closing stages of the northern East African Orogen. *Journal of African Earth Sciences*, *61*(3), 167–232.

Josso, P., Hall, A., Williams, C., Le Bas, T., Lusty, P., & Murton, B. (2023). Application of random-forest machine learning algorithm for mineral predictive mapping of Fe-Mn crusts in the World Ocean. *Ore Geology Reviews*, 105671.

Khaleghi, M., Ranjbar, H., Abedini, A., & Calagari, A. A. (2020). SYNERGETIC USE OF THE SENTINEL-2, ASTER, AND LANDSAT-8 DATA FOR HYDROTHERMAL ALTERATION AND IRON OXIDE MINERALS MAPPING IN A MINE SCALE. *Acta Geodynamica et Geomaterialia*, *17*(3).

Lachaud, A., Adam, M., & Mišković, I. (2023). Comparative study of random forest and support vector machine algorithms in mineral prospectivity mapping with limited training data. *Minerals*, *13*(8), 1073.

Li, C., Shen, P., Zhao, Y., Li, P., Zhang, L., & Pan, H. (2022). Mineral chemistry of chlorite in different geologic environments and its implications for porphyry Cu±Au±Mo deposits. *Ore Geology Reviews*, *149*, 105112.

Loughlin, W. (1991). Principal component analysis for alteration mapping. *Photogrammetric Engineering and Remote Sensing*, *57*(9), 1163–1169.

Louis, J. (2016). Sentinel 2 MSI–Level 2A Product Definition. *Eur. Sp. Agency*, *49*.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259.

Pour, A. B., Hashim, M., Hong, J. K., & Park, Y. (2019). Lithological and alteration mineral mapping in poorly exposed lithologies using Landsat-8 and ASTER satellite data: North-eastern Graham Land, Antarctic Peninsula. *Ore Geology Reviews*, *108*, 112–133.

Prado, E. M. G., de Souza Filho, C. R., Carranza, E. J. M., & Motta, J. G. (2020). Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: Dealing with imbalanced training data. *Ore Geology Reviews*, *124*, 103611.

Rochim, A., Widyaningrum, K., & Eridani, D. (n.d.). Comparison of Kernels Function between of Linear. *Radial Base and Polynomial of Support Vector Machine Method Towards COVID-19 Sentiment Analysis*.

Shanks III, W. P. (n.d.). *11. Hydrothermal Alteration*.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325.

Shirmard, H., Farahbakhsh, E., Beiranvand Pour, A., Muslim, A. M., Müller, R. D., & Chandra, R. (2020). Integration of selective dimensionality reduction techniques for mineral exploration using ASTER satellite data. *Remote Sensing*, *12*(8), 1261.

Sillitoe, R. H. (2010). Porphyry copper systems. *Economic Geology*, *105*(1), 3–41.

Teklay, M. (1997). *Petrology, geochemistry and geochronology of Neoproterozoic magmatic arc rocks from Eritrea: Implications for crustal evolution in the southern Nubian Shield*. Department of Mines-Ministry of Energy Mines and Water Resources-State of ….

Van der Meer, F., Van der Werff, H., & Van Ruitenbeek, F. (2014). Potential of ESA's Sentinel-2 for geological applications. *Remote Sensing of Environment*, *148*, 124–133.

Werff, H. V. der, & Meer, F. V. der. (2015). Sentinel-2 for mapping iron absorption feature parameters. *Remote Sensing*, *7*(10), 12635–12653.

Woldehaimanot, B., & Behrmann, J. (1995). A study of metabasite and metagranite chemistry in the Adola region (south Ethiopia): Implications for the evolution of the East African orogen. *Journal of African Earth Sciences*, *21*(3), 459–476.

Yajima, T., Yamamoto, K., Yamamoto, K., & Hayashi, T. (2007). Identification of hydrothermal alteration zones for exploration of porphyry copper deposits using ASTER data. *Journal of The Remote Sensing Society of Japan*, *27*(2), 117–128.

Zerai, F. T., Gorsevski, P. V., Panter, K. S., Farver, J., & Tangestani, M. H. (2023). Integration of ASTER and soil survey data by principal components analysis and one-class support vector machine for mineral prospectivity mapping in Kerkasha, Southwestern Eritrea. *Natural Resources Research*, *32*(6), 2463–2493.

Zhongyang Wang. (n.d.). *Model Selection and Boosting*. ICDT. https://icandothese.com/docs/tech/machine_learning/all_in_one/09_model_selection_and_boosting/intro/

Appendix 1 - PCA reports

A. Iron oxide

```
#                    COVARIANCE MATRIX

#    Layer             1             2
#  ---------------------------------------------------------------------------
         1            0.03020       0.01841
         2            0.01841       0.02815
#  ===========================================================================


#                    CORRELATION MATRIX

#    Layer             1             2
#  ---------------------------------------------------------------------------
         1            1.00000       0.63155
         2            0.63155       1.00000
#  ===========================================================================


#                 EIGENVALUES AND EIGENVECTORS

# Number of Input Layers     Number of Principal Component Layers
            2                                 2
# PC Layer             1             2
#  ---------------------------------------------------------------------------
# Eigenvalues
                     0.04762       0.01073
# Eigenvectors
# Input Layer
         1           0.72649       0.68717
         2           0.68717      -0.72649
#  ===========================================================================

#                 PERCENT AND ACCUMULATIVE EIGENVALUES

# PC Layer   EigenValue   Percent of EigenValues   Accumulative of EigenValues
         1       0.04762          81.6068               81.6068
         2       0.01073          18.3932              100.0000
#  ===========================================================================
```

Appendix 1.A, PCA report of Iron oxide- Layer 1 and 2 represent iron oxide and ferric oxide respectively

B. Argillic

```
#                    COVARIANCE MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
        1       4.087062e-03  1.816823e-03 -1.143107e-05
        2       1.816823e-03  3.048838e-03  2.145304e-03
        3      -1.143107e-05  2.145304e-03  2.421909e-03
#  ========================================================================


#                    CORRELATION MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
        1            1.00000       0.51468      -0.00363
        2            0.51468       1.00000       0.78948
        3           -0.00363       0.78948       1.00000
#  ========================================================================


#                 EIGENVALUES AND EIGENVECTORS

# Number of Input Layers     Number of Principal Component Layers
            3                              3
# PC Layer             1             2             3
#  ------------------------------------------------------------------------
# Eigenvalues
                     0.00602       0.00337       0.00016
# Eigenvectors
# Input Layer
        1            0.62702      -0.71042       0.31962
        2            0.67004       0.28254      -0.68646
        3            0.39736       0.64458       0.65316
#  ========================================================================

#                 PERCENT AND ACCUMULATIVE EIGENVALUES

# PC Layer   EigenValue   Percent of EigenValues   Accumulative of EigenValues
        1       0.00602          62.9986               62.9986
        2       0.00337          35.3100               98.3086
        3       0.00016           1.6914              100.0000
#  ========================================================================
```

Appendix 2.B, PCA report of Argillic - Layer 1 ,2, 3 represent Clay, Kaolinite and Alunite respectively

C. Phyllic

```
#                    COVARIANCE MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
         1            0.00128       0.00062       0.00190
         2            0.00062       0.00149       0.00211
         3            0.00190       0.00211       0.00401
#  ========================================================================


#                    CORRELATION MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
         1            1.00000       0.45084       0.83919
         2            0.45084       1.00000       0.86378
         3            0.83919       0.86378       1.00000
#  ========================================================================


#                 EIGENVALUES AND EIGENVECTORS

# Number of Input Layers     Number of Principal Component Layers
            3                                3
# PC Layer             1             2             3
#  ------------------------------------------------------------------------
# Eigenvalues
                     0.00603       0.00076       0.00000
# Eigenvectors
# Input Layer
         1           0.38348       0.72084      -0.57735
         2           0.43253      -0.69252      -0.57735
         3           0.81601       0.02832       0.57735
#  ========================================================================

#                 PERCENT AND ACCUMULATIVE EIGENVALUES

# PC Layer   EigenValue   Percent of EigenValues   Accumulative of EigenValues
         1       0.00603          88.8573               88.8573
         2       0.00076          11.1427              100.0000
         3       0.00000           0.0000              100.0000
#  ========================================================================
```

Appendix 3.C, PCA report of Phyllic - Layer 1 ,2, 3 represent Sericite, Muscovite and Ser.-Musc.-Ill.-Smec respectively

D. Propylitic

```
#                    COVARIANCE MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
         1            0.00227       0.00465       0.00300
         2            0.00465       0.01068       0.00755
         3            0.00300       0.00755       0.00680
#  ========================================================================


#                    CORRELATION MATRIX

#    Layer             1             2             3
#  ------------------------------------------------------------------------
         1            1.00000       0.94510       0.76354
         2            0.94510       1.00000       0.88675
         3            0.76354       0.88675       1.00000
#  ========================================================================


#                 EIGENVALUES AND EIGENVECTORS

# Number of Input Layers     Number of Principal Component Layers
            3                              3
# PC Layer             1             2             3
#  ------------------------------------------------------------------------
# Eigenvalues
                     0.01842       0.00119       0.00013
# Eigenvectors
# Input Layer
         1           0.32311      -0.46265       0.82557
         2           0.75323      -0.40240      -0.52031
         3           0.57292       0.78996       0.21846
#  ========================================================================

#                 PERCENT AND ACCUMULATIVE EIGENVALUES

# PC Layer   EigenValue   Percent of EigenValues   Accumulative of EigenValues
         1       0.01842          93.3001               93.3001
         2       0.00119           6.0456               99.3458
         3       0.00013           0.6542              100.0000
#  ========================================================================
```

Appendix 4.D, PCA report of Propylitic - Layer 1 ,2, 3 represent Epi.-Chl.-Amp, MgOH Amp. and Carb.-Chl.-Epi. respectively