Advancing Lithological, Hydrothermal Alteration, and Artisanal Mining Mapping Using Multi-Sensor Remote Sensing and Machine Learning in the Red Sea Hills, Northeast Sudan

Thesis for the Degree of Doctor of Philosophy (PhD)

Abdelmajeed Adam Elrasheed Ali

Supervisor: Prof. Dr. Szabó Szilárd

UNIVERSITY OF DEBRECEN
Doctoral Council for Natural Sciences and Engineering
Doctoral School of Earth Sciences
Debrecen, 2025

Hereby I declare that I prepared this thesis within the Doctoral Council for Natural Sciences and Engineering, Doctoral School of Earth Sciences, University of Debrecen in order to obtain a PhD Degree in Natural Sciences Debrecen University.
The results published in the thesis are not reported in any other PhD theses.

Debrecen, , , 2025

…….
signature of the candidate

Hereby I confirm that Abdelmajeed Adam Elrasheed Ali candidate conducted his studies with my supervision within the Physical Geography and Geoinformatics Doctoral Program of the Doctoral School of Earth Sciences between 2021 and 2025 The independent studies and research work of the candidate significantly contributed to the results published in the thesis.
I also declare that the results published in the thesis are not reported in any other theses.
I support the acceptance of the thesis.

Debrecen, , , 2025 ……
signature of the supervisor

Advancing Lithological, Hydrothermal Alteration, and Artisanal Mining Mapping Using Multi-Sensor Remote Sensing and Machine Learning in the Red Sea Hills, Northeast Sudan

Dissertation submitted in partial fulfillment of the requirements for the doctoral (PhD) degree in Earth Sciences

Written by Abdelmajeed Adam Elrasheed Ali certified by degree of Master of Economic Geology

Prepared in the framework of the Doctoral School of Earth Sciences of the University of Debrecen, Physical Geography and Geoinformatics Doctoral Program

Dissertation advisor: Prof. Dr. Szabó Szilárd

The official opponents of the dissertation:

Dr. …….. ……

Dr. …….. ……

The evaluation committee:

Chairperson: Dr. …….. ……

members: Dr. …….. ……

Dr. …….. ……

Dr. …….. ……

Dr. …….. ……

The date of the dissertation defence:……..20….

DEDICATION

To my **father**, whose noble spirit continues to guide me from beyond this world your memory is a light that never fades.

To the everlasting souls of my beloved **sister** and **brother**, whose presence I feel in every step of my journey may your love and strength forever dwell in my heart.

To the **children**, **families**, and **innocent people** of my nation who have lost their lives in our current catastrophe may your suffering never be forgotten, and may your courage and resilience continue to inspire hope, justice, and peace for generations to come.

CONTENTS

MOST IMPORTANT ACRONYM AND ABBREVIATIONS

ANS	Arabian Nubian Shield
RSH	Red Sea Hills Region
MLAs	ML Algorithms
CC	Color Composite
BR	Band Ratio
RBD	Relative Absorption Band Depth
DPCA	Directed Principal Component Analysis
CEM	Constrained Energy Minimization Method
PS	pseudo-labelling
ASM	Artisenal and Small-scale Mining
SVM	Support Vector Machine
RF	Random Forest
XGB	The extreme gradient boosting
RSH	Red Sea Hills
FCC	False Color Composite
TCC	True Color Composite
PA	Producer's Accuracy
UA	User's Accuracy
OA	Overall Accuracy
VNIR	visible and near-infrared
SWIR	Short-Wave Infrared
SDGs	United Nations Sustainable Development Goals

VMS	Volcanogenic Vassive Sulphide
MARS	Multiple Adaptive Regression Splines

Chapter 1: Introduction

Mining continues to be important for the growth of society because it provides the raw materials necessary for the development of infrastructure, transformation of the economy, and advances in technology. The extracting of minerals has facilitated the development of civilization from earlier metallurgical discoveries to the technological revolution and present digital businesses. As a source of crucial materials, including lithium, cobalt, copper, and rare earth elements, which are required for electronic devices, clean energy, and industrial facilities, the mining process has remained an essential component of the global industry today. In numerous regions, notably rural and largely undeveloped ones, mining revenue supports growth and development, healthcare, and educational opportunities. Despite its importance, the mining industry, particularly, Artisanal and Small-Scale Mining (ASM), presents serious socio-environmental problems. ASM tends to operate informally, causing unsafe working conditions, habitat destruction, water contamination, forest loss, and greenhouse gases emissions. These effects become more pronouced in regions with less regulatory control, where vulnerable populations are particularly affected financially and medically. Therefore, the major challenge is balancing the economic necessity of mining with the value of social justice and environmental sustainability.

In modern mineral exploration, advanced technologies, including high-resolution remote sensing, hyperspectral and multispectral imagery, machine learning algorithms (MLAs), and geospatial analytics, have been increasingly used. These tools are crucial for mapping complex lithologies, identifying hydrothermal alteration zones, and tracking ASM activities. Their use is more relevant in remote areas where the geology is cropped out, underexplored, and complex terrains such as the Red Sea Hills (RSH). Accordingly, the objectives of this dissertation are: (1) Examine how machine learning (ML) might improve lithological mapping in comparison to more conventional techniques; (2) Establish updated lithological maps of selected areas in the RSH; (3) Map hydrothermal alteration zones linked to gold mineralisation; (4) Identify

and track the growth and environmental impact of ASM. Moreover, we intended to test the consequences of using training data collected through different approaches and introduce a simple method for extending the available input training data by applying region-growing segmentation to augment the inputs. Furthermore, no direct comparison has been made between spatially explicit area-based accuracy assessment (with full coverage of possible occurrences) and traditional point-based approaches. In this context, and under varying training data sizes, our aims were to (i) determine whether a limited amount of reference data is sufficient for reliable ASM identification, (ii) evaluate the effects of training data sizes on classification performance, (iii) assess the potential of hyperspectral (PRISMA) and multispectral (Landsat) imagery for ASM mapping, (iv) evaluate the efficiency of a developed seeded region-growing algorithm, and (v) determine area-based accuracy and compare it with the traditional point-based technique.

Three study sites within the RSH were selected for this investigation. Each method assesses the methodological robustness in actual mineral exploration environments using combined remote sensing datasets, field surveys, petrographic analysis, and ML. The RSH, part of the Arabian-Nubian Shield (ANS), best illustrates this paradox. The area has a long history of mining, both formal and informal, in the past and present, and is rich in gold, base metals, critical minerals, and other minerals. ASM expansion has been expedited, especially in relatively stable states such as the Nile River and Red Sea regions, with the increase in world gold prices, the secession of South Sudan in 2011 (which reduced Sudan's oil revenues), and the ongoing political instability. These dynamics have intensified the pressure on both landscapes and communities. The ANS, particularly the Sudanese part, was chosen because of its geological characteristics and the potential of its mineral resources. This Neoproterozoic shield was formed during the Pan-African orogeny and is composed of accreted island arcs, ophiolites, metamorphic sequences, and granitoids. In comparison to other Precambrian shields around the world, it continues to be overlooked despite its mineral wealth.

Therefore, the RSH is an appropriate natural laboratory for testing cutting-edge lithological mapping, identifying mineral alteration technologies, and evaluating the effects of ASM.

The lithological diversity, orogenic gold systems, and the presence of both formal and informal mining make the RSH of especial importance. As indirect indicators of mineral exploration, ASM sites that are visible in satellite imagery offer surface evidence of subsurface mineralization and serve as indirect indicators for exploration. Furthermore, the arid climate of the area limits vegetation cover, enhances the accuracy of the spectral analysis. Our research discusses the large demand for sustainable mineral exploration in fragile socio-ecological systems, in addition to making a methodological contribution. Through the integration of ML, remote sensing, and field and petrographic data, this study offers scalable tools for mapping lithology and mineral deposits, identifying ASM, and understanding its landscape impacts.

Supporting Sudan's economic development serves as the second, but equally important, motivation for this research. Mining and gold exports are essential sources of income and contribute to the country's recovery and growth. Formalizing the industry, enforcing law, reducing environmental damage, and aligning mining practices with national and international development goals, such as the African Mining Vision and the United Nations Sustainable Development Goals (SDGs), basically depend on the methodological mapping of new gold occurrences and monitoring ASM.

Guided by insights from the initial literature analysis, I formulated the following research assumptions to guide this study's direction and analytical framework:

- The combining of remote sensing and ML with field data, can notably facilitate lithological mapping compared to traditional (field-based) methods, particularly in complex geological settings such as the Red Sea Hills.
- Enhancing spectral features derived from multispectral and hyperspectral imagery using advanced image enhancement tools and ML can effectively detect hydrothermal alteration zones associated with gold mineralization.

- ASM activities are continuously increasing in the RSH and can be detected and monitored over time using medium- and high-sapatial-resolution satellite data and ML classifiers.
- ASM has a huge effect on the geology and environmental integrity of the RSH, including the disturbance of geological features, which can be detected and assessed through the application of remote sensing.
- Spatial patterns of ASM activity show a statistically significant correlation with the locations of known subsurface mineralization and can therefore be used as a cost-effective surface guide to delineate high-potential zones for undiscovered ore deposits.
- The iterative application of pseudo-labelling significantly improved the model accuracy and spatial coherence for ASM and lithological mapping in data-scarce environments.
- Employing adaptive data augmentation strategies, as opposed to static ones, enhances model accuracy and spatial coherence by simulating a wider and more geologically realistic spectrum of feature variations.

Research Scope and Novelty

This dissertation builds upon existing studies of the ANS and RSH, which focused on regional tectonics, and Landsat-based exploration is expanded upon in this dissertation. However, the use of high-resolution remote sensing, hyperspectral, and ML in lithological mapping and ASM dynamics and impacts was overlooked in these studies. To fill this gap, this dissertation adopts an interdisciplinary approach that considers ASM as a stand-in for subsurface mineralization, in addition to being a threat to ecosystems. Therefore, this study offers several new insights and contributions. In addition, this study introduced adaptive data augmentation methodologies, mainly pseudo-labelling and segmentation-based training data expansion to tackle the prevailing shortage of reference data in geological and ASM studies. For the first time, ASM sites were mapped and tracked in detail using Planet imagery. Furthermore, hyperspectral PRISMA data were used for the first time in applications related to ASM. Additionally, a major breakthrough in

remote-sensing-based mineral prospecting was achieved when gossan zones linked to VMS were successfully detected using PRISMA hyperspectral datasets.

Chapter 2: Literature review

2.1 Lithological mapping

Lithological mapping is the process of identifying and spatially delineating rock units and is fundamental to mineral exploration, geological modelling, groundwater assessment, and environmental planning (Abrams and Yamaguchi, 2019). EL-Omairi and El Garouani (2023) defined lithological mapping as a process of identifying and outlining different rock types and their spatial relationships, providing vital information on the composition, structure, and stratigraphy of the Earth's surface for a particular study area. According to Bucci *et al. (*2022), Lithological maps, are depictions that include details about the various rock types found in a certain area. The basic concept of lithological mapping is fundamentally the same throughout the literature, although geoscientists may use somewhat different terminology, such as "geological unit classification," "rock type mapping" or "bedrock delineation". It's commonly acknowledged that lithological mapping is a crucial technique for understanding subsurface features and guiding informed decision-making in resource management (Sabins, 1999; van der Meer *et al.*, 2012; Shebl *et al.*, 2023). Lithological maps provide essential information for comprehending activities at both the surface and subsurface. Geoscientists can use these maps to better understand subsurface structures, assess mineral potential, direct drilling operations, and support environmental and land-use planning (van der Meer et al., 2012). Moreover, lithological map data are used to construct reliable geological models, particularly in areas with complex tectonic or metamorphic histories (Rowan and Mars, 2003). Therefore, lithological maps have a direct impact on the outcomes of exploration efforts and sustainable resource management plans (Pour et al., 2018). Lithology is an essential component of geohazards. Furthermore, it has a major

impact on hydrological dynamics, such as surface water flow paths and groundwater recharge (Gleeson et al., 2011), and it assists landscape evolution models by providing information on erosion and deposition patterns (Coulthard, 2001). They are used to predict the likelihood of landslides (Alvioli et al., 2021; Sarro et al., 2020) and investigate the effects and amplification of seismic locations (Mori et al., 2020). It is also important to understand the geochemical evolution of river systems and processes such as silicate weathering, which consumes CO_2 from the atmosphere (Gibbs and Kump, 1994; Donnini et al., 2020). Furthermore, lithological data affect flood risk mapping (Vojtek and Vojteková, 2019), groundwater level variations (De Graaf et al., 2017), soil classification, and erosion modelling (Vanmaercke et al., 2021). Likewise, it helps to assess subsurface energy resources, such as geothermal capacity (Moeck, 2014; Sunan Sakti Syah Alam et al., 2019) and hydrocarbon reserves (Han et al., 2018) and even to improve geomorphological categorization systems (Alvioli et al., 2020).

In the past, lithological mapping primarily relied on direct field observations, hand specimen analysis, systematic geological traverses, and interpretation of aerial photographs (Szaniawska, 2018). Although these traditional approaches provide high-resolution data and informative contextual information at localized scales, they are labor-intensive, time-consuming, and often require considerable logistical support (Mahmood et al., 2019). Their effectiveness is further diminished in remote and inaccessible locations and topographically challenging areas, such as deserts, densely forested areas, or high-temperature environments (Drury, 2001; Shebl et al., 2022). Furthermore, large-scale or regular lithological updates are limited by access restrictions and the high cost of field excursions to the sites. Consequently, it is now more crucial than ever to involve modern technologies, such as remote sensing, GIS, and ML to increase the global reach and efficiency of lithological investigations (Shirmard *et al.*, 2022). RS and ML have revolutionized lithological mapping by enabling the identification of surface mineralogy and texture from satellite data. This

information can then be processed using classification algorithms to differentiate between various rock types based on their spectral characteristics (Abdelkareem *et al.*, 2021; Shirmard *et al.*, 2022).

2.2 Artisanal and Small-Scale Mining

Artisanal and small-scale mining (ASM) refers to mining activities conducted by individuals, groups, families, or cooperatives with little or no mechanization, often operating in the informal or even illegal sector of the economy (Hentschel, 2002; Hilson and Maconachie, 2020; OECD, 2017). ASM includes a wide range of extraction processes, such as panning, hard-rock excavation, sluicing, and rudimentary mineral processing techniques, such as amalgamation and manual sorting. Miners engaged in these activities usually target various commodities, such as gold, diamonds, tungsten, tin, cobalt, semi-gemstones, and gemstones (M. M. C. Fritz et al., 2018; Hilson and McQuilken, 2014; Telmer and Veiga, 2009). Among these, gold remains the most commonly extracted mineral, which has led to the term Artisanal and Small-Scale Gold Mining (ASGM) becoming dominant in the literature and policy discourse(Telmer and Veiga, 2009; UNEP, 2012).

ASM is growing rapidly worldwide, particularly in developing countries (Hentschel, 2002). Despite its informal nature, ASM contributes significantly to global mineral production, accounting for approximately 20% of the world's gold and up to 25% of tantalum, tin, and tungsten (ICMM, 2021; UNEP, 2012). Therefore, it provides livelihoods for an estimated 40–45 million people worldwide, with over 150 million more indirectly dependent on the sector (OECD, 2017). ASM plays a critical socio-economic role in developing regions such as sub-Saharan Africa, Latin America, and Southeast Asia. For example, in Ghana, the Democratic Republic of Congo (DRC), and Burkina Faso, ASM contributes between 15% and 25% of national mineral exports (Hentschel, 2002; Hilson and McQuilken, 2014). It provides an essential source of income for rural populations, particularly where formal employment opportunities are scarce because of conflict, economic

instability, or structural poverty. The World Bank's 1987 report on Small-Scale Mining: A Review of the Issues (Noetstaller, 1987) significantly changed in the way that ASM was presented within the rhetoric of international development. By emphasizing how labor-intensive small-scale businesses are in comparison to large firms, it highlights the sector's potential to create jobs in rural, job-scarce parts of sub-Saharan Africa. Using business-oriented terminology, such as "entrepreneurship" and "business development," the report placed ASM within the larger framework of small firm development. This reflects a policy mindset that sees ASM largely as a means of facilitating quick, market-driven economic engagement. Throughout the 1990s, the World Bank's efforts to reform the mining industry were greatly influenced by this prespective. ASM is becoming increasingly important to the economy, although it is still not a top priority for development in sub-Saharan Africa. Hilson and McQuilken. (2014) conducted a comprehensive review study involving four decades of ASM support in sub-Saharan Africa. Their paper highlighted that ASM is an important way for people to make a living and a big part of the rural economy, but it is still not included in larger development plans or measures to reduce poverty in the DRC. It concludes that support interventions are likely to remain ineffective and poorly coordinated if ASM is not included in official regional development planning. ASM poses significant environmental risks and is a global concern. According to Telmer and Veiga, (2009), mercury emissions from ASM regions because of ineffective amalgamation methods are considered a major source of mercury pollution, with estimates of yearly outputs ranging from 800 to 1000 tonnes. According to the study, despite efforts to control its usage, substantial amounts of mercury are still lawfully imported into numerous nations, frequently from the European Union, which ironically still produces virgin mercury for export. The authors underlined that mercury's low price in relation to gold and the absence of global regulatory control continue to motivate its usage. Moreover, in mining communities, this arrangement presents significant hazards to the

environment and public health, underscoring the need for more robust political action to regulate mercury trade and lessen its effects. Given these dangers, the 2013 Minamata Convention on Mercury, a global agreement, specifically named ASM as the leading cause of anthropogenic mercury emissions and called on signatory nations to create National Action Plans to minimize and, if possible, completely eradicate mercury use in the industry (UNEP, 2012).

Numerous reports from the United Nations Industrial Development Organisation (UNIDO, 2006b) and the World Health Organisation (WHO, 2016) have detailed the serious environmental health risks associated with ASM, especially when it comes to the use of dangerous chemicals like cyanide and mercury. These groups draw attention to the fact that uncontrolled mining operations frequently result in mercury pollution of soil, water, and air, directly endangering miners and the nearby populations. Furthermore, because of their extended and unprotected exposure to harmful substances, women and children who are frequently active participants in ASM are disproportionately affected (Gibb and O'Leary, 2014). In remote mining areas, inadequate public health infrastructure, restricted access to protective gear, and lax regulatory enforcement contribute to these health issues.

The environmental impacts of ASM vary considerably depending on the geographical and ecological context in which it develops. ASM has unique effects in arid and semi-arid areas, such as the Red Sea Hills of Sudan. These include extreme land degradation, soil compaction, and the destruction of important geological features, such as dips, faults, and ore body structures, which are essential for comprehending regional tectonics and mineralization patterns (Adam et al., 2024). These activities often occur in fragile landscapes where vegetation is sparse and water is limited, making the ecological recovery of mined sites particularly difficult. Furthermore, contamination from the use of dangerous chemicals such as cyanide and mercury, occurs in both contexts, whereas the persistence and spread of contaminants vary

because of hydrological and climatic conditions. Therefore effective mitigation techniques must be customized to each region's unique geological and environmental features. Beyond environmental health concerns, ASM has significant social repercussions that have garnered increasing attention. According to Dorner (2012), ASM can both reduce and deepen poverty by providing opportunities for short-term income while promoting long-term economic instability caused by unregulated labor practices, volatile commodity prices, and restricted access to markets and finance. Schwartz et al. (2021) examined the dichotomy of ASM as a livelihood strategy and a cause of social instability. In resource-rich but institutionally weak situations, ASM frequently contributes to land disputes, rural-urban migration, and youth disempowerment. Furthermore, Hilson and Maconachie (2020) contend that miners are unable to participate in sustainable development projects or obtain formal legal protection due to the informal and usually unlawful nature of ASM operations.

ASM in Sudan has deep historical roots, with evidence of gold extraction dating back to ancient civilizations, including the Nubians, Egyptians, and later colonial powers, such as the Turks and the British (Zeinelabdein and Nadi, 2014). However, since 2011, the scale and distribution of ASGM surged dramatically due to the economic crisis caused by South Sudan's secession, which caused 80% of Sudan's oil revenue to be lost, and the sharp rise in global gold prices, which rose from $14,000 USD/kg in 2006 to over $62,000 USD/kg after 2020 (Price, 2024). ASM has now emerged as a vital livelihood strategy in areas with inadequate infrastructure and a lack of official employment, such as Darfur, Kordofan, the Blue Nile, River Nile State, and the Red Sea Hills, (Ahmed El Tohami, 2018; Elwaleed et al., 2024). With over one million miners directly involved and more than two million people supported indirectly the ASM sector plays a vital socio-economic role, particularly in remote rural areas (“Minamata Convention,” 2015; Seccatore et al., 2014). ASM obliterates key geological features such as fault lines, mineralized zones, and structural outcrops necessary for mineral

exploration, disrupts delicate desert ecosystems, and causes extensive land degradation, in contrast to tropical mining zones, where the main effect is forest clearing (Fadlallah et al., 2020; Adam *et al.*, 2024). Additionally, informal ASM has surpassed industrial production, despite the existence of legal industrial mining, such as in the Hassai mine in the Red Sea Hills (Ahmed et al., 2019; Mhangara et al., 2020). Despite this significance, ASM remains largely informal and unregulated, and its environmental and geological impacts are poorly documented, mostly because of a lack of rigorous monitoring. Therefore, the environmental effects and spatial impact of this increase are still poorly understood.

2.3 Remote Sensing and Machine Learning for Lithological Mapping

The integration of remote sensing data with ML has considerably advanced lithological mapping, hydrothermal alteration mapping, and ASM tracking. This combination enhanced the accuracy and efficiency of lithological mapping compared to traditional geological survey practices, especially in remote and inaccessible areas. Numerous studies have underscored the effectiveness of ML-driven remote sensing in traditional lithological mapping practices. Shirmard *et al.* (2022), for instance investigated the latest advances in ML applications for lithological mapping, highlighting innovative methodologies that help in overcoming traditional field mapping challenges and improving the mineral exploration outcomes. From a regional perspective, Shereif et al. (2024) classified granitic rocks in Egypt's Central Eastern Desert using Support Vector Machine (SVM) and dimensionality-reduction techniques, including principal component analysis (PCA) and independent component analysis (ICA). Their findings showed notable improvements in accuracy, confirming the usefulness of ML in specific geological contexts.

The integration of remote sensing and ML, especially using algorithms such as SVM, random forest (RF), and artificial neural networks (ANN), has also improved mapping accuracy in mineral

exploration, water management, and environmental assessment, particularly in semi-arid and resource-challenged areas, as shown by Amusuk *et al.* (2024) and EL-Omairi and El Garouani (2023). These studies underscore the advantages of hybrid approaches that use information from multiple sources, such as Sentinel-2, ASTER, Landsat, and Hyperion. These combinations make mapping faster, cheaper, and more accurate than traditional approaches. Previous studies have explored the use of machine learning for lithological mapping in densely vegetated regions, highlighting its capacity to extract lithological information despite canopy interference. They deployed algorithms including MD, CART, SVM, RF, and GBDT to Sentinel-2 and SRTM DEM data, applying algorithms including MD, CART, SVM, RF, and GBDT. The GBDT_Sen+DEM model delivered the highest overall accuracy of (63.18%) because of the combination of spectral and terrain data. Similarly, Nugroho *et al.* (2023) improved lithological mapping in Komopa, Indonesia, by combining remote sensing and airborne geophysical data with an RF classifier.

The potential of multi-source data integration was demonstrated through this fusion approach, which adopted magnetic, electromagnetic, and radiometric data to enhance the classification accuracy by up to 13% in the F1 score compared to remote sensing alone. The benefits of using Landsat 8 OLI with ALOS/PALSAR and ASTER data were validated by Bachri *et al.* (2019) and Ali *et al.* (2024). They enhanced lithological discrimination in arid and challenging terrains and refined geological maps by achieving classification accuracies of 75-85% using the SVM classifier. Elrasheed and Szabó (2024) and Shebl *et al.* (2023) studied the value of well-known ML, RF, XGB, and SVM for lithological mapping purposes and introduced PRISMA Hyperspectral lithological mapping using ML. In establishing comprehensive lithological maps for their research, they validated the usefulness of PRISMA through fieldwork. The ML classifiers used revealed a high level of accuracy. As the majority of studies highly recommend integration, the objective of ML with multispectral or hyperspectral remote sensing data is to achieve

more accurate lithological mapping, notably prior to mineral exploration programs. For more accurate lithological mapping, most studies highly recommend integrating the objectivity of ML with multispectral and hyperspectral remote sensing, especially prior to mineral exploitation operations in comparable geological settings. Consequently, we considered incorporating these methods into our study because they provide a revolutionary avenue for contemporary geological investigations. Geological features can be identified using extensive of remote sensing datasets (El-Omairi and Garouani, 2023). Valuable insights can be drawn from the geographic data obtained by sensors based on the distinctive properties of the local area (Nair et al., 2023). ML offers an efficient method for processing remote sensing data. After the training phase, algorithms can identify rock types, faults, or mineral deposits if the spectral resolution of the images is sufficient. These algorithms can handle large datasets, allowing the discovery of plausible patterns in complex geological data (Han et al., 2023). However, ML requires high-quality training data to build reliable models. Additionally, independent testing data (known as reference data) are essential for validating the performance of a model (James et al., 2013).

ML provides an important method for generating geological information by processing remote sensing data, although its validity relieses heavly on the quantity and quality of the reference data (Foody, 2009; Maxwell et al., 2018). Reference data is the most crucial constituent of all models. Training subsets were used for model building, whereas testing subsets were used to assess accuracy. In remote sensing, we may have millions of pixels (i.e., data points), which may give the impression that the delineation of reference data is easy; however, this is not the case for all tasks. For example, in land cover mapping, the traditional approach is to classify surface objects into simple classes, such as forests, grasslands, and water bodies. These classes can include thousands of pixels as reference data because simple visual interpretations of images can provide sufficient information. However, when the aim is a more specific problem, such as identifying plant or tree

species, detecting plant diseases, or classifying roof types, reference data collection requires field observations and/or ground measurements, which makes this step labor intensive and time consuming. Lithological mapping faces the same problem: field observations are essential for a reliable reference dataset. If the number of labelled instances is insufficient, the model cannot be adequately trained because of the small size of the reference dataset.

The amount of reference data required depends on the algorithms used and has been widely discussed in the literature. Although there are basic rules, generally, the more reference data, the better. Studies have shown that accurate, balanced, and large training datasets are often more important than the choice of algorithms in terms of output (Collins et al., 2020; Li et al., 2014; Maxwell et al., 2018). There is no universally accepted minimum number of training data points. However, Foody (2009) defined a method for calculating the minimum amount of testing data based on the desired overall accuracy, acceptable standard and classification errors. For example, for four categories with an 85% target accuracy and a standard error of 0.02 (2%), at least 1,275 testing data samples are required, which is approximately 320 per class. If the target features are limited owing to unique characteristics (e.g., rare species, specific roofing types, or uncommon rock types), collecting a sufficient number of data points may be unfeasible. Even with fewer reference data than the optimal, accurate outcomes are still possible depending on the data distribution and representativeness of the data which introduces a higher risk of uncertain results.

An alternative method, known as self-training or pseudo-labeling (PL), involves selecting high-probability data from the predictions of an initial model conducted with fewer data points, and model training is repeated with data from a classified map with the highest probability. The effectiveness of this method has been proven to improve classification results, including soil class mapping (Zhang et al., 2021), detection of geochemical anomalies (Chen et al., 2023), prediction of invasive species

distribution (Cruz et al., 2023; Kim et al., 2024), and identification of sporadically distributed species (Likó et al., 2024). However, PL does not always improve classification accuracy or result in worse predictions. Although PL appears to be a good solution for overcoming the issue of limited reference data, users cannot assume that the outcome will be better than that of the model output based on the original data. Geological mapping presents a unique challenge because both natural and anthropogenic processes (e.g., physical and chemical weathering and mining) can alter rock characteristics. However, the spatial pattern can easily be validated by visual interpretation (e.g., a given rock type can occur at a given location). To the best of our knowledge, this is the first study to focus on the effectiveness of PL in geological mapping. We aimed to determine the effectiveness of PL in geological mapping. The selected study site was in an arid environment, where the lack of vegetation made it possible to observe the rock types in satellite imagery. We had the following questions related to the probabilities and PL: (i) what was the probability of the related rock type and how did it change with the increased training data; (ii) what was the standard deviation (SD) of classes in the changed areas, and how large were the areas influenced; and (iii) what was the direction of the changes in terms of probability? Accordingly, we formulated the following hypotheses: (i) high probability values ensure high model performance, and (ii) pseudo-labeling improves map quality by providing additional training data for the modelling process.

2.4 Remote Sensing for Artisanal and Small-Scale Mining Detection and Monitoring

Remote sensing technology has been helpful in mapping and monitoring artisanal and ASM because it offers reliable, consistent, and wide-area observations. Several studies have used multispectral and hyperspectral imaging (Dorner, 2012; Hinton, 2003b; Schwartz et al., 2021; Telmer and Veiga, 2009) to identify the spectral signatures of mining sites and evaluate environmental deterioration, such as land

degradation, vegetation loss, and water pollution. A review article by Nursamsi *et al.* (2024) investigated scalable remote sensing methods for mapping ASM. The study highlights that although local ASM mapping has been successful, regional differences in research, context-dependent method selection, and the necessity for innovative data combinations and community engagement to achieve precise and ongoing regional monitoring make it challenging to scale these techniques globally. This study draws several conclusions regarding the remote sensing of ASM. Most notably, the local context of ASM activities, the intended scale of analysis, and real-world resource limitations, such as the availability of validation data and funding for high-resolution imagery, all have an essential impact on the selection of remote sensing methods used for mapping ASM.

Moomen *et al.* (2022) discussed opportunities for applying remote sensing technologies to the ongoing global concerns about ASM activity. By combining ML and deep learning (DL) techniques, data fusion (optical, radar, UAV, lidar, in-situ, and crowd-sourced data), time-series analysis, and analysis-ready data structured in data cubes, this study aimed to enhance ASM land use/land cover methodologies. According to the authors, remote sensing technologies provide an effective and reliable means of tracking ASM activities in developing nations, providing important data for policymaking, public health interventions, and environmental management. Similarly, Alessi et al. (2023) enhanced ASM detection and monitoring by using Synthetic Aperture Radar (SAR) to enable semi-automated identification of dredges with minimal false positives, improving regulation and data collection. Based on their findings, remote sensing is a useful way to improve data collection, regulation, and monitoring of the more mobile, growing, and elusive ASM activities and the damage they cause. Remote sensing has been discussed in a number of studies in densely vegetated areas, where soil exposure or deforestation driven by mining provides a strong visual signal (Simionato et al., 2021; Ngom *et al.*, 2023). In arid and lithologically diverse environments such as the Red Sea Hills, remote

sensing applications pose additional challenges due to complex geological backgrounds and subtle spectral variability. These characteristics emphasize the necessity of sensor evaluation and region-specific adaptation for ASM detection.

ML, such as RF, and SVM have also been used to enhance ASM detection capabilities in challenging environments, especially in vegetated or tropical environments, where mining footprints stand out against the background (Nursamsi et al., 2024). Amri et al. (2023) employed multi-sensor data (Sentinel-1 and Sentinel-2) and a supervised random process on cloud computing platforms to detect and map ASM footprints in the Kuantan Singingi, Riau region. This study successfully mapped ASM footprints covering an estimated area of 10,044.38 hectares, with an accuracy of 89.23%. Nyamekye *et al.* (2021) investigated the performance of individual Sentinel-2 bands using four distinct ML and deep learning models, including ANN, RF, SVM, and CNN, in order to analyse changes within ASM landscapes in Ghana in 2017. The research outcomes showed that band 5 Vegetation Red Edge (VRE) 1 was the most effective in classifying ASM, and ANN was determined to be the superior classifier for ASM. Furthermore, despite the restriction on ASM expansion, the classification results revealed that the majority of farmlands and peripheral areas of forests were converted to ASM.

In order to assess the consequences of illicit ASM (Galamsey) on vegetation in Ghana's high forest zone, Kwang and Afele (2024) explored various ML classifiers. The study examined different algorithms, namely, SVM, RF, ANN and Maximum Likelihood Classification (MLC), to identify the most effective model for detecting vegetation cover changes driven by Galamsey activities. Their findings demonstrated that RF consistently outperformed the other classifiers with an overall accuracy of 88%, and the Kappa Coefficient values of 0.84 emerged as the top performer. Despite the fact that previous studies have provided significant information in the field by integrating remote

sensing and ML in ASM, there are still substantial gaps in our understanding of ASM expansion, its socioeconomics, and its environmental impact. The impact of ASM on geological integrity and its unintentional exposure to hydrothermal alteration zones features which can contain important hints for further mineral exploration, are also notably understudied. These gaps must be filled for more sustainable mining governance systems that incorporate geological, environmental, and socioeconomic factors. In this study, we adopted ML with a very high spatial resolution, hyperspectral remote sensing data, and multispectral data to address these challenges. Although the combination of ML and remote sensing offers an efficient tool for identifying ASM sites, reliable reference data remain a key question. Wadi deposits often resemble ASM features, making field verification the best practice for accurate site identification. However, field observations are not feasible in remote regions or countries where onsite observations involve high risks. Consequently, the number of verified locations is usually limited, creating the need to extend field observations with visually interpreted reference points to meet the minimum requirements for classification algorithms.

The amount and quality of reference data are crucial factors that influence the accuracy of classification results in remote sensing applications. Inadequate training samples often lead to biased predictors and insufficient standardization in complex scenarios (Foody 2009; Maxwell et al. 2018). Pal and Mather (2005) and Ghosh *et al.* (2014) reported that large and representative training datasets improved classification efficiency. For example, Pelletier *et al.* (2017) showed that when applying RF for land cover classification, a larger reference dataset increased the classifier accuracy, although this effect decreased after a certain point. Similarly, Belgiu and Drăguţ (2016) pointed out that although ensemble methods, such as RF, can tolerate limited training data, they still require a minimum dataset size to avoid misclassification and overfitting. In the case of high-dimensional data, such as hyperspectral imagery, training data scarcity can notably affect classifier

performance owing to the Hughes phenomenon (Camps-Valls and Bruzzone, 2005; Hughes, 1968). According to Al-Mufti (2022), the adaptability of algorithms such as SVM and RF is strongly influenced by spectral resolution, and even with limited samples, advanced sensors can yield reliable results if reference data are carefully constructed. Likewise, Belgiu and Drăguţ (2016) and Volpi *et al.* (2011) demonstrated that active learning methods and semi-supervised learning approaches mitigate reference data scarcity by refining sample selection. Because ground truthing is often inaccessible, these findings are highly relevant for ASM mapping and monitoring in unstable or remote regions (Werner et al., 2019). Consequently, to develop practical, flexible, and robust ASM mapping methodologies, it is essential to understand how variations in the training sample size affect the classification outcomes.

Because ASM areas have a limited spatial extent and can be characterized by well-defined boundaries, accuracy assessments can be based not only on data points but also on their real extent. This principle is also applicable to all distinct phenomena with definite borders, such as buildings, rooftops, or land-use classes. This raises an important methodological question regarding accuracy: if the points reflect accurate results, how well do they reflect the accuracy metrics of the true spatial distribution. In a binary approach, the true and false positives of the ASM areas can serve as more informative indicators of the model performance. Area-based accuracy assessments are commonly employed in deep learning and semantic segmentation studies, often using the Jaccard index (Maxwell et al., 2021). However, area-based solutions exist, and many are not explicitly spatial. For example, Friedl et al. (2001) proposed a probabilistic approach tied to per-pixel performance, which provided a conceptual background for confidence estimates. Lewis and Brown *et al.* (2022) used pixel-level fractional class memberships (i.e., statistical area-based method), but their approach did not directly cover polygon-level extent, instead relying on a generalized confusion matrix with proportional error, and the method was designed primarily for sub-pixel classifiers. Zhen et al. (2013) applied an area-

based solution using object-based classification. However, despite being area-based and repeated 30 times, the accuracy was still evaluated using a confusion matrix.

2.5 Remote Sensing and ML of Hydrothermal Alteration and Gossan Mapping

Hydrothermal alteration zones are essential markers of mineralization processes, particularly in gold mineralization and base metals such as VMS deposits. These zones are developed when hydrothermal fluids change the mineralogy and chemistry of host rocks, forming unique mineral assemblages such as iron oxides, clays, and carbonates that are detectable using remote sensing because they have distinctive absorption characteristics in the VNIR and SWIR spectral regions (Rajan Girija and Mayappan, 2019; Sabins, 1999). A number of hydrothermal alterations are commonly recognized in mineral exploration, including argillic, advanced argillic, phyllic, propylitic, potassic, greisen, chloritisation, silicification, and carbonatization (Shim et al., 2021). The spatial distributions of these alteration zones vary, sometimes extending for kilometers around deposits such as iron-oxide-copper-gold, VMS, and porphyry copper deposits. Detecting and mapping these zones is of special significance, particularly in arid, inaccessible, or challenging areas such as the ANS, where traditional fieldwork is constrained by terrain, security, or cost (Zeinelabdein and Nadi, 2014; Gupta, 2025). Shirmard *et al.* (2022) investigated the application and adaptation of popular and recently developed ML methods for remote sensing data processing, including satellite optical and radar, airborne, and drone-based data. They discussed the challenges and potential of ML applications for processing remote sensing data to explore different ore deposits. This study concluded that ML techniques are useful for analyzing a variety of remote sensing data to understand the relationship between the reflectance continuum and features of interest. They validated both the usefulness of remote sensing data in

overcoming field obstacles and the increasing importance of ML for the reliable processing and robust analysis of these data.

The detection and mapping of hydrothermal alteration and gossan zones have been notably enhanced through the combination of remote sensing with ML techniques. This integration enables effective mineral exploration over large areas. Shim *et al.* (2021), for instance, described how high accuracy can be achieved in detecting hydrothermal alteration minerals such as sericite, chlorite, and pyrophyllite using ML and spectral indices. This study used a RF model for classification, achieving an overall accuracy of 97% for mineral content above 5 wt%. Additionally, it selects spectral variables using least absolute shrinkage and selection operator (LASSO) logistic regressions, which yield high overall accuracies (above 93.6%) and R^2 values ranging from 0.57 to 0.71. Using ASTER data to identify alteration zones in Iran's Sar-Cheshmeh mining area. Bahrami *et al.* (2024) assessed many ML algorithms, including RF, SVM, gboost (GB), XGB, and ANN. Their results indicated the efficiency of automated, non-invasive mapping over large areas, showing that feature extraction-free models outperformed the others. Yalcin *et al.* (2020) also used remote sensing techniques for finding hydrothermal alteration minerals spread over large areas, reducing both time and cost. This study specifically used Crosta and Spectral Angle Mapping (SAM) techniques of ASTER satellite data to locate hydrothermal alteration zones. The outcomes demonstrated how alteration zones can be distinguished. Remote sensing and ML offer efficient delineation of hydrothermal zones using spectral techniques such as band ratios (BR) (Labdaoui et al., 2023), principal component analysis (PCA), decorrelation stretch (DCS), and target detection algorithms such as Constrained Energy Minimization (CEM) (Shebl *et al.*, 2023). These techniques work particularly well to identify minerals that are characteristic of kaolinite, muscovite, chlorite, and iron oxides, which typify argillic, phyllic, and propylitic alteration zones (Bahrami *et al.*, 2024; Abdoulatif et al., 2025). These features have been extracted, and mineralized zones have been mapped with high precision using

Landsat, Sentinel-2, ASTER, Hyperion, and more recently, PRISMA hyperspectral imagery, which offers important advantages over lower-resolution multispectral data (Abubakar et al., 2019; Anifadi et al., 2019; Mrinmoy Dhara et al., 2017; Yousefi et al., 2021).

In addition to alteration zone mapping, remote sensing and ML have also been used in geological studies beyond alteration zones for structural analysis (Abdelkareem et al., 2021; Sayed et al., 2023) and the identification of surface manifestations of ore deposits, such as gossans (Abdelsalam et al., 2000; Gahlan and Ghrefat, 2018). In our research area, the ANS gossans are the oxidized caps of VMS deposits. However, in comparison to broad lithological units or regional alteration zones, their spatial imprints are often much smaller and more subtle. Studies using moderate-resolution multispectral data, such as Landsat, have reported promising but limited results, mainly due to the inadequate spectral resolution for isolating key gossan indicators. In light of these studies, technologies such as remote sensing data (e.g., PRISMA, Landsat, Sentine-2) and advanced ML models (RF, SVM, XGB) are necessary. Enhanced, time-and cost-effective alteration zone identification in a remote, topographically rugged large area with diverse geological characteristics, such as our study area, the RSH, and the ANS. Moreover, these techniques are especially helpful in settings where ASM inadvertently reveals hydrothermal alteration zones, enabling remote sensing to detect mineralization hints that would otherwise be obscured by overburden or worn surfaces. Subsequently, in this dissertation, we applied this approach, which has the potential to facilitate early stage prospecting and reduce the need for exhaustive fieldwork. These advances not only enhance the scientific understanding of hydrothermal systems but also support more targeted and environmentally responsible exploration strategies in the field.

Chapter 3: Material and Methods

3.1 Study sites

3.1.1 Study Sites within regional geological context

The Arabian-Nubian Shield (ANS), also known as the East African–Antarctic Orogen (EAAO), is located in the northern part of the East African Orogen (EAO) and represents a fundamental region in the Pan-African Belt, formed during the Neoproterozoic by multifaceted tectonic events (Hamimi, 2021; Hamimi et al., 2014). The evolution of the ANS, which spanned approximately 700 million years from 1200 to 500 Ma, included the successive cratonization of multiple intra-oceanic island arcs. It consisted of sporadic magmatic, metamorphic, and sedimentary processes identical to those found in subduction zones today (Reymer and Schubert, 1984). The closure of the Mozambique Ocean and the pronounced occurrence of arc suturing and terrane accretion throughout northeast Africa are the result of the convergence and collision of East and West Gondwana, which drives the plate-tectonic model of crustal growth (Hamimi et al., 2014). The lithospheric architecture of the shield was substantially affected by key orogenic pulses that occurred approximately 1000, 800 Ma, and 600 Ma, when massive structural deformation and ophiolite obduction occurred (Abdelsalam and Stern, 1993; Reymer and Schubert, 1984). Hamimi (2021) reported that the ANS has been described by some researchers as an enormous tract of juvenile Neoproterozoic crust, whereas others considered it to be a massive suture zone that traces the boundary between formerly discrete Gondwanan blocks (Hamimi, 2021). This tectonic evolution places the ANS within a globally transformative geological setting, which also emerged simultaneously as broader changes to the Earth system, such as the separation of Rodinia, the Snowball Earth glaciations, raised atmospheric oxygen levels, and the emergence of complex multicellular life placing the ANS within a globally transformative geological context (Miller and Stern, 2021). The extent and complexity of the ANS have

been explored in several studies, which have also described the different mineral deposit types and occurrences that are inextricably connected to each of these geological phases (Kröner, 1985; Johnson and Woldehaimanot, 2003; Be'eri-Shlevin *et al.*, 2009; Botros, 2021).

The Sudanese sector of the Arabian Nubian Shield (ANS) is represented by the RSH in northeast Sudan (Figure 3-1), where our study sites are located. According to Johnson and Woldehaimanot (2003), the ANS is a vast, continuous basement exposure found in northern Ethiopia, Eritrea, Jordan, Saudi Arabia, Yemen, Egypt, and Sudan. It comprises juvenile continental crust that emerged between 900 and 550 Ma, during intra-oceanic arc terranes welded together due to ophiolite draped arc-arc sutures to the Nile Craton during the Neoproterozoic (Abdelsalam and Stern, 1993; Vail, 1978). ANS and RSH are part of the Pan-African, which is dominated by Neo-Proterozoic sequences of stratified orogenic accumulations of oceanic island-arc and plate margin Andean-type volcanics and volcanogenic and shallow-water sediments (Gass, 1977). The rocks are mostly metamorphosed into low greenschist facies and intruded by a variety of older granitoids and minor post-orogenic intrusions (Almond and Ahmed, 1987). Kröner et al. (1987) divided the RSH into five geologically distinct crustal segments (terranes) namely: 1)-Tokar/Baraka, 2)- Haya, 3)- Gebeit 4)- Gabgaba, and 5)-Gerf. They are separated by ophiolite-draped suture zones. Furthermore, El-Nadi (1989) provided genetic nomenclature for the main basement rock groups rather than geographic to facilitate correlation across the Arabian-Nubian Shield. Several authors have studied the geology of the RSH (Genna *et al.*, 2002; Qiu et al., 2006; Elsheikh *et al.*, 2015; Sudan Satti *et al.*, 2020) and reported that the RSH is a Neoproterozoic sequence of volcanic, volcaniclastic and sedimentary rocks, largely metamorphosed into low green-schist facies, and intruded by a variety of older granitoids and minor post-orogenic intrusions.

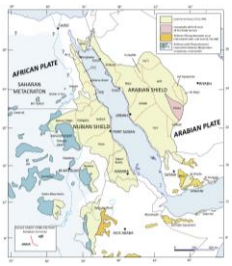


Figure 3-1. ANS extent map after Johnson *et al.* (2011)

3.1.2 Study site 1

Our first study location was approximately 90 km west of Port Sudan, northeastern Sudan (Figure 3- 2). This area coveres roughly 250 km² within the Sudanese sector of the ANS, a juvenile continental crustal block formed between 900 and 550 Ma by the accretion of intra-oceanic arc terranes along ophiolite-decorated arc-arc sutures (Abdelsalam and Stern, 1993; Elrasheed and Szabó, 2024; Vail, 1978; Whiteman, 1972). Geologically, the study area belongs to the Gebeit Terrane, one of the five distinct crustal segments identified in the RSH by Kröner et al. (1987), which also includes the Tokar/Bark, Haya, Gabgaba, and Gerf terranes. These terranes are separated by major suture zones marked by ophiolitic assemblages. The Gebeit Terrane is a well-preserved example of juvenile arc crust, consisting of metavolcano-sedimentary sequences and syntectonic igneous intrusions, structurally enclosed by the Nakasib and Oko shear zones (Almond and Ahmed, 1987; Bierlein et al., 2016; Klemenic and Poole, 1988). The RSH and particularly the Gebeit Terrane, exhibits a highly complex geological history shaped by multiple deformation events. The rocks in this region have undergone at least four major deformation phases (D_1–D_4), resulting in intense folding, shear, and metamorphism. The dominant metamorphic grade is greenschist facies, affecting the original volcanic, volcaniclastic, and sedimentary rocks (Abdelsalam and Stern, 1993; El-Nadi, 1989; Gass 1977).

Based on detailed field mapping and petrographic analysis, the lithological units at this site include low-grade metavolcanics (e.g., meta-basaltic andesite, meta-andesitic basalt, and meta-dacite), metasedimentary rocks (mainly marble), ophiolitic fragments, and syn- to post-tectonic granitoid intrusions (mostly granodiorite). Sheared quartz veins and superficial deposits are widespread. These lithologies are heavily dissected by shear zones, and the complex tectonic setting has created favorable conditions for gold mineralization, which has attracted significant ASM activity throughout the area. Artisanal miners primarily target quartz veins and shear-hosted gold occurrences, making this site a key area for studying the impacts and patterns of ASM in geologically complex terrain. The geological complexity, diverse lithology, and active ASM presence make the Gebeit Terrane an ideal setting for investigating the application of remote sensing and ML techniques in mapping mining-induced land-cover changes and structural features.

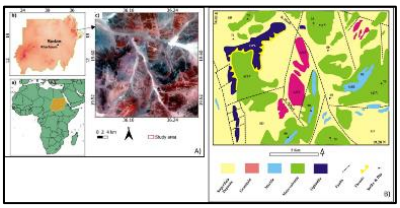


Figure 3-2. Location map of the study site #1. A/a) map of Africa, A/b) DEM map of Sudan, A/c Landsat 9 image of the study area #1. B) geological map based on our fieldwork

In addition to evaluating the performance of the algorithms, enhanced satellite imagery, previous geological maps, and fieldwork were conducted for ground truthing. While we were conducting the fieldwork, the area was divided into 10 traverses, which were designed to cover the entire area and cross the strikes. In each traverse, 10 waypoints were visited to check the lithology, sampling, and field relationships between the different rock units. Figure 3-3 shows photographs representing the main 6 lithological units in the area under investigation. Figure 3-3

reveals: a) ophiolitic rocks, b) thrust contact between ophiolite (Oph), meta-volcanics (Mtv), and superficial deposits (Spf), c) highly sheared meta-volcanics. d) shows the relationship between four lithological units: meta-volcanics (Mtv), marble (Mar), altered rocks (Alt), and superficial deposits (Spf), e) sharp contact between meta-volcanics (Mtv), marble (Mar), f) brecciated granodiorite (Gra), and superficial deposits (Spf), g) highly sheared, fragmented, and kaolinitized altered rocks (Alt), and ophiolite (Oph), h) highly silicified, ferruginated, and altered rocks (Alt), which are positive features for mineralization and i) contact between meta-volcanics (Mtv), marble (Mar), altered rocks (Alt), and superficial deposits (Spf).

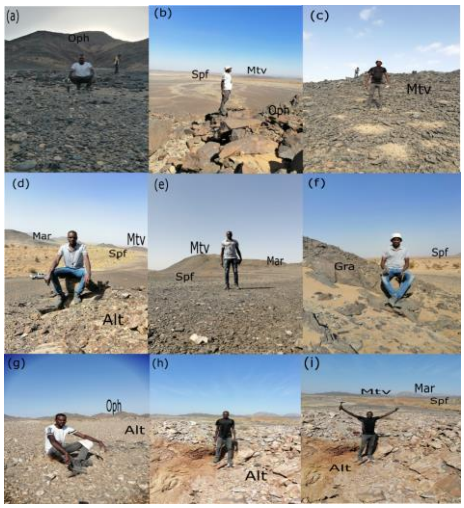


Figure 3-3. Reveals: a) ophiolitic rocks, b) thrust contact between ophiolite (Oph), meta-volcanics (Mtv), and superficial deposits (Spf), c) highly sheared meta-volcanics. d) shows the relationship between four lithological units meta-volcanics (Mtv), marble (Mar), altered rocks (Alt) and superficial deposits (Spf). e) sharp contact between meta-volcanics (Mtv), marble (Mar). f) brecciated granodiorite (Gra) and superficial deposits (Spf); g) highly sheared, fragmented, and kaolinitized altered rocks (Alt), and ophiolite (Oph), h) highly silicified, ferruginated, and altered rocks (Alt) which are positive features for mineralization; and i)

contact between meta-volcanics (Mtv), marble (Mar), altered rocks (Alt), and superficial deposits (Spf).

Petrographic studies of thin sections were conducted for petrographic descriptions. Several thin sections were prepared for meta-volcanics, meta-sediments, and associated acid and basic intrusives to validate our remote sensing findings and better characterize the lithological units. Figure 3-4 shows photomicrographs of the lithological units in the area under investigation. Sections (a) and (b) correspond to the meta-volcanic rocks, specifically meta-andesitic basalt, that have undergone strong hydrothermal alteration. This alteration is characterized by the presence of anhedral quartz grains, sericite needles, sparse muscovite, biotite, kaolinite, and streaky hematite, reflecting silicification, phyllic, potassic, and argillic alterations, along with ferrugination, all features commonly associated with mineralization. Sections (c) and (d) represent meta-volcanics containing quartz, sericite, biotite, and plagioclase. The plagioclase-phyric texture displays intense silicification, as evidenced by sub-oriented angular quartz fragments produced by shearing. Sections (e) and (f) show plagioclase–pyroxene–phyric metabasalt, where the phenocrysts are set in a partially sericitized and devitrified groundmass containing minute plagioclase laths. Hydrothermal alteration is evident through widespread sericitization, silicification with angular quartz grains, and potassic alteration marked by slender flakes of biotite. Sections (g) and (h) illustrate the meta-andesite composed of kaolinitized plagioclase phenocrysts within a fine-grained, devitrified groundmass containing acicular plagioclase crystals. The presence of scarce tremolite, possibly after pyroxene, indicates a low-grade metamorphism. Section (i) depicts a fine-grained, highly sheared marble cut by quartz–calcite stringers, while (j) shows a brecciated microgranodiorite containing plagioclase phenocrysts in a silicified, carbonatized, and chloritized groundmass.

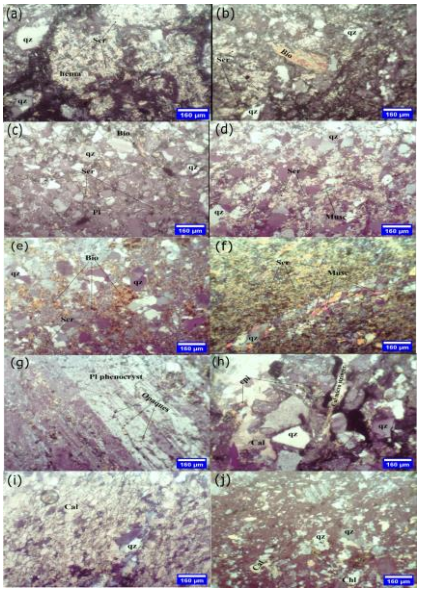


Figure 3-4. Representative thin sections from the petrographic study are shown. (a–b) Meta-volcanics (meta-andesitic basalt) showing strong hydrothermal alteration with quartz, sericite, biotite, kaolinite, and hematite, indicating silicification, phyllic, potassic, argillic, and ferruginous alterations. (c–f) Meta-volcanics with plagioclase-phyric textures exhibiting intense silicification and sericitization. (g–h) Meta-andesite with kaolinitized plagioclase phenocrysts and scarce tremolite, suggesting low-grade metamorphism. (i) Sheared marble with quartz–calcite stringers. (j) Brecciated microgranodiorite showing silicification, carbonatization, and chloritization.

3.1.3 Study site 2

The Ariab area, located in the Haya Terrane of the Red Sea Hills in northeastern Sudan, is a geologically significant site within the ANS. The Haya Terrane is largely concealed beneath the Phanerozoic sedimentary cover to the west and is structurally separated from the Gebeit island-arc terrane (830–720 Ma) by a northeast-trending shear zone marked by the

Oshib-Nakasib ophiolite complexes (Abu-Fatima et al., 2021). The dominant lithologies within the Haya Terrane include greenschist facies metavolcano-sedimentary assemblages, collectively known as the Nafirdeib Series, as well as batholithic granitoids. The geological succession of the Ariab area is classified into four main lithological groups arranged from oldest to youngest:(i) the ultramafic complex of Oshib, (ii) Ariab volcano-sedimentary sequences, (iii) the Awat Asotriba Group, and (iv) intrusive granitoid bodies (Abdelrahman et al., 2024). These lithologies are illustrated in the geological map of the Ariab area (Figure 3-5). The ANS is renowned for its numerous gold occurrences and evidence of mining activity extending back over 5,500 years. The principal types of gold deposits in this region are orogenic gold and those associated with volcanogenic massive sulphide (VMS) systems (Johnson et al., 2017).

Ariab is especially noted for its auriferous VMS deposits, which have been the focus of numerous studies (Bakheit and Matheis, 2017; Kröner et al., 1987; Shanti and Roobol, 1979). In the Ariab mining district, gold is extracted from gossans that have developed over VMS deposits formed in a Tonian ensimatic arc setting of the Neoproterozoic Haya Terrane(Abu-Fatima et al., 2021). These VMS-related deposits are hosted within felsic and intermediate volcanic and volcaniclastic rocks, with common intercalations of chert and jasper. The hydrothermally altered felsic rocks contain sulfide bodies that are largely oxidized at the surface, forming extensive gossan zones, particularly notable in Hadal Awateb (Abdelrahman et al., 2024). The gold-bearing gossans in Ariab have been classified into four main facies based on their mineralogical and textural characteristics (Abu-Fatima et al., 2021; Daoud et al., 2025).These facies, from the outer to the inner zones, include: (i) iron-stained clays (mainly kaolinite) and iron oxide-impregnated schistose host rocks ("foliated gossan"); (ii) bleached schistose rocks, often with jarosite-rich patches; (iii) jarosite-rich halos (5–20 m thick) surrounding the core; and (iv) silica-barite rock (SBR) cores that typically host concentrated gold mineralization (Figure 3-6). Jarosite-rich zones also

contain gypsum, alunite, and Fe- and Cu-sulfates. In this study, we focused on the uppermost parts of the gossan, which are enriched with jarosite and exhibit the widest horizontal distribution in the Ariab area. Our research, titled "Assessing the Potentiality of PRISMA Hyperspectral Remote Sensing Data in Detecting and Mapping Gossan in Ariab Area, Red Sea Hills, NE-Sudan: Insights from BR, Density Slice, and RF," explores the capacity of hyperspectral remote sensing for mineral exploration. Specifically, we employed PRISMA hyperspectral data and high-resolution PlanetScope imagery to detect and map gossan zones. This represents the first known application of PRISMA and Planet data for this purpose in this region.

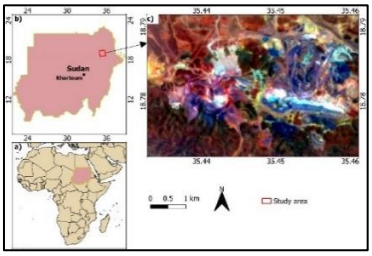


Figure 3-5. Location map of the study site #2. a) map of Africa, b) map of Sudan, c) Sentinel-2 image of the study area #2.

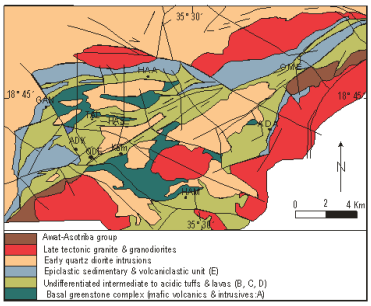


Figure 3-6 Simplified geological map of the Ariab area (study area #2) after Abdelrahman et al. (2024)

3.1.4 Study site 3

The third study site was located in the eastern part of the Bayouda Desert, within the western margin of the ANS in northern Sudan (Figure 3-7). Geographically, the site lies approximately 15 km from Albawga, 40 km from Berber, and 60 km north of Atbara in a region characterized by a relatively high population density (Figure 3-8). This location is east of the Nile River, placing it in a key transitional zone between the Nile Valley and the inner ANS. Geologically, low-grade volcano-sedimentary rocks and island-arc/back-arc ophiolitic assemblages predominantly underlie the area (Schandelmeier et al., 1987; Zeinelabdein and Nadi, 2014; Zeinelabdein and Elsheikh, 2015). These units represent a typical accretionary complex associated with Neoproterozoic arc development and back-arc basin closure, which is characteristic of much of the ANS. The dominant lithologies at this site are greenschist facies metavolcanic rocks, including metabasaltic andesite, meta-andesite, and meta-dacite (Abdelsalam and Stern, 1993; Almond and Ahmed, 1987; Küster and Jean-Paul, 2001).

These volcanic and volcaniclastic rocks exhibit extensive deformation, including intense folding, faulting, and shearing, as well as variable degrees of metamorphism, resulting in a structurally complex terrain. Field-based geological observations confirmed the predominance of these metavolcanic units, with some areas showing clear signs of hydrothermal alteration, which likely contributed to local gold mineralization. ASM is widespread at this site, with mining activities mainly concentrated within meta-andesite and, in some zones, meta-basaltic rocks. These lithologies, particularly where intersected by shear zones and quartz veins, appear to host gold mineralization, which has attracted intensive ASM operations. The proliferation of mining pits, waste dumps, and surface disturbances provides an important opportunity to assess and monitor ASM impact using remote sensing techniques. The site's geological diversity, structural complexity, and high intensity of mining activity make it a critical zone for evaluating the

performance of remote sensing and ML methods in detecting artisanal mining and its environmental footprint in low-grade metamorphic terrans.

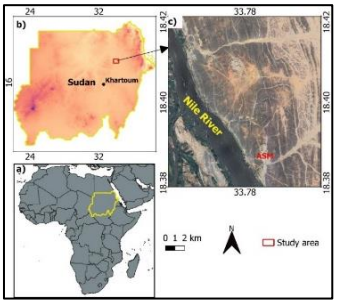


Figure 3-7. Location map of the study site #3. a) map of Africa, b) DEM map of Sudan, c) Planet image of the study area #3.

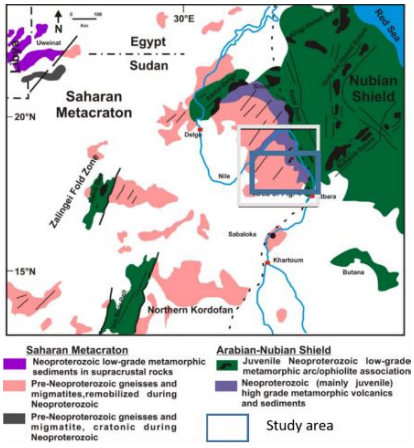


Figure 3-8 regional and local geological map of study area #3 after Evuk (2013)

3.2 Dataset

We used various of geospatial datasets to accomplish the aims of this dissertation. These datasets include geological maps, remote sensing, field data, and petrographic analyses. Their combination and integration enables a more thorough and precise examination of lithological features, mineral alteration zones, and ASM(ASM) operations. For a reliable cross-sensor assessment and complementarity of spectral and spatial information, we used a broad range of satellite-based datasets. These include optical and hyperspectral sources. The optical datasets included multispectral images from Landsat-7 Enhanced Thematic Mapper Plus (ETM+), Landsat-8, Landsat-9 Operational Land Imager (OLI/OLI-2), and Sentinel-2. We also employed PlanetScope imagery, which provides high spatial resolution, as one of the first applications of non-commercial data in ASM mapping and monitoring. This study is the first to use PRISMA hyperspectral data to identify mineral alterations related to ASM practices. Notably, gossan formations belonging to VMS deposits were detected in both hyperspectral PRISMA datasets, presenting novel insights into the potential of using space-borne hyperspectral sensors in early stage mineral exploration.

The pectral and spatial characteristics of the remote sensing sensors used in this study are showen in (Table 3-1). These datasets were sourced from official portals: Landsat datasets were obtained through the USGS Earth Explorer (https://earthexplorer.usgs.gov/), Sentinel-2 data from Copernicus Open Access Hub (https://dataspace.copernicus.eu/); PlanetScope data from Planet Labs (https://www.planet.com/), and PRISMA from the Italian Space Agency (https://prisma.asi.it/missionselect/). In order to describe the surface morphology and lithological context of the research areas, we used existing topographic and geological maps (Abdelrahman et al., 2024; Abu-Fatima et al., 2021; Kenea, 1997). These maps provide important baseline knowledge about regional structures, rock types, and previously mapped mineralized zones. To ensure better validation of remote sensing

in situ fieldwork and observation were included to better understand lithological and structural complexities. Moreover, intrusive contacts, cross-cutting relationships, and the relative timing of geological units were identified. Field documentation also recorded evidence of legacy and ongoing small-scale placer mining activities. Representative rock samples were collected from key locations, including areas with mineralized quartz veins and hydrothermal alteration zones. Samples were processed into thin, thick, and polished sections for petrographic, mineralogical, and textural analyses, providing crucial microscale information that supports remote sensing interpretations and confirms the presence of target mineralization.

Table 3-1. Spectral and spatial characteristics of remote sensing sensors used in this study.

Sensor	Spatial Resolution	Spectral Resolution	Spectral Range (nm)	Key Applications
Landsat 7/8/9	30 m (15 m PAN)	8–11 multispectral bands	430–12,500	Baseline lithological mapping; thermal anomaly detection near ASM sites.
Sentinel-2A	10–60 m (10–20 m VNIR)	13 multispectral bands	443–2,190	Vegetation stress monitoring; red-edge band analysis for ASM-impacted areas.
PlanetScope	3 m	8 multispectral bands	455–860	Fine-scale ASM footprint delineation (first

Sensor	Spatial Resolution	Spectral Resolution	Spectral Range (nm)	Key Applications
				application in ASM monitoring).
PRISMA	10–30 m (resampled)	237 hyperspectral bands	400–2,500	Mineral alteration mapping (e.g., gossans); ASM-exposed VMS detection.

3.2.1 Image Preprocessing and Data Preparation

We subjected all satellite images to systematic preprocessing prior to analysis to ensure consistency, comparability, and high-quality results. Landsat-7 ETM+, Landsat-8 OLI, and Landsat-9 OLI-2 images from 2003, 2013, 2023, and 2024, respectively, were selected to represent long-term temporal changes, whereas Sentinel-2 MSI data from 2015, 2021, and 2024, PlanetScope imagery from 2016, 2021, and 2024, and PRISMA hyperspectral data from 2023 and 2024 provided high-resolution and hyperspectral perspectives for recent observations. These datasets were collectively employed to mapp ASM activities, lithological discrimination, and detection of hydrothermal alteration zones. For preprocessing, Landsat-7 ETM+ data were corrected to surface reflectance using the FLAASHA atmospheric correction algorithm, whereas Landsat-8 OLI and Landsat-9 OLI-2 Level-2 products were used directly, as they had already undergone atmospheric and radiometric correction by the USGS. Sentinel-2 Level-1C imagery was converted to surface reflectance (Level-2A) using Sen2Cor, and PRISMA hyperspectral data were corrected using the QUAC algorithm. PlanetScope imagery was used in its surface reflectance form, as

provided by Planet Labs. All datasets were co-registered to the WGS 84/UTM Zone 36N/37N coordinate system and resampled to a uniform 30 m spatial resolution to ensure compatibility across sensors, where applicable. Cloud and shadow masking were applied prior to the analysis. These preprocessing steps ensured that subsequent image enhancement and analytical procedures such as band ratio generation, principal component analysis (PCA), and supervised classification (RF and SVM) were based on radiometrically consistent and geometrically aligned datasets.3.3 Applied Methods

3.3.1 Image Processing Techniques

Image processing is an enhancement technique that offers powerful means for the visual examination and analysis of multispectral or hyperspectral images utilising individual bands or BR (Mwaniki et al., 2015). Therefore, image processing techniques are increasingly used in geological studies to differentiate lithological features and mineral alterations from satellite imagery (Gupta, 2025). Tj *et al.* (2022) also emphasised the key role of image processing in the extraction of geological features, mainly when it comes to geological mapping using satellite imagery. They focused on the information extraction of geological features using computer-aided techniques to manipulate or modify imagery and concluded that image processing aids in the collection of clearer and more helpful observations for mapping purposes.

3.3.1.1 Colour Composite

Color composites (CC) are an essential component in processing remote sensing data, mostly satellite images, in order to improve the visual analysis, understanding, and interpretation of satellite data (Diwyacitta et al., 2022). It merges several spectral bands to deliver visually interpretable images of the Earth's surface. It has been deployed for several of roles, including classifying land use and land cover and monitoring environmental conditions (Singh and Shanmugam, 2016).

True color composite (TCC) and false color composite (FCC) represent these two composites. TCC normally combines spectral bands, often red, green, and blue, in order to create images that have appearances similar to what the human eye sees. Conversely, an FCC can be created by combining visible and infrared wavelengths or by exclusively utilizinginfrared bands. The way these bands are allocated to the three primary colors (red, green, and blue) for presenting the images determines how CC images are interpreted. This, in turn, depends on the spectral properties of the region under investigation (Ibrahim et al. 2016). This method is frequently applied in geological research, such as gossan detection (Gahlan and Ghrefat, 2018), lithological mapping (Qiu et al., 2006; Bajwa et al., 2020; Hassan et al., 2020), and hydrothermal alteration zone identification (Muavhi, 2022; Williamson et al., 2015). Accordingly, we used a number of TCCs and FCCs for Landsat, Sentinel-2, PlanetScope, and PRISMA in this dissertation to improve the visual inspection of lithological units, understand the alteration zone, and detect ASM.

3.3.1.2 Band Ratio

Band ratios (BR) are an important remote sensing data processing technique that relies on dividing the reflectance values of one spectral band by another in order to improve certain features that are relevant while reducing noise from topographic changes, illumination differences, and weather conditions (Bachri et al., 2019; Lillesand and Kiefer, 2015). BR has become essential in geological studies because it applies to the diagnostic spectral absorption attributes of minerals and rocks, which differ throughout wavelengths depending on electronic transitions (as with iron oxides) or vibrational operations (like hydroxyl-bearing clays) (Crosta, 1989). Furthermore, BRs are vital for mineral exploration and environmental monitoring because they are particularly useful for determining mineral assemblages, defining alteration zones, and differentiating lithological units (Sabins, 1999; van der Meer et al., 2012). Subsequently, we applied several BRs. For example, in mapping mineral alteration, ratios including Landsat's band 4/band 2 (which

highlights iron oxides) and band 5/band 7 (which isolates clays that contain hydroxyl) implement spectral contrasts in the visible to shortwave infrared (SWIR) regions to identify hydrothermal zones (Hills, 2014; Pour et al., 2019). For lithological mapping, Sentinel-2's band 11/band 12 targets carbonate signals, whereas Landsat's band 5/band 2 distinguishes iron-rich mafic rocks from felsic units (Gad and Kusky, 2007). Hyperspectral remote sensing allows for much more precise BR use because it provides over 100 contiguous and smaller bands through the electromagnetic spectrum compared to multispectral sensors. Therefore, the ability to detect minerals with similar spectral signatures is advanced by this data, which enables the utilization of delicate absorption properties associated with distinct minerals. Hyperspectral BR, such as 2200 nm/2300 nm and 2340/2200 nm, is effective in separating kaolinite, alunite, and other clay minerals due to their distinct hydroxyl absorption features (Kruse et al., 2003). Likewise, carbonate can be identified from sulphate and phyllosilicate groups using BR 2330 nm/2210 nm or 2310/2160 nm (van der Meer et al., 2012). Additionally, alteration minerals like chlorite, epidote, and sericite, which are commonly known as hydrothermal alteration system indicators, can be observed using hyperspectral of BR (Rowan and Mars, 2003; Abubakar et al., 2019; Pour *et al.*, 2019). Subsequently, when we compared BR obtained from hyperspectral with multispectral data, it significantly enhanced the spectral contrast, which facilitates the geologist's tasks in finding alteration assemblages, even mixed pixels, or in complex geological environments. We used BR of both multispectral and hyperspectral, which allow us to identify hydrothermal alteration zones and separate mineral phases with high precision. This is important in mapping lithological units, ASM sites, and hydrothermal zones.

3.3.1.3 Density slicing

Density slicing is a useful method for grouping and evaluating land surface features based on their spectral properties. By applying this technique, various information can be extracted from a single satellite image by splitting it into different classes based on the value of the pixel

intensity (Hamandawana et al., 2006). Density slicing is also used to extract information from processed images, such as BR images. For example, Abdelsalam et al. (2000) and Madani (2009) applied density slicing to BR images of Landsat 5 and Landsat 7, which effectively identified mineral alteration related to gossan mineralization. To map anomalies, they adopted the threshold rule: mean + standard deviation. The usefulness method has been demonstrated by several studies (Sabins, 1999; Saadi and Watanabe, 2009; Abdoulatif et al., 2025) to map mineralized hydrothermal zones. Furthermore, according to Abrams *et al.* (1983), alteration zones usually show unique spectral signals that can be properly detected by leveraging BR, followed by the density slicing method using threshold M + SD, where M refers to the mean and SD to the standard deviation. Jensen (2016) was able to map alteration zone anomalies associated with iron oxides and clay minerals through using the statistical mean and standard deviation method on BR images. Accordingly, this approach is recommended as useful for classifying mineralized zones by isolating pixels with high reflectance values (anomalies) as mineralization indicators in the given area. In this dissertation, we followed a multiple-threshold approach to refine the detection of alteration zones by identifying weakly altered areas from highly mineralized regions:

(i) M+ SD (Moderate anomaly threshold),
(ii) M+1.5*SD (intermediate anomaly threshold),
(iii) M+2*SD (high-confidence anomaly threshold),
where M = Mean of pixel values for a given BR and SD = standard deviation

3.3.1.4 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method commonly applied for dimensionality reduction and pattern detection, and it is used in different fields such as remote sensing, social sciences, and economics (Kelley, 2010). PCA was formalized by Hotelling (1933) based on Pearson's (1901) concept of PCA. It reduces the number of

variables in the datasets while maintaining variance, thereby simplifying the analysis and visualization (Karamizadeh et al., 2013; Kelley, 2010). According to Mesa et al. (2018), PCA is powerful for data analysis, identifying patterns in the data, and expressing them in a way that highlights their similarities and differences. This pattern in data can be challenging to find when the data are in arrays with high dimensions, where classical data display is difficult. Mohammed *et al.* (2016) reported that PCA is used as an effective tool for reducing redundancy in the data when the data are highly correlated to uncorrelated principal component and is used notably in areas such as image processing, classification, and anomaly detection.

In remote sensing and geological applications, PCA is essential, especially for analyzing and interpreting geological features from satellite data. For instance, PCA was useful in enhancing geological structure measurements by identifying the surface orientation of geological features (Quinn and Ehlmann, 2019). Tianshi *et al.* (2021) also used PCA and Landsat data to delineate hydrothermal alteration anomalies. They concluded that PCA was useful in highlighting subtle differences in mineral response throughout the spectral bands and minimized noise from water bodies and vegetation. Furthermore, PCA improved visual inspection of geology by transforming the spectral bands into new images, which enhanced lineament extraction and the identification of geological structures (H. et al., 2016). It is also used to optimize lithological mapping, as reported by Gasmi *et al.* (2016). The authors demonstrated the validity of PCA and ASTER data in enhancing the differentiation between lithological units in Tunisia. They found that PCA was helpful in detecting lithological boundaries and identifying new units. In this study, we applied PCA to reduce data redundancy and feature extraction for hyperspectral data (PRISMA). Our objective was to use all the information in the data while minimizing redundancy and the computational process.

3.3.1.5 Directed Principal Component Analysis

Directed Principal Component Analysis (DPCA), also known as selective or feature-oriented PCA, was developed based on classic PCA to handle certain issues in high-dimensional data by modifying factors to better guide specific decisions (Kao and Van Roy, 2014). DPCA is effective in handling time-correlated and distributed datasets (Ye and Zhu, 2020). According to Crosta (1989) and Liu *et al.* (2011), DPCA is an advanced spectral enhancement method that uses eigenvector loadings from a carefully chosen subset of correlated spectral bands to isolate specific geological features.

In contrast to conventional PCA, which processes all available bands indiscriminately, DPCA deliberately selects narrowband combinations that match the diagnostic spectral signatures of target materials, including alteration minerals (Hdeid et al., 2024; Loughlin, 1991). Therefore, DPCA increases the visibility of desirable targets by highlighting bands that are sensitive to their distinct spectral signatures. For example, the eigenvector loadings of visible/near-infrared (VNIR) and SWIR-dominant bands are best used to identify hydroxyl-bearing minerals, which demonstrate absorption diagnostics in the 2.10–2.28 μm shortwave infrared (SWIR) (Amer et al., 2010). Regions enriched with target mineralogical alterations are indicated by positive (bright pixels) and negative (dark pixels) eigenvector values in the resultant components (Novaes et al., 2024; van der Meer et al., 2012). Accordingly, we adopted for DPCA to Landsat 7/8/9 and Sentinel-2 bands sensitive to iron oxide absorption features (e.g., Red, SWIR). Positive eigenvector loadings in these bands show iron-enriched areas, consistent with the spectral absorption characteristics described by Crosta (1989) and Van Der Werff and Van Der Meer (2016), which were originally developed for Landsat Thematic Mapper (TM) bands 1, 4, 5, and 7. In this study, we adapted DPCA to Landsat 7 ETM+, Landsat 8/9 OLI (bands 1, 3, 4, 5 and 1, 4, 5, 7), (2, 4, 6 and 7, and 2, 5, 6, 7) for Landsat7, 8, and 9. Sentinel-2 MSI (bands 2,4, 8a, 11 and 2, 8a, 11, 12). These band combinations make it possible to map gold mineralization and associated minerals linked to

alteration zones across a variety of geological environments by increasing the sensitivity to hydroxyl-bearing phases and iron oxides (Mielke et al., 2014).

3.3.1.6 Constrained Energy Minimization (CEM)

To isolate specific hydrothermal alteration minerals, the Constrained Energy Minimization (CEM) technique employs a Finite Impulse Response (FIR) filter, designed with a constraint that enhances the response of the desired targets while suppressing interference from background materials(Shebl et al., 2021). This approach requires the use of a correlation or covariance matrix to model complex and often unknown backgrounds. Here, the output of the CEM analysis is represented as a pixel-based map, each corresponding to a target mineral, with each pixel reflecting a defined spectral endmember. The CEM method was implemented using reference laboratory spectra from the USGS (https://www.usgs.gov/labs/spectroscopy-lab/science/spectral-library), spectral library specifically targeting the identification of goethite, hematite, jarosite, illite, kaolinite, chlorite, and calcite. The technique was applied to PRISMA hyperspectral data to enhance the detection and mapping of key alteration minerals, particularly those associated with broader alteration zones previously identified using multispectral data.

3.4 Reference data collection for training and testing

Reference data collection can be considered a fundamental stage in classification and modelling; therefore, obtaining high-quality training and testing data is essential. This processing is even more challenging when it comes to being employed in highly deformed, complex geological environments such as the ANS due to the inhomogeneity and similarity of rock unit composition and the itinerant nature of ASM. More importantly, the reliability of the data used as a reference has a major impact on the certainty of the outcomes for both lithological mapping and ASM, because it determines the accuracy and efficiency of

the applied methods. Therefore, having more precise training and testing data improves the capacity of the algorithms and remote sensing methods to create accurate predictions, leading to more reliable analyses and conclusions (Elahi et al., 2022). Consequently, we have diligently developed this dissertation by meticulously reviewing the existing geological map (Kenea, 1997), conducting a detailed visual analysis, inspecting processed remote sensing data, and analyzing high spatial resolution imagery from Google Earth. Additionally, extensive fieldwork was conducted, including the establishment of strategic traverses and checkpoint locations throughout the study area (Figure 3-9) to ensure comprehensive coverage of all lithological units present. This method of reference data collection coincided with that of several studies (Butler et al., 2024; El-Omairi and El Garouani, 2024; Zeng et al., 2023). We applied the same strategy for ASM and mapping lithologies to ensure the reliability of the results. Moreover, to study ASM with Landsat 9 and PRISMA, we applied five versions of the training dataset, each with different aims and data volumes (Figure 3-10). First, we used only 95 data points (orig), deliberately excluding reference data from wadi areas, as these regions contain a mixture of materials eroded from surrounding rock units, including ASM-impacted zones, which increases the risk of misclassifying between the ASM and non-ASM features. Next, we added the missing information (sampling points from the wadis) in two steps: (ext1) with only a limited, 126 data points (i.e. 31 additional data to the original), and (ext2) with 232 data points (137 additional data to the original), and (ext_12) merging the two additional datasets ('ext1' and 'ext2') resulting in 263 data. Finally, (aug) we developed an algorithm in R 4.4.3 (R Core Team) (ud-geoai, 2025) using a pseudolabeling approach tailored to the challenges of limited reference data. The resulting dataset consisted of 794 records selected from the ext1 subset, which, despite its limited size, offered sufficient spatial coverage and diversity of geological features to maintain model stability. This deliberate use of fewer but higher-quality samples was based on the recognition that the spatial representativeness and geological diversity of the training data

can compensate for smaller sample sizes. During testing, models trained on spatially well-distributed reference points achieved classification accuracies comparable to those generated with substantially larger augmented datasets, indicating that spatial coverage plays a more decisive role than data volume. Moreover, the addition of excessive pseudo-labeled data did not necessarily enhance the model accuracy but primarily improved the spatial coherence of the classified units. Consequently, our workflow prioritized a balanced spatial representation of lithological assemblages and geomorphic contexts, ensuring that the ext1 dataset, although numerically limited, served as an optimal foundation for producing a robust and interpretable geological classification.

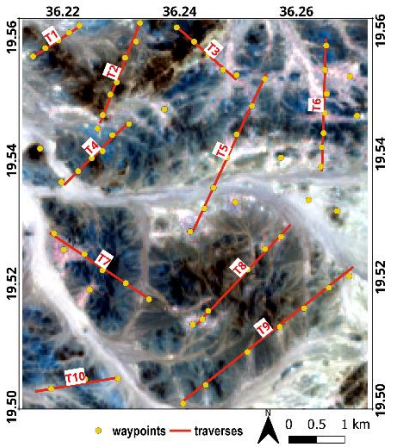


Figure 3-9. Showing the traverses and waypoints locations across the study area

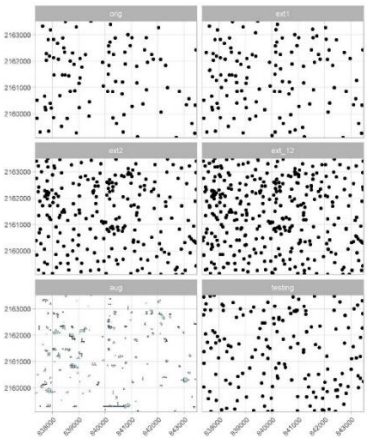


Figure 3-10. Six sampling point patterns of the training dataset with different aims and amounts of data (orig: 95 data, ext1: 126 data, ext2: 232 data, ext1_2: 263 data, aug: 794 data, testing: 162 data).

ASM sites were identified through visual analysis and delineated using manual vectorization. The resulting map served as the basis for spatially explicit area-based accuracy assessment (Figure 3-11) .

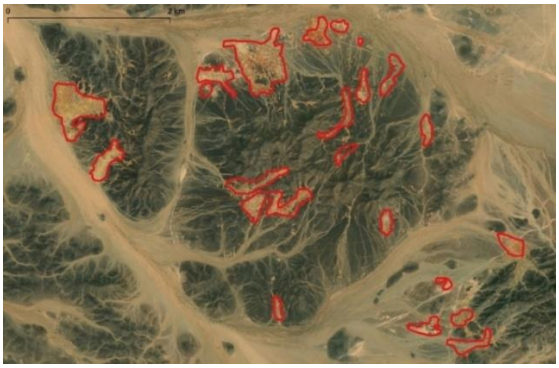


Figure 3-11. Artisanal and Small-scale Mining (ASM) sites (polygons with red outline) in the study area

3.5 Machine Learning Algorithms

ML is a crucial part of the artificial intelligence field and statistical models, which allow machines (computers) to learn from data and improve over time without the need for explicitly written code, and are inherently included in the number fields of research (Ramos et al., 2023). It involves training a model on labelled data, where the input-output pair is known (Venkata Mahesh Babu Batta, 2024). Based on Rao *et al.* (2017), ML has invented a number of algorithms that are frequently used for interpreting text, identifying patterns, and several applications in business. In addition to providing the foundation for the statistical computational principles of learning processes. Furthermore, the authors reported that ML has sparked additional investigation in data mining to reveal hidden regularities or irregularities in social data that are expanding by the second.

ML can be classified into various types, with having a particular goal and use. According to Aslam (2023), the main ML groups are supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, and it is essential to understand the different categories to select the best algorithm for a given application. Supervised learning includes algorithms such as linear regression, SVM, Naïve Bayes, RF, and decision trees, which learn from labelled data to perform prediction or classification (Kumar et al., 2024). Unsupervised learning refers to algorithms that detect patterns in unlabelled data, including K-Means Clustering, Hierarchical Clustering, and Association Rule Learning (Aslam, 2023; Ling, 2023). ML can be classified into parametric and non-parametric models. Parametric methods include logistic regression and Naive Bayes, which assume a predefined functional relationship between input variables and outputs, optimizing a fixed set of parameters. These methods excel in scenarios with well-understood data distributions and limited computational resources, offering interpretability at the cost of flexibility.

In contrast, non-parametric models, such as RF, SVM, and XGB, make no a priori assumptions about the data structure, dynamically adapting to inherent complexities through ensemble learning or kernel-based transformations (Breiman, 2001; Maxwell et al., 2018). Their capacity to handle high-dimensional, noisy, and nonlinear relationships common in hyperspectral imagery and geophysical datasets renders them indispensable for resolving subtle mineralogical features in heterogeneous terrains (Chen et al., 2019; Mangeruga et al., 2020; Zebker, 2021). This capability makes them particularly well-suited for our dataset, where patterns and relationships are expected to be intricate and challenging for parametric models. In geospatial data analysis, ML is considered as an enhanced tools (Lary et al., 2016), similar to lithological mapping (Hebri et al., 2024; Shanthi et al., 2024, p. 2022202; Wang et al., 2024), allowing the identification of complex patterns in high-dimensional datasets. In this study, we used both parametric and non-parametric ML approaches to address the challenges of lithological mapping, hydrothermal alteration detection, and ASM monitoring in a geologically intricate region (Belgiu and Drăguţ, 2016; Foody and Mathur, 2004; Pal, 2005).

3.5.1 Naive Bayes

The Naïve Bayes (NB) model is a probabilistic supervised ML classifier based on Bayes' theorem, which assumes conditional independence across the predictor variables considering the class label (Maron, 1961). NB determines the subsequent chances for every class using Gaussian distributions for continuous variables or frequency counts for discrete features by multiplying the prior probabilities by the likelihoods obtained from the training data (Ananthkumar, 2023). This classifier works well for processing remote sensing data, including initial exploration, modelling, and benchmark assessment. It is computationally powerful, adaptable to large huge datasets, and resilient to irrelevant features (Ren et al., 2022). Nonetheless, when the correlation between features is high or in the case of insufficient data, its performance could

decrease and lead to skewed probability estimations (Bai et al., 2023). However, NB continues to be incredibly helpful for classification tasks, particularly among high-dimensional domains, including medical diagnostics and remote sensing data (Domingos & Pazzani, 1997; Tariq et al., 2023). In mineral resource exploration and lithological mapping, it enhances the accuracy of lithological predictors by utilizing observed data and experiences; therefore, this classifier represents an important tool, mostly for lithological mapping and mineral prospecting (EL-Omairi and El Garouani, 2023; Ruiz et al., 2013). Furthermore, they emphasized that the integration of NB with remote sensing data and other ML algorithms improved lithological mapping, offering a robust classification of geological features over large areas. In this study, lithological units in the area of interest where mapped using NB, and zero-frequency problems in sparse categorical data were mitigated using Laplace filtering.

3.5.2 Multiple Adaptive Regression Splines (MARS)

Friedman and Steppel (1974) discovered Multiple Adaptive Regression Splines (MARS). It is an effective classifier that handles high-dimensional regression problems and resolves the issue of dimensionality by focusing on lower-order relationships (Kooperberg, 2012). The algorithm uses flexibly defined spline functions, and owing to its flexibility, MARS is usually able to model functions that show high variation in one part of the predictor space and are smoother in other parts (Friedman, 1991).

The algorithm functions according to the data structure and is a valuable tool in various disciplines, including environmental studies, genetics, and mapping. For instance, Park *et al.* (2017) compared logistic regression (LR) with MARS to map groundwater potentiality. They found that MARS was more adequate than LR, with an area under the curve of 0.867 for success rate. Similarly, Zabihi *et al.* (2016) assessed MARS and RF for groundwater potential using R. This study revealed successful predictions using various explanatory variables. While MARS

has been increasingly used in geoscience applications, including mapping erosion susceptibility (Javidan et al., 2019), landslide hazard mapping (Felicísimo et al., 2013), and modelling the relationship between rock quality designation and permeability (Qureshi et al., 2022), its applications in lithological mapping and ASM have not been properly explored yet. This dissertation represents a major attempt to evaluate the utility of MARS in these fields. We illustrate the significant potential of this method for precisely differentiating lithological units and detecting ASM-related perturbations by using its prowess in modelling intricate, nonlinear interactions between spectral, spatial, and geological variables. According to our research, MARS can be a useful tool for improving geospatial mineral exploration procedures, especially in settings with complex structures and scarce data. The main hyperparameters of MARS are the 'number of terms' (nprune) and the 'degree,'which specifies the possible highest degree of interactions. To optimize performance, I conducted a systematic grid-based search to find the best combination with 20 evenly spaced nprune values between 2 and 100 rounded to an integer, while 'degree' was tested from 1 to 10 in increments of one.

3.5.3 Support Vector Machine (SVM)

Vapnik (1979) and Vladimir Vapnik (1995) developed the SVM as a supervised ML algorithm for classification and regression problems relying on the statistical learning theory. It was primarily designed to handle binary objects; however, it has been extended to multiclass tasks (Melgani and Bruzzone, 2004). The SVM classifier depends on identifying the most suitable hyperplane for splitting the data into classes with the a larger variance (Vapnik, 2013). It is widely used because of its accuracy, effectiveness, and adaptability (Han et al., 2025). It functions as a general relationship weight to enhance the accuracy of the classifier by investigating the data dimensionality, applying a kernel function to weight them, and guiding the classification pressures based on total correlation (Friedman, 2009; Zhu et al., 2023).

The potential of SVM to manage classification problems in complex geological settings such as RSHs, was demonstrated by Yu *et al.* (2012) and Shebl *et al.* (2023). This research investigated several algorithms, including different input datasets, in order to identify suitable data for upgrading the classification accuracy. SVM exhibited better accuracy than RF, maximum likelihood, and XGB. The study concluded that while SVM adaptation with larger datasets could be affected by computational challenges, it is an effective means for directing the first phase of lithological maps, mainly in areas where the geological units are already known (Burges, 1998). Therefore, due to its high accuracy, reliability, and robustness, SVM is frequently applied and recommended in various geoscience applications, including remote sensing (Mountrakis et al., 2011; Sebtosheikh *et al.*, 2015; Abbaszadeh et al., 2022). In lithological mapping, Othman and Gloaguen (2014) and Shebl *et al.* (2023) examined the usefulness of SVM in mapping chromite deposits through classifying the lithological units of the Mawat ophiolite complex in NE Iraq utilising ASTER data and Egypt. Likewise, SVM potentiality in ASM and mineral prospecting has been demonstrated by Pal and Mather (2005) and Ngom *et al.* (2023) using three datasets: geological, geochemical, and geophysical. Mapping hydrothermal alteration zones using AVM and remote sensing data has also been evaluated through analyzing Hyperion images (Wang and Zheng, 2010). I adjusted the SVM hyperparameter tuning and radial basis kernel using grid search and the sigma parameter to manage the hyperplan's nonlinearity (0.01 to 0.1 by 0.01 steps). To penalize the misclassification, we used the C parameter (1 to 10 by 1 step).

3.5.4 Random Forest

Random forest (RF) is an ensemble supervised ML method known for its robustness due to the algorithm's ability to minimize overfitting by averaging the different tree results, allowing it to handle large data problems (Kulkarni and Sinha, 2014; Zhao et al., 2024). RF functions by building several decision tree bootstrap replicates of training data and

aggregates their outputs in order to increase classification accuracy and mitigate predictive modelling (Rose and Hassen, 2019; Schonlau and Zou, 2020). According to Kar and Singh (1979) and Breiman (2001), RF is a combination of tree predictors in which each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. RF effectively manages challenges such as imbalanced data and high-dimensional feature spaces, thereby enhancing fraud detection accuracy (Jaiswal et al., 2024). RF has been used in several fields for high-accuracy classification in medical diagnostics (Maindola et al., 2024), crop yield prediction, soil quality assessment (Goel and Abhilasha, 2017), and large-scale data analysis (Bhadula et al., 2024). It has also been used to process remote sensing data for lithological and mining mapping (Elrasheed and Szabó, 2024; Mohamed Taha et al., 2022; Shebl et al., 2023b; Wang et al., 2024; Xi et al., 2022).

In my thesis, I evaluated RF classification for lithological and ASM mapping. I developed and assessed the model by splitting the datasets into training (70%) and testing (30%) subsets using random sampling to ensure a representative distribution of the data points. Subsequently, I used the training data to construct the model and the test data to retain an unbiased accuracy assessment. In order to build the RF model, I utilised a 10-fold cross-validation approach (train control method set to "cv") that mitigates overfitting by iteratively training and validating the model on a number of subsets of the training data. For model optimization, I implemented RF using R (R Core Team, 2024) and the scikit-learn library in Python (Pedregosa et al., 2011) with hyperparameter tuning via a randomized search coupled with cross-validation (Dayley, 2009; Olivier Grisel et al., 2011). Additionally, I merged a hyperparameter tuning procedure by extending the grid of the mtry parameter (the number of variables considered at each split) across a range of values based on the numberof bands of the sensor. Finally, the predictive performances of all models were assessed using the test data. Evaluation metrics, including overall accuracy, user and producer accuracy, F-score,

and confusion matrix, were computed following standard procedures (Foody, 2002).

3.5.5 eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGB) is a ML algorithm that is effectively used in different fields due to its accuracy, rapid computation, and sequential learning (Yin et al., 2023). It represents an ensemble algorithm constructed to upgrade model performance, especially on multiclass imbalanced datasets using boosting techniques, which allows its applicability in several fields, such as finance, healthcare, and image recognition (Pristyanto et al., 2023). Arif Ali *et al.* (2023) described XGB as a recently released ML algorithm that indicates exceptional capacities for modelling complex systems and superior prediction accuracy, interpretability, and classification versatility.

This algorithm demonstrated high performance when used to enhance heart attack prediction (Feng et al., 2024), human motion classification (Peng and Zhang, 2018), enhance ML model performance on multiclass imbalanced datasets (Pristyanto et al., 2023), and yield estimation (Huber et al., 2022), among other applications. XGB has also been frequently used in remote sensing and geological studies (Bhagwat and Uma Shankar, 2019; Samat *et al.*, 2020; Lin *et al.*, 2023; Shebl *et al.*, 2023). Based on these studies, XGB showed high accuracy and efficiency for mapping lithological units using either multispectral or hyperspectral remote sensing data. In this study, I demonstrated the XGB classifier to map ASM by applying a structured and reproducible approach. A grid search strategy was adopted to fine-tune major hyperparameters, for example, the number of boosting rounds, learning rate, maximum tree depth, and subsampling and column sampling ratios. The train function with the method set to "xgbTree" offered a complete inspection of XGB's functionality and calculated key metrics to determine the most critical variables for the classification procedure. The workflow I followed maintained the consistency of using the XGB model, demonstrating its capability for mapping ASGM areas and

associated land cover types. Finally, I applied the predict function to assess the model's performance using the test data for predicting the class labels.

3.6 Accuracy assessment

The process of testing models is an important step in all predictions (Bui and Mucsi, 2022). We utilized a previously developed module, programmed for an automated accuracy assessment, the Classification Assessment Tool (Szabó et al., 2024), which calculates accuracies using advanced solutions by taking random subsamples from the entire testing dataset based on a predefined ratio obtained from the testing data [0-1] and the number of repetitions of random sampling. We applied 0.7 for the fraction (70% of the data were used at a time with stratified random sampling), and for repetitions, we applied 10. Boxplot diagrams were used to visualize the differences between classes. The following class-level metrics were calculated: precision (or user's accuracy, UA), sensitivity (pproducer's accuracy, PA) (Barsi et al., 2018; Congalton, 1991), specificity (True Negative Rate), F1-scores (or Dice Similarity Coefficient), Jaccard Index (or Intersection over Union, IOU) (Willem, 2017; Grandini et al., 2020), and Matthews correlation coefficient (MCC) (Cao et al., 2020; Chicco and Jurman, 2020). The Kappa coefficient (κ) was also calculated to measure the level of agreement between the classified and reference data beyond chance (Congalton, 1991). It provides a robust statistical indicator of classification reliability, complementing the overall and class-specific accuracy metrics of the classification. Values of κ close to 1 indicate strong agreement, whereas values below 0.4 suggest weak or poor classification performance (Landis and Koch, 1977).

3.7 Pseudo labeling and statistical evaluation

For pseudo-labeling, users must select a probability map based on an accuracy assessment (e.g., the highest accuracy). A threshold value should be set to define pixels with a minimum probability of delineating the area where the newly labelled classes will be collected. Although the

threshold can be between 1 and 100%, only higher values (e.g., 0.9 [>90%]) are useful; accordingly, we chose 95%. This step assigned the relevant areas to the probability maps. We selected 1000 spatially random data points from the assigned area as PL-data (abbreviated as PL1000), which were involved in the new training phase merged with the original training data, and possible duplicates were removed. Next, the classification step was repeated with an accuracy assessment performed. Furthermore, we compared the class-level accuracy metrics (independent variables: Precision, Sensitivity, Specificity, MCC, F1, IOU) of the original maps with the map where the model was trained with the largest number (1000) of pseudo-labelled, resampled data (PL1000) (i.e., input datasets as dependent variables). A multivariate method, Hotelling's T-squared test, was used to test H0 (the group means for all independent variables were equal). The analysis was conducted using R 4.4.2 with the Hotelling package (Curran and Hersh, 2012). Beside the p-values, the effect sizes of partial $\eta 2$ and Mahalanobis Distance Squared (D2) were also determined to express the magnitude of the differences between the two values of the dependent variable. For partial $\eta 2$, 0.01-0.06 is considered small, 0.06-0.14 as medium, and >0.14 large effect; for D2, 0.25-0.50 is considered as small, 0.50-1.00 as medium, and >1.00 as large effect (Cohen, 2013; Matcharashvili et al., 2019; Shaker, 2023).

Maps produced with the original training and PL1000 data were compared using cross-tabulation and quantified by cross-entropy and visual analysis. Cross-entropy is a robust index for identifying hotspot areas of change (the higher the value, the larger the change) (Shim, 2024). Values below the upper quartile were blanked to enhance the relevant differences between the two maps, and the differences were quantified in the comparison (agreement) tables. The two maps were also compared using the Interspersion and Juxtaposition Index (IJI), which showed the isolation of the intermixing of patches (i.e., rock types) (Mead et al., 1981). Cross-entropy was determined using the ‘spatialEco’ package (Evans et al., 2023), and IJI was determined using the

'landscapemetrics' package (Hesselbarth et al., 2025). We also performed a difference analysis between the original and PL1000 maps, focusing directly on the changes and probabilities. We determined the probabilities of the classifications and the relationship between the accuracy metrics and the probabilities at the class level by rock type. We focused on the areas where the classification differed between in the two approaches, and investigated the probability of the pairs (e.g., in the original approach, a pixel was art, and in the PL1000 it was WDi). Finally, we compared the F1-scores and mean probabilities by rock type using the Spearman correlation.

3.8 Field Observations

Although remote sensing technologies provide robust analytical capabilities, ground truthing remains essential to ensure the accuracy and credibility of remote sensing–based interpretations. This helps mitigate potential errors in the reference data that could mislead assessments or weaken classification reliability (Sun et al., 2017). To validate our methodology and assess the performance of ML, remote sensing methods, fieldwork was conducted, and the results were compared with existing geological maps in areas #1 and #3 (Abu-Fatima et al., 2021; Kenea, 1997). The study areas were divided into a number of traverses depending on the size of the study area, and representative samples were collected to capture the full range of rocks, focusing particularly on locations where remote sensing study was applied, such as gossan signatures identified through our PRISMA-based analysis, Planet, and Landsat data. Field observations were performed in accordance with local regulations and no entry was made into private or protected lands. No human or environmentally sensitive data were collected, ensuring full compliance with the institutional research ethics and data-use guidelines.

Chapter 4: RESULTS

The results based on the subject and study sites are explained below. Although the used methods are similar, I focused on lithological mapping at site1 and ASM at sites #1 and #3, while hydrothermal alteration was at sites #1 and #2; therefore, the applicable methodologies and techniques should be independently interpreted. I included the discussion of the most significant findings per topic and study area, and later in the next chapter, I will highlight the important results in the conclusion section.

4.1 Lithological mapping results using ML in the study area #1

4.1.1 RF and SVM results for lithological mapping using Landsat and Sentinel-2 data

Six lithological units were classified: ophiolite, marble, metavolcanic, altered rock, and superficial deposits. For this classification, we used Landsat 8, Landsat 9, and Sentinl-2 acquired in June 2023 as the primary input datasets. Figure 4-1 display the accuracy levels observed across all rock units using the different sensors, RF, and SVM. After carefully reviewing the obtained accuracy assessment metrics and the six classified maps, we concluded that SVM was slightly better than RF. This further underlines the lithological unit with the best classification outcomes. According to the obtained accuracy metrics, the most accurately classified units with an accuracy above 95% were ophiolite, marble, and metavolcanic. Additionally, Landsat 8, Landsat 9, and Sentinel-2 provide offer adaptable sources for handling this task of detecting and mapping lithologies in complex, challenging, and highly deformed geological environments. As a classification data source, Landsat data are used more often than Sentinel-2 data. When we applied the classification using the same size of reference, data Sentinel-2 indicated higher overall accuracy, F-score, and user accuracy than Landsat 8 and Landsat 9 (Figure 4-2).

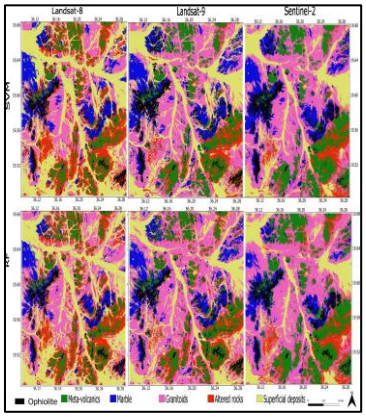


Figure 4-1. Represents six classified lithological maps obtained by applying RF and SVM on Landsat-8, Landsat-9 and Sentinel-2 datasets.

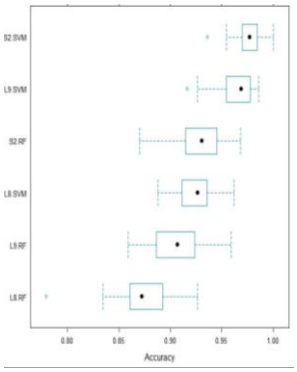


Figure 4-2. OAs obtained by SVM and RF for Landsat 9, Landsat 8, and Sentinel-2

4.1.2 Lithological mapping results using RF and NB applied on PRISMA and Landsat 9 data in study area #1

Section (4.1) described the potential of multispectral sensors for lithological mapping, implying that they offer a few number of bands because of the shortage of consistent wavelength continuation. Therefore, it is likely that a variety of features may be overlooked due to this limitation in band coverage. The introducing of hyperspectral data (PRISMA) could most probably be useful in avoiding these challenges. Certainly, there is a compromise in order to obtain this data, which could be cost, high-dimensionality, and computational time in comparison to the multispectral data. Therefore, we integrated both multi-spectral and hyperspectral data, a combination that is well suited to our research objectives.

Study area #1 is covered mainly by ophiolite, metavolcanic, marble, granitoids, altered rocks, and superficial deposits. The results showed that, PRISMA hyperspectral data obtained better classification results than Landsat 9 multispectral data using both classifiers. For example, the NB classifier provided Kappa values of 0.90 and 0.80, while the RF provided 0.95 and 0.90 for PRISMA and Landsat 9, respectively. Moreover, the OA was 0.96 and 0.92 for PRISMA and 0.92 and 0.83 for Landsat 9, respectively. Based on the classification results shown in the (Figure 4-3), the study area was mostly covered by metavolcanic. An ophiolite complex was located along the western boundary of the map area. The metasediments were mainly marble. It is discretely foliated, generally dipping NW, and brecciated and sheared in places. Altered rocks represent both hydrothermally altered rocks and areas where intensive artisanal activities are involved. Superficial deposits, which are scattered throughout the study area, are mainly recent sediments and drainage lines. The two datasets achieved reasonable overall classification accuracy (OA), though PRISMA yielded higher for both OA and Kappa (OA: 0.96 and 0.92) than Landsat 9 (OA: 0.92 and 0.83).

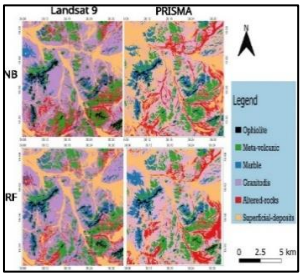


Figure 4-3. Represents six classified lithological maps obtained by applying RF and NB on Landsat-9 and PRISMA datasets.

4.2 Lithological mapping using Landsat 9 data and RF and MARS classifiers in light of probabilities and self-training in site #1

4.2.1 Geological maps with original training data

The MARS and RF algorithms provided seemingly similar maps of rock types; however, there were also significant differences. The main difference was in the case of Wdi, where MARS considerably overestimated WDi and underestimated all other types (Table 4-1). Considering the possible occurrence of WDi, the RF model was more reliable, as MARS indicated deposits at irrelevant locations (NW and NE corners of the area, (Figure 4-4). However, the class-level accuracy metrics showed different results: WDi exhibited the best performance with MARS. (Figure 4-5). In the case of other rock types, the accuracies were similar (or at least slightly, 2-3%, better with the RF), and the RF had a narrower range based on the repetitions of the testing procedure; that is, it acted more reliably with the existing testing data.

Table 2. Estimated area of rock types by two classification models: multiple adaptive regression spline (MARS) and RF. art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits.

Rock type	Area based on MARS model (km^2)	Area based on RF model (km^2)
art	54.09	58.98
gra	133.08	150.57
marb	47.56	60.64
mtvo	54.93	64.81
ophi	29.60	22.88
WDi	139.59	100.96

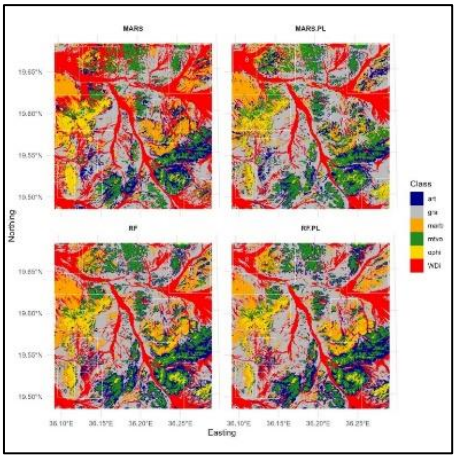


Figure 4-4. Geological maps of RF and Multiple Adaptive Regression Spline (MARS) models and their pseudo labelled (PL) versions (art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits; PL: 1000 pseudo label data with 95% probability).

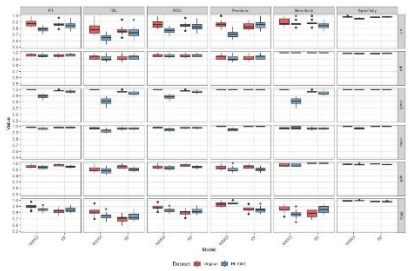


Figure 4-5. Accuracy metrics of RF and Multiple Adaptive Regression Spline (MARS) classifications (Original: models trained with the original training dataset, PL1000: models trained with an additional 1000 pseudo-labelled data sampled from 95% probability pixels; art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits; IOU: Intersection over Union, MCC: Matthews correlation coefficient).

4.2.2 Classification accuracies and probabilities

Regarding the probabilities, maps showed that MARS was supposed to be more accurate as 371.56 km^2 of the area had >99% probability, whereas in the case of RF, it was only 94.85 km^2 (Figure 4-6). Although near-100% pixels dominated the MARS model, it did not perform better; only the probabilities were overoptimistic, and the accuracy measures were only slightly better than RF. The median (derived from the accuracy assessment data) differences between the two models in the case of robust F1 and MCC did not demonstrate the superiority of any of the models: for art, marb, and ophi, RF performed better than MARS with 4.6-2.2-6.6% (MCC) and 4.0-1.8-5.9% (F1), whereas the MARS was shown to be a better model for gra, mtvo, and WDi with 2.9-1.4-12.2% (MCC) and 2.7-1.0-10.9% (F1).

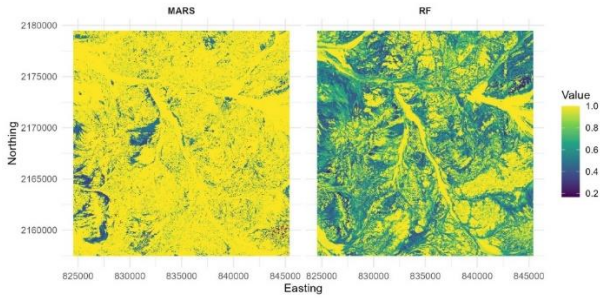


Figure 4-6. Probability layers of classification models of the Multiple Adaptive Regression Spline (MARS) and RF models

Class level accuracy and the F1-score had no connection with the two classifiers; the Spearman correlation coefficient was 0.21 ($p=0.51$, i.e., not significant). In half of the rock types, the RF performed better, and in the other half, the MARS performed better based on the F1-scores (Figure 4-7). Although the values theoretically followed a linear relationship, there were outliers in both models: in the case of $MARS_{art}$, the mean accuracy (F1) was one of the lowest (0.81), while the probability was 92%, and RF_{marb} had a large F1-score (0.98) with the lowest probability (58%).

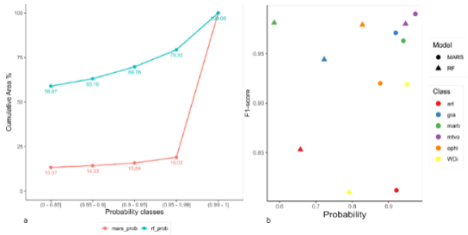


Figure 4-7. Cumulative probabilities and accuracies of the Multiple Adaptive Regression Spline (MARS) and RF models: a) proportions of pixels by probability thresholds; b) probabilities and F1-scores by rock type (art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits).

We compared the maps produced with the original and PL1000 (95% probability) training datasets and found that the original training set performed similarly to the case with an additional 1000 data points, at least on the level of accuracy metrics. Comparison matrices showed smaller agreements for MARS (art, marb, and ophi had <50%, WDi 58%) and slight differences for the RF (all rock types had >80% agreement, except marb having only 65%).

Visual analysis brought controversial observations: although both the RF and MARS maps differed by 25% from the PL versions regarding the hot-spot changing areas (in addition to the simple expression of the area, the cross-entropy levels differed significantly), the changes caused by the additional data improved the maps (Table 4-2). In area #1, the PL versions outperformed the original classifications for both RF and MARS; in area #2, RF.PL outperformed all other versions, and MARS.PL outperformed the original MARS. For area #3, MARS.PL provided the best outcome, but RF.PL was also better than the original version. In areas #4 and #5, the RF.PL was the most reliable solution. For area #6, MARS was the least accurate, and RF.PL provided the best solution for ophi, and MARS.PL for marb. MARS and MARS.PL mapped mtvo better than the other models, and RF.PL mapped WDi the best in the case of area #7. Generally, RF.PL was the best for area #8. In case area #9, RF.PL had the best classification for art and MARS.PL for mtvo and WDi. Accordingly, the PL model versions performed well, and based on the visual analysis, the spatial patterns were determined in several cases as the best outcomes. Although cross-entropy showed that the hot-spot areas of differences were 25% for both model pairs (original vs. PL), in the case of RF, the values were higher, indicating that there was a difference in sparse pixels; that is, the level of mixing of standalone pixels had changed (Figure 4-8). IJI confirmed that interspersion became more uneven with the RF.PL with 2.0% (77.2) related to the RF (79.2), while in the case of MARS it was 86.2, and for MARS.PL it was 89.5.

Table 4-2. Summary of best models based in visual inspection of spatial patterns (area codes depicted RF: Random Forest; MARS: Multiple Adaptive Regression Spline; PL: pseudo labelled; art: artisanal, gra: granite, marb: marble, mtvo: metvolcanic, ophi: ophiolite, WDi: wadi deposits)

Area	Characteristic rock type	Rock type	Best model
#1	ophi, WDi	generally	RF.PL and MARS.PL
#2	gra, WDi	generally	RF.PL
#3	gra, mtvo, WDi	generally	MARS.PL and RF.PL
#4	art, ophi, WDi	generally	RF.PL
#5	ophi, marb, gra	generally	RF.PL
#6	marb, mtvo, ophi	Ophi	RF.PL
#6	marb, mtvo, ophi	marb	MARS.PL
#7	art, marb, mtvo, WDi	mtvo	MARS and MARS.PL
#7	art, marb, mtvo, WDi	WDi	RF.PL
#8	gra, WDi	generally	RF.PL
#9	art, marb, mtwo, WDi	Art	RF.PL
#9	art, marb, mtwo, WDi	WDi, mtvo	MARS.PL

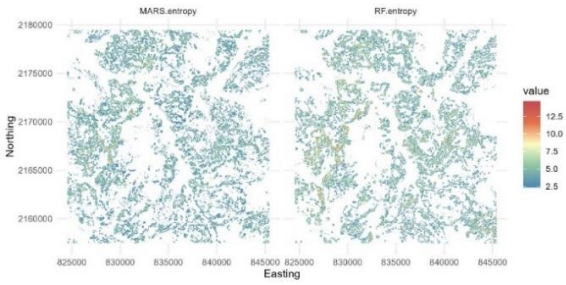


Figure 4-8. Differences of the pseudo-labelled (PL1000) to the original models using the cross-entropy (values <2.5, i.e., upper quartile, was blanked; RF: Random Forest, MARS: Multiple Adaptive Regression Spline)

The class-level evaluation of the probability levels in the two classification approaches showed that the classifiers reacted differently. In the case of MARS, the probabilities were initially high in the changing pixels of the rock-type pairs, and usually, the larger probabilities had lower SDs. A small decrease in the mean probability levels caused an increase in SD (Figure 4-9/a). The relationship between the mean probability and SDs was almost perfect (e.g., as a second-order polynom), but the marb-WDi pair was an influential data point with a low mean and SD. Although the changes in probabilities were not significant according to the Wilcoxon test, they exceeded 10% in 14 of 30 cases (Figure 4-9/b). In 12 instances, the probability increased. The results were different for RF; typically, lower mean probabilities had lower SDs and followed a linear relationship (Figure 4-10/a). The probabilities of the original approach were lower than those of MARS, and the probabilities of the PL versions were even lower, with significant differences (Figure 4-10/b). The number of cases in which the change was >10% was nine, and the number increased by only five. The magnitudes of the changes also differed; the maximum increase was 62% in the case of MARS, and 0.09 for RF (additionally, the largest change regardless of the direction of changes was only 0.27).

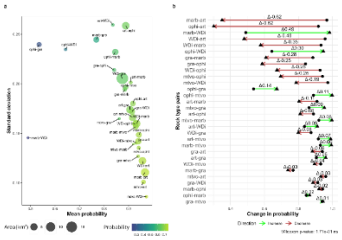


Figure 4-9. Changes of rock type classifications in light of probabilities in the case of MARS (Multiple Adaptive Regression Spline) classifier: a) original probabilities by changing rock types; b) magninudes of

probabilities of rock type pairs (art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits).

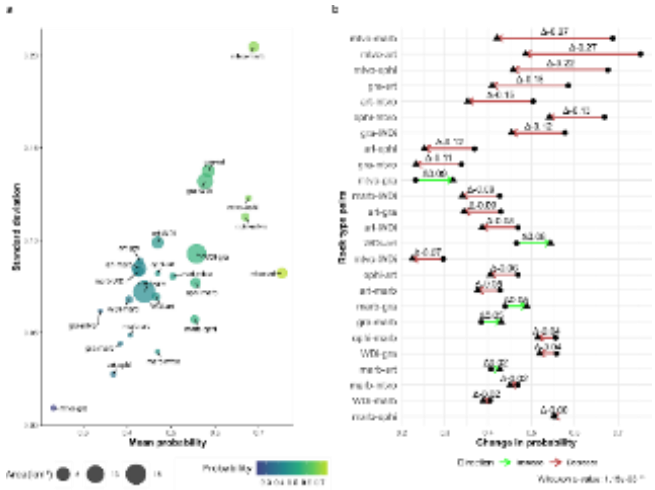


Figure 4-10. Changes of rock type classifications in light of probabilities in case of RF (Random Forest) classifier: a) original probabilities by the changing rock types; b) magnitudes of probabilities of the rock type pairs (art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits).

4.2.3 Multivariate comparison of the MARS models

The MARS models performed better with PL1000 than with the original training data, except for three out of 36 cases (without specific rock types). The original data provided more accurate results, and among the three exceptions, the mean differences were below 2% (mostly <0.5%). The decreases in F1, IOU, and MCC were up to 18.8%, with a maximum increase of 1.8%. For RF models, PL1000 was not very useful; nevertheless, it provided better metrics in 17 out of 36 cases, particularly for gra and WDi rock types (Figure 4-11).

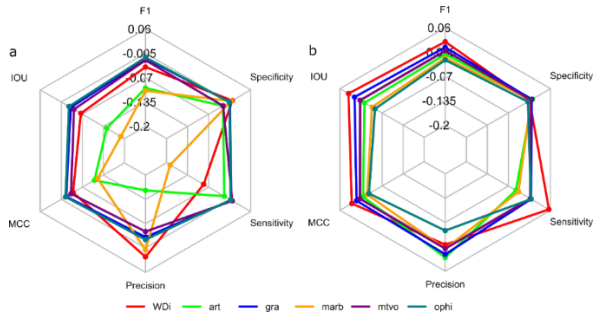


Figure 4-11. Difference in model accuracies (Original – PL1000 training data) by rock type and accuracy metrics a) MARS: Multiple Adaptive Regression Spline (MARS) and b) RF: Random Forest (IOU: Intersection over Union; MCC: Matthews correlation coefficient; art: artisanal, gra: granite, marb: marble, mtvo: metavolcanic, ophi: ophiolite, WDi: wadi deposits).

The difference between the models based on the original and PL1000 training data was significant according to the Hotelling test ($T^2 = 447.77$, $F(6, 593) = 74.63$, $p < 0.001$). $\eta^2 = 0.4302$ indicated a very large effect size and accounted for 43.02% of the multivariate variance by group differences in the independent variable. The large effect size was also justified by $D^2 = 0.746$. Both indicated a strong association between the grouping variables and the set of dependent variables. Effect sizes suggested that there were substantial differences in how the models performed across different datasets, considering all performance metrics simultaneously. Accordingly, the groups were well-separated in the multivariate space defined by the dependent variables.

4.2.4 Multivariate comparison of the RF models

For the RF models, the accuracy metrics showed varied results related to the MARS, and the PL1000 training dataset provided better accuracy measures than the original in 15 out of 36 cases. The increase was 5.5% and the maximum decrease was 4.8%. The Hotelling test

revealed a significant difference ($T^2 = 35.31$, $F(6, 593) = 5.89$, $p < 0.001$), but the effect sizes were not as large as in the case of MARS, indicating less pronounced differences between the groups. $\eta^2 = 0.056$ was close to the threshold for a medium effect (0.06), suggesting moderate significance of the group difference. D2=0.0588 indicated a relatively small separation between the groups; accordingly, the difference was statistically significant, but the effect was not large.

4.2.5 Discussion the results of lithological mapping

PRISMA and Sentinel-2 showed better spectral separability compared to Landsat data, especially between marble, metavolcanics, and altered rocks, while Landsat allowed clearer visual interpretation. Misclassifications mainly occurred between granitoids, altered rocks, and superficial deposits, reflecting their genetic relationships and shared spectral characteristics. Comparison with geological maps confirmed high agreement for ophiolite and metavolcanic units, though Sentinel-2 consistently produced the most reliable classifications. RF provided more realistic probability distributions, while MARS tended to overestimate classification accuracy. The PL method modestly increased classifier confidence and training sample size but did not consistently improve accuracy, likely due to geological complexity and limited testing data.. .

4.3 ASM mapping using remote sensing and ML algorithms

I used the RF and XGB classifiers due to their excellent accuracy, resilience, robustness and abilities to avoid overfitting with hyperspectral (PRISMA) and multispectral (Landsat 7, 8 and 9, Sentinel-2, and Planet) data. The application of hyperspectral PRISMA and very high spatial resolution Planet in ASM research, as well as for the XGB classifier, is pioneered by this work.

4.3.1 ASM Binary classification results

4.3.1.1 ASM Binary classification results Site #1 using original training data samples and Landsat 7, 8 and 9 data

The binary method produced a minimum OA of 0.88 for Landsat 9 and a maximum of 0.91 for Landsat 8, respectively (Appendix 4-1). The minimum OA for the multiclass method was 0.80 for Landsat 7 in 2003 and the maximum was 0.88 for Landsat 9 in 2023. The UA, PA, and F1-scores fluctuated between 0.85, 0.82 and 0.88 (minimum) and 0.93 (maximum), respectively, with the binary method (Table 2). However, in multiclass classification, the minimum values were less than 0.516, 0.611, and 0.610 for the three metrics. Nonetheless, we identified four lithological classes in areas with a relatively high class accuracy. The marble and metavolcanics classes provided the highest accuracy (0.93), whereas the ASM class achieved the lowest accuracy (0.516). In addition, superficial deposits showed an acceptable minimum accuracy of 0.765. By comparing the obtained classified maps (Figure 4-12) of Landsat 7, 8, and 9 data and field observations, we found that the binary classification resulted in a slightly (1-10%) higher accuracy than the multi-class classification.

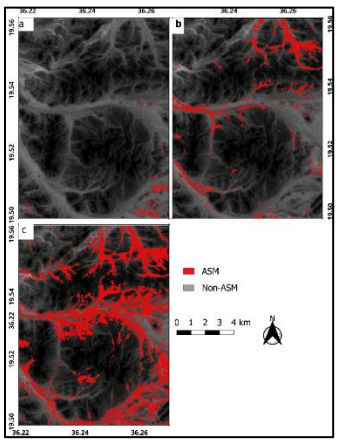


Figure 4-12. Distribution of ASM according to binary RF classification (a: 2003, b: 2013, c: 2023).

4.3.1.2 ASM Binary classification results in Site #3 using Planet data and Sentinel-2

The binary classification results derived from PlanetScope (2016, 2021, and 2024) (Figure 4-13) and Sentinel-2 (2017 and 2024) (Appendix 4-2) datasets revealed a clear and consistent pattern of ASM expansion over time in Site 3. Using the RF classifier, the PlanetScope data showed that ASM activities were initially limited to the northern part of the study area in 2016, occupying approximately 50 ha. By 2021, the mined area expanded to about 90 ha and further increased to 125 ha by 2024, spreading toward the central-eastern and southern sectors. Similarly, the Sentinel-2 binary classification results obtained from RF and XGB algorithms confirmed this progressive expansion. In 2017, ASM was confined mainly to the north-northeastern zones, covering less than 35 ha (RF) and 30 ha (XGB), but by 2024, the ASM footprint had extended markedly toward the southeast and northwest. The RF classifier showed a twofold increase in ASM extent between 2017 and 2024 (from 37 ha to 74.3 ha), whereas XGB indicated a more pronounced sixfold expansion, from 30 ha to approximately 166 ha. Both datasets consistently demonstrated that ASM development followed a spatially directional trend, originating from the northern and central-eastern zones and gradually expanding across the study area over the examined years.

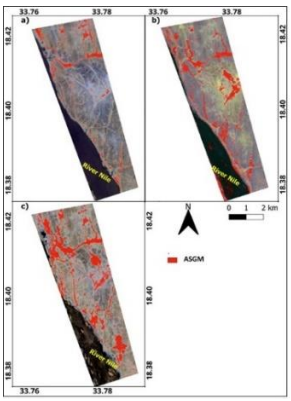


Figure 4-13. ASM expansion over time using binary classification (a: 2016, b: 2021, c: 2024)

4.3.2 ASM Binary classification results in Site #1 based on point-based and spatially explicit testing

4.3.2.1 ASM Binary classification results in Site #1 with Landsat imagery based on point-based and spatially explicit testing

MLAs produced varying outputs (Figure 4-14) and accuracies based on class-level metrics with the Landsat image however, the values generally exceeded 85%. The classifier performance for ASM site detection varied substantially depending on the training dataset. The RF classifier performed the worst with the original dataset ('orig'), showing a high false positive rate that was reflected across all accuracy metrics (Figure 4-15). In contrast, the SVM classifier, although not flawless, generated the most accurate map, whereas the MARS model performed mediocrely well. Interestingly, other findings indicated the opposite trend: RF achieved the best performance across datasets and even obtained near-perfect scores with the original data, indicating discrepancies arising from the collected reference data. All classifiers performed better when trained on the first extended dataset ('ext1'), with the SVM producing the most accurate classification. Despite having additional data, the second extension ('ext2') appeared to contain confusing spots that degraded classification quality. Neither visual interpretation nor accuracy metrics showed reliable ASM detection with 'ext2'. Combining the two enlarged datasets ('ext_12') resolved some issues and yielded better results than "ext2' alone though still far from optimal. The augmented dataset ('aug') performed marginally better than both 'ext2' and 'ext_12' although it did not outperform SVM trained on the original dataset. Nevertheless, all three models benefited from augmentation, which appeared to improve the overall classifier robustness. In contrast, the extended datasets (E.1 and E.2, i.e., ext1 and ext2) performed inconsistently, with several metrics, especially Sensitivity, dropping below the 0.85 benchmark, possibly indicating domain shift issues. In all cases, the classification of non-ASM areas performed consistently well (typically above 0.95), suggesting that all the models were generally more effective at identifying non-ASM regions than ASM sites.

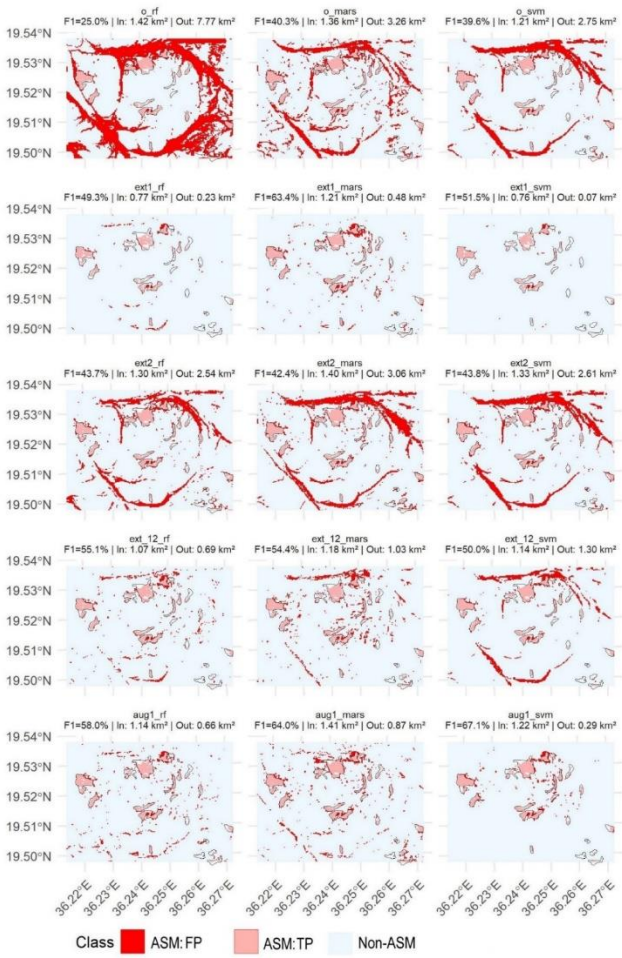


Figure 4-14. Artificial Small-scale Mining (ASM) sites of the study area using the Landsat image (rf: Random Forest, mars: Multiple Adaptive Regression Splines, svm: Support Vector Machine, orig, ext1, ext2, ext_12, aug: training datasets; ASM TP: ASM true positive, ASM FP: ASM false positive, F1: area-based F1-score, In: true positive area of ASM, Out: false positive area of ASM).

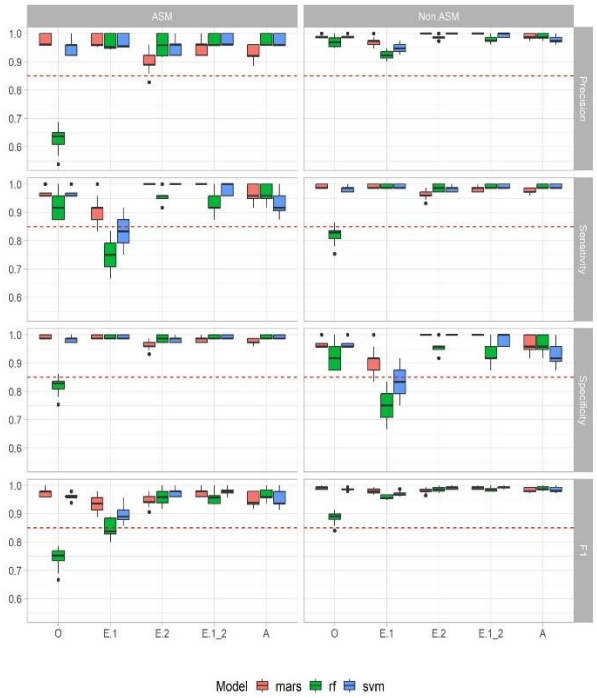


Figure 4-15. Accuracy metrics of the maps of Landsat image (ASM: ASMareas; rf: random forest, svm: support vector machine, mars: multiple adaptive regression splines; O: original training dataset with the least data, E.1: extended data, E.2: extended data, E.1_2: merge of E.1 and E.2, A: augmented data with the region growing algorithm; red dashed line: 85% accuracy benchmark)

4.3.2.2 ASM Binary classification results in Site #1 with PRISMA imagery based on point-based and spatially explicit testing

Visual interpretation showed that the best maps were obtained with the RF and SVM classifiers, particularly when larger training data were used (ext_12, aug) (Figure 4-16). The RF classifier trained with the ext_12 and aug datasets, as well as the SVM trained with the ext_12

dataset, provided the most accurate ASM site coverage with fewer false positives (Figure 4-17).

In terms of performance metrics, the original dataset 'O' (orig) yielded moderately good results, with considerable variability. In particular, precision and sensitivity had wide ranges that varied substantially across the classifiers. Metric values across classifiers ranged from 0.75 to 1.00. The extended datasets 'E.1' and 'E.2' (ext1 and ext2) demonstrated excellent performance, which consistently obtained values >0.90 for most metrics. The merged dataset 'E.1_2' (ext_12) showed the best overall performance with most metrics approaching near-perfect scores (0.95-1.00). This result indicates that combining external datasets provided complementary information that improved the model robustness, confirming that a larger training dataset is crucial for obtaining the best results. The augmented dataset 'A' (aug) performed marginally worse than the combined external dataset and showed greater variability, particularly in the Sensitivity metric. While all ASM test points were accurately identified, the extended datasets (E.1 and E.2, i.e., ext1 and ext2) performed well in terms of Sensitivity, showing no false negatives. Sensitivity was 8-10% lower for the augmented dataset, especially for the MARS and SVM classifiers, indicating that region-growing data augmentation might represent noise that prevents true positive detection. Overall, the E.1_2 dataset consistently obtained the highest F1-scores (>0.95), indicating an ideal balance between false positives and false negatives. The performance ranking was E.1_2 > E.2 ≈ E.1 > O > A, underscoring the importance of authentic external training data as the most critical factor for model accuracy.

From the perspective of classifier performance, notable differences were observed; however, the accuracy metrics generally exceeded 85% in nearly all cases. While RF classifier usually showed reliable performance, with both false positives (Precision) and false negatives (Sensitivity) adequately managed. Nevertheless, with the original dataset

('orig'), the RF exhibited the lowest Sensitivity among the three algorithms. The performance of the MARS classifier was weak in terms of Precision and Specificity, and required extended datasets to reach accuracy levels comparable to those of RF. SVM displayed the most varying performance across the three classifiers: it provided acceptable performance, but with notable fluctuations, indicating inconsistent false-positive control. The Sensitivity was the lowest with augmented dataset (A), where false negatives caused some issues. In contrast, the SVM performance was outstanding when trained with extended datasets (E.1, E.2).

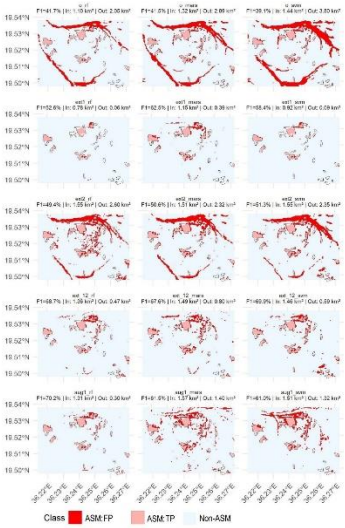


Figure 4-16. Artificial Small-scale Mining (ASM) sites of the study area using the PRISMA image (rf: Random Forest, mars: Multiple Adaptive Regression Splines, svm: Support Vector Machine, orig, ext1, ext2, ext_12, aug: training datasets; ASM TP: ASM true positive, ASM FP: ASM false positive, F1: area-based F1-score, In: true positive area of ASM, Out: false positive area of ASM).

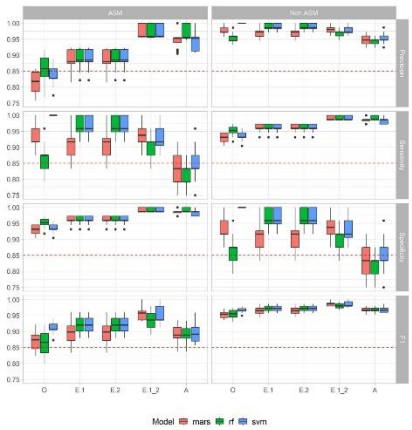


Figure 4-17. Accuracy metrics of the maps of PRISMA image (ASM: ASMareas; rf: random forest, svm: support vector machine, mars: multiple adaptive regression splines; O: original training dataset with the least data, E.1: extended data, E.2: extended data, E.1_2: merge of E.1 and E.2, A: augmented data with the region growing algorithm; red dashed line: 85% accuracy benchmark)

4.3.2.3 Accuracies in case of the Landsat image

For ASM detection, the RF models yielded the lowest mean accuracies, whereas MARS and SVM consistently achieved values of ≥0.9. The highest F1-scores belonged to the SVM classifier using the E.2 training set. Because our primary objective was to develop a tool capable of generating more accurate ASM maps, and the augmented training set (A) showed a relevant accuracy improvement, from 85% with E.1, the training dataset used for augmentation to 97% with A. For MARS and SVM, the augmentation also improved the performance, although to a lesser extent (MARS: ~1%, SVM: 5%). For the non-ASM class, the F1-scores were generally higher across classifiers. Augmentation had the strongest effect on the RF classifier, while the MARS performance was almost insensitive, and SVM was improved by ~2% (Figure 4-18). The GLM results showed that the model was significant and explained 41.6% of the variance (adjusted R^2=0.416), although the factors were not significant ($p>0.05$) (Appendix 4-3). Nevertheless, the effect size estimates (ω^2) indicated medium to large effects. Accordingly, classification accuracies were dependent on both the classifier type and

training dataset (medium effect), while their interaction showed the highest influence, showing a large effect ($\omega^2 = 0.179$).

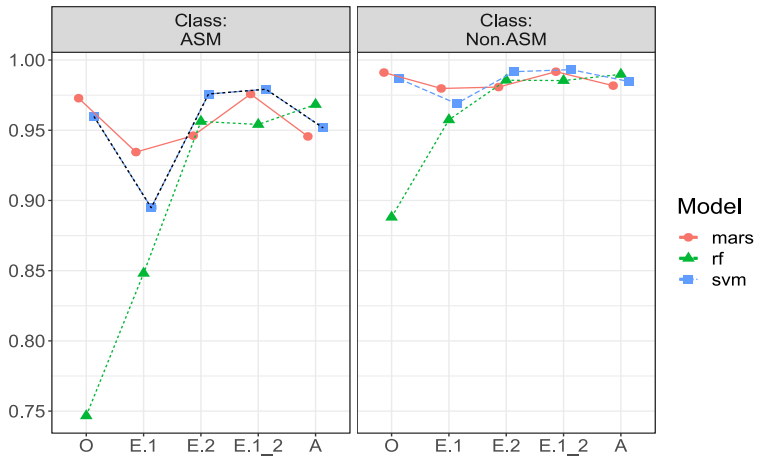


Figure 4-18. Mean F1-scores of the models derived from the Landsat image by ML models, and training datasets (ASM: ASMareas, O: original dataset, E.1: extended dataset #1, E.2: extended dataset #2, E.1_2: merged extended datasets, A: augmented dataset, mars: multiple adaptive regression splines, rf: Random Forest, svm: Support Vector Machine)

4.3.2.4 Accuracies in case of the PRISMA image

ASM-detection was not as accurate as the Landsat image, which also belonged to RF, at 0.86, which was 11% better than the worst Landsat value (Figure 4-19). The best results were just above 0.95. The training set with the largest number of training data, E.1_2, provided the best values, and the augmentation was not successful with the PRISMA image. For the non-ASM class, the accuracies were always >0.95; thus, the identification was more accurate. Regarding PRISMA, the GLM demonstrated that neither the training datasets nor the ML models significantly contributed to the explanation of the accuracies (Appendix 4-4). In most cases, all the models responded similarly across the different training datasets, and the effect size indicated "no effect".

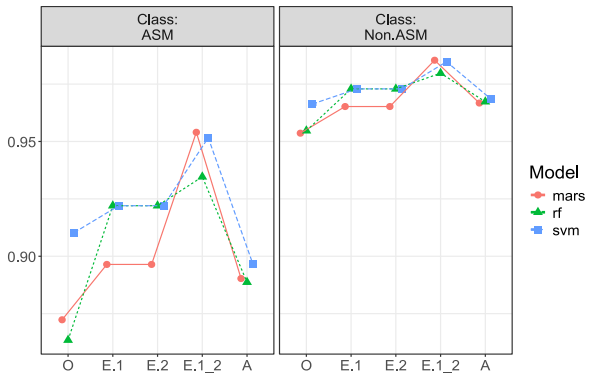


Figure 4-19. F1-scores of the models derived from the PRISMA image by ML models, and training datasets (ASM: ASMareas, O: original dataset, E.1: extended dataset #1, E.2: extended dataset #2, E.1_2: merged extended datasets, A: augmented dataset, mars: multiple adaptive regression splines, rf: Random Forest, svm: Support Vector Machine).

4.3.2.5 Comparison of point-based and area-based F1-scores

We also compared point- and area-based F1-scores and found relevant differences depending on both the training sets and ML algorithms. Statistical evaluation was not possible with the area-based metrics because of the limited amount of data (unlike the point-based F1-scores, there were a 10-times repetition, but randomization was not possible here). When points were spatially close, the performances were similar; however, when corresponding points had a large distance. For example, in the case of 'o' models with Landsat, the performance had large anomalies: o_rf had low F1-scores with both approaches, while the accuracy was better for o_svm and o_mars based on the point-based F1, but still low with the area-based one. For PRISMA, all 'o' models produced the weakest results for both F1 approaches. In contrast, with Landsat, the 'aug1' models had high values for both F1-scores. A similar trend was observed with PRISMA; however, in this case, the best values were obtained with the 'ext12' models (Figure 4-20).

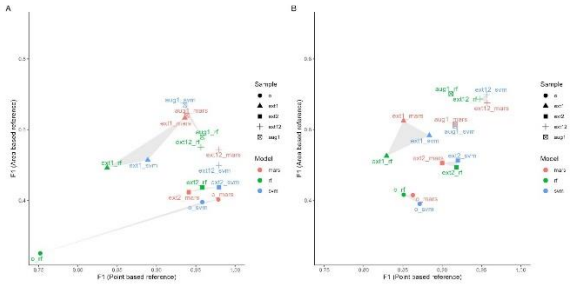


Figure 4-20. Accuracy of Artisanal Small-scale Mines (ASMs) based on point based and area based F1-scores using Landsat (A) and PRISMA (B) images (mars: Multiple Adaptive Regression Splines, rf: Random Forest, svm: Support Vector Machines; o: original dataset, ext1: extended dataset #1, ext2: extended dataclassifier1.2: merged extended datasets, aug1: augmented dataset).

4.3.2.6 ASM area and accuracy

Visual interpretation and area-based accuracy assessment showed that maps with the smallest ASM-area predictions had the most realistic underestimation and overestimation of ASM sites. The detected ASM area varied depending on the satellite data and applied ML algorithms, and the classification performance differed by class (i.e., ASM and non-ASM). However, the reasonable size of the identified ASM sites and point-based accuracy (F1-scores) did not correlate or did not correlate, as expected (i.e., high accuracy was not related to the most reasonable area extent; Table 4-3). We found that the two satellites had different patterns based on the training sets using the ML-methods. Thus, for MARS, and particularly for SVM, there was a significant positive correlation (Table 4-3), while RF showed no connection with Landsat. The RF showed no relationship in PRISMA, which is similar to what we experienced with Landsat. In contrast, MARS and SVM exhibited a strong negative correlation and a weaker positive connection, respectively, when applied to PRISMA. Thus, for PRISMA, lower (i.e., more reliable) values were associated with higher accuracies, whereas for Landsat, high accuracies did not necessarily guarantee reliable outcomes. The ext1_rf performed

best in the case of PRISMA, whereas ext1_svm performed excellently in the case of both satellites, according to the area-based accuracy evaluation (Figure 4-21). The worst models were the 'o' and 'ext2' models, using the original and second type of extension training set, with the worst F1-scores (<0.5) and SCR-scores (<0.5); thus, both types of evaluation showed weak performance. The 'ext1' and 'ext12' models usually were between SCRs of 0.5-0.85, indicating good performance, and the F1s were also >0.5. The developed segmentation-based algorithm, with the 'aug1' dataset resulted SCRs >0.5, and F1s >0.5, and in case of Landsat, the SVM, and in case of PRISMA, the RF had the second, and third best performance, based on the SCRs (Figure 4-21).

Table 4-3. Spearman' rhos of detected Artisanal Small-scale Mine (ASM) areas and F1-scores by satellites and ML classifiers (MARS: Multiple Adaptive Regression Splines, RF: Random Forest, SVM: Support Vector Machines)

Classifier	**Landsat**	**PRISMA**
MARS	0.43	-0.68
RF	-0.12	0.16
SVM	0.64	-0.37

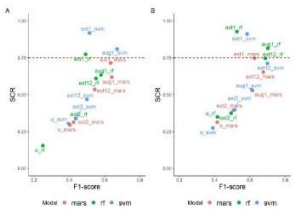


Figure 4-21. Spatial Completeness Ratio (SCR) and area-based F1-scores using Landsat (A) and PRISMA (B) satellites by classifiers (mars: Multiple Adaptive Regression Splines, rf: Random Forest, svm: Support Vector Machines; o: original dataset, ext1: extended dataset #1, ext2: extended dataset #2, ext1.2: merged extended datasets, aug1: augmented dataset; ---: 75% SCR level).

4.3.3 ASM results using multiclass RF classification

4.3.3.1 With Landsat data on study area #1 in 2003, 2013 and 2023

The minimum OA for the multiclass method was 0.80 for Landsat 7 in 2003 and the maximum was 0.88 for Landsat 9 in 2023. The UA, PA, and F1-scores fluctuated between 0.85, 0.82 and 0.88 (minimum) and 0.93 (maximum), respectively, (Appendix 4-5). However, in multiclass classification, the minimum values were less than 0.516, 0.611, and 0.610 for the three metrics. Nonetheless, we identified four lithological classes in areas with a relatively high class accuracy. The Marble and Meta-volcanics classes provided the highest accuracy (0.93), whereas the ASM class achieved the lowest accuracy (0.516). In addition, superficial deposits showed an acceptable minimum accuracy of 0.765. By comparing the obtained classified maps (Figure 4-22) of Landsat 7, 8, and 9 data and field observations, we found that the binary classification resulted in a slightly (1-10%) higher accuracy than the multiclass classification.

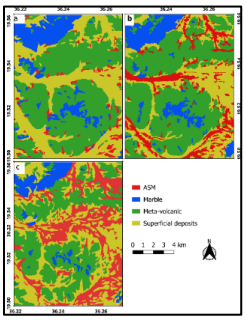


Figure 4-22. Distribution of ASM according to multiclass RF classification (a: 2003, b: 2013, c: 2023).

4.3.3.2 ASM Multiclass Classification Results in Site #3 Using PlanetScope and Sentinel-2 Data

The details about this part in (Elrasheed et al., 2025). The multiclass RF classification results obtained from PlanetScope (2016, 2021, and 2024) (Figure 4-23) and Sentinel-2 (2017 and 2024) (Figure 4-24) datasets consistently showed a sharp spatial expansion of ASM activities

in Site 3. Utilizing PlanetScope data, ASM was initially focused in the central-eastern and northern parts of the study area in 2016, where mining mainly targeted meta-andesite and, to a lesser extent, meta-basalt and adjacent stream sediments. By 2021 and 2024, the extent of ASM had increased considerably, expanding toward the eastern and western sectors while remaining dominantly associated with meta-andesite outcrops, reflecting the continued exploitation of gold-bearing metavolcanic rocks. Similarly, the Sentinel-2 multiclass RF classification results corroborated the observed temporal expansion between 2017 and 2024. In 2017, ASM sites were primarily confined to the north and central-eastern parts of the area, where miners exploited meta-andesite and surrounding alluvial channels. By 2024, ASM activity had spread extensively to the southern and northwestern zones, with a pronounced concentration over meta-andesite units, as evident in the comparative lithological distribution between 2017 and 2024. These findings align with regional studies, which reported a significant increase in ASM activities across sub-Saharan Africa particularly in Sudan driven by economic necessity and livelihood opportunities. Overall, the multiclass classification maps effectively delineated the spatial evolution of ASM and highlighted the preferential targeting of metavolcanic lithologies as the principal host for gold mineralization.

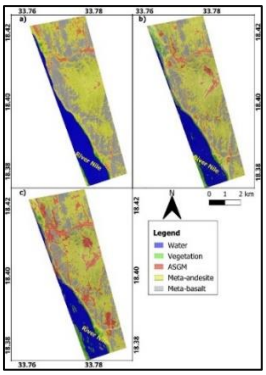


Figure 4-23. ASM expansion over time using multiclass classification with Planet (a: 2016, b: 2021, c: 2024)

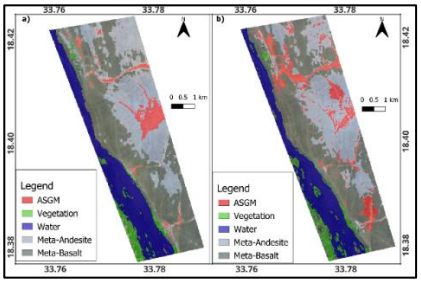


Figure 4-24. Distributions of ASM according to multiclass RF classification with Sentinel-2 in a) 2017 and b) 2024

4.3.4 ASM change analysis

4.3.4.1 ASM change analysis Landsat data in the study area #1

The binary classification revealed that, in 2003, the study area was almost free of ASM activity, occupying an area of 30 ha. The area witnessed an exponential increase in ASM zones in the last 20 years: in 2013, it was 8.3-times more (265 ha), and in 2023, it was 33-times more (992 ha) than it was in 2003. The results also showed that the ASM started in the southeastern part of the study area in 2003; in 2013, this activity shifted to the northeast and west sectors. In 2023, the ASM dominated the central part according to the distribution of gold-bearing rocks (Figure 4-25). The binary method's finding, that ASM areas are increasing over time, was supported by the multiclass classification results, which also showed which rock units ASM has been targeting over the past 20 years. ASM areas were slightly different from the multi-class approach: 31 ha in 2003, 204 ha in 2013, and the same in 2023 (992 ha). ASM activity started to involve some portions of marble and meta-volcanic deposits in 2013, whereas in 2003, it was limited to stream sediments (superficial deposits), according to information obtained from the lithological units. The highest change was found between Superficial and ASM (501 ha) between 2003 and 2013, and Meta-volcanic and ASM (77 ha), and the lowest was between Marble and ASM (420 ha) (Figure

4-26). Between 2003 and 2023, spatial analysis revealed that 77 ha of meta-volcanic rock areas were transferred into ASM, while only 4 ha of marble areas experienced conversion. In addition, notable exchanges were observed between ASM and superficial deposits (Appendix 4-6) and (Table 4-4).

Table 4-4. Areas (ha) of Lithologies Transferred to ASM across the Study Period.

Lithologies transferred to ASM	Area (ha) 2003-2013	Area (ha) 2003-2023	Area (ha) 2013-2023
Marble - ASM	4.20	28.23	9.11
Meta-volcanic - ASM	77.19	218.45	145.02
Superficial deposits- ASM	501.37	748.61	506.73
ASM	40.86	79.51	413.92

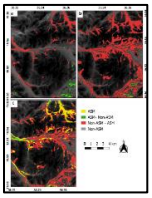


Figure 4-25. ASM area changes identified by the binary RF classification in pairings of dates (a: 2003-2013, b: 2003-2023 and c: 2013-2023)

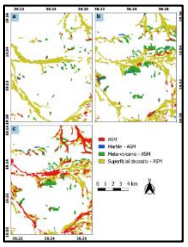


Figure 4-26. ASM related rock types over time detected by the multi-class approach (a: 2003-2013, b: 2003-2023 and c: 2013-2023)

4.3.4.2 ASM Change Analysis Using PlanetScope and Sentinel-2 Data in Study Area 3

Multiclass RF classification of PlanetScope (2016–2024) and Sentinel-2 (2017–2024) data revealed a clear expansion of ASM activities in Study Area 3 (Appendix 4-7) and (Appendix 4-8). Using PlanetScope imagery, ASM areas increased from 50 ha in 2016 to 125 ha in 2024, mainly at the expense of meta-andesite (75 ha) and meta-basalt (21 ha), while non-ASM areas declined (Figure 4-27). Sentinel-2 results showed similar trends, with ASM expanding from 46.4 ha in 2017 to 85.2 ha in 2024, primarily converting meta-andesite and meta-basalt (Figure 4-28). Minor interchanges between these metavolcanic units (27.3 ha and 18.2 ha, respectively) were attributed to both mining-induced disturbance and spectral similarity. In both datasets, transitions between ASM and vegetation or water were negligible, confirming that ASM expansion occurred mainly on exposed rocky terrains rich in meta-andesite, the primary lithological target of miners.

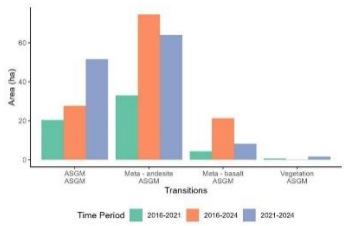


Figure 4-27. Geological units and land cover transitions to ASM in the years 2016, 2021, and 2024 of Planet data

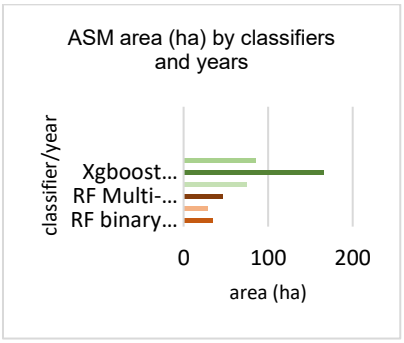


Figure 4-28. Bar graph showing OF ASM area in hectares in 2017, and 2024 from change analysis of binary and mult-class XGB and RF classification results of Sentinel-2.

4.4.1 Field observations of ASM activities and its impacts

ASM activities have a relevant effect on the study area, accelerating land degradation, obliteration of important geological features, and obstruction of natural drainage systems. Based on the field observations, the general structural trend of the region indicated a predominant strike direction from N to NE, which was also observed in the Landsat images, particularly in 2003 and 2013. In 2023, due to the expansion of ASM sites, these features were not always obvious. Besides, the field observations provided more nuanced insights into the geological characteristics of the region: we found several strike directions, including NE, E, and SE with dipping angles varying from approximately 80° in ophiolitic units to as low as 10° within areas affected by ASM activities. The meta-volcanic rocks displayed an even broader range of strike directions (N, NE, E, SE, and S) and dipping angles, with maximum dips (80°) predominantly occurring in regions unaffected by ASM in the central part of the study area and minimum dips (10°) observed in ASM-impacted zones in E-NE. Furthermore, field studies identified considerable discontinuities in the lithological units and mineralized

quartz veins, particularly in areas of intensive ASM activity. For example, in the western part of the study area (Figure 4-29) we found disruptions to the mineralized quartz veins cutting through the meta-volcanic rocks, which are critical geological features, appeared and disappeared intermittently during the mapping. Although geological deformation may be the cause of this discrepancy, artisanal miners' removal and exploitation of the quartz veins, who target these structures for their mineral content, is mostly responsible in this instance. (Figure 4-30) illustrates distinct discontinuities and variability in strike and dip angles of ore bodies in areas impacted by ASM activities. These features are spatially associated with ASM zones, where critical geological structures such as dip, strike and joints exhibit observable alterations. Field observations corroborate these findings, documenting noticeable degradation of such features within ASM-affected areas. This evidence indicates localized structural modifications directly associated with ASM extraction processes.

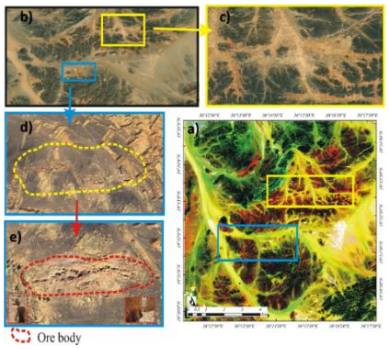


Figure 4-29. Structural changes caused by the ASM (a: Landsat 9 image of the area; b: high resolution Google Earth image of ASM sites; c: deliberation of the geological features (strike, dip and joints) of meta-volcanic as a result of intensive ASM activities; d and e: relevant disruptions discontinuity of the mineralized quartz veins caused by ASM).

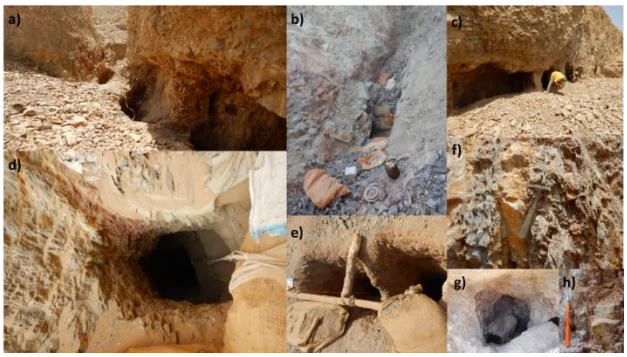


Figure 4-30. Photographs of ASM sites in the area (a and b: digging within a shallow pit ASM site where miners are excavating a mineralized quartz vein; c: ASM mining spot where a highly deformed, mineralized quartz vein is exposed, scattered rock debris indicating active excavation; d and e: deep mining pit with rudimentary wooden structures used by the artisans to extract ore; f, g and h: a deep pit in a meta-volcanic rock used to mine selectively the mineralized quartz veins and veinlets zones).

ASM operations in the area relevantly affected the landscape, geology, soil, drainage system, and water. Field observations across the study area revealed extensive environmental degradation resulting from ASM activity. Evidence collected from field observations demonstrates relevant impacts on land stability, natural drainage systems, and geological integrity. Mining activities utilize basic wooden support structures within excavated pits, reflecting inadequate safety protocols that endanger workers while simultaneously causing severe landscape deformations (Figure 4-31). Systematic removal of surface materials fundamentally alters the original topography, leading to areas of increased instability and accelerated erosion. According to our observations near the Nile, adjacent to mining sites, there are extensive debris fields and topographic disruptions with significant soil stripping and sediment displacement patterns. These alterations substantially compromise land-use sustainability and ecosystem functions in the affected regions. Examination of the collapsed mining pits revealed exposed subsurface geological layers, demonstrating the destruction of current exploitation practices. ASM systematically weakened the

structural integrity of the soil and increased the land subsidence risk, thereby forming environmental vulnerability zones. Abandoned mining locations exhibited advanced stages of erosion and structural instability, providing clear evidence of the long-term geomorphological consequences associated with ASM activities. These degraded sites represented persistent sources of sediment runoff that directly influence the quality and ecological balance of adjacent water bodies and terrestrial ecosystems.

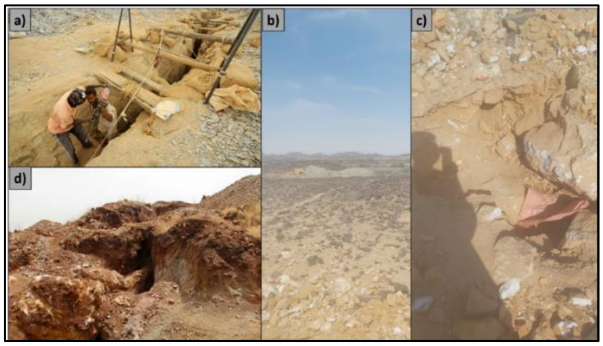


Figure 4-31. Environmental issues of artisanal small-scale mining (a: mining pit, b: scattered debris, c: collapsed mining pit, d: abandoned mining site)

Ore washing and panning relied on constructed artificial pools where miners concentrated sediment-laden materials, leading to localized pollution that adversely affected both terrestrial and aquatic ecosystems. Mechanical ore processing involving crushing operations generated substantial particulate matter emissions, and the resulting dust clouds contain respirable mineral particles that pose health risks, particularly respiratory complications and pulmonary disorders, for exposed workers. Airborne dust contaminants contributed to both immediate health hazards and long-term chronic exposure. Waste management at ASM sites was also insufficient: illegal activities did not require standards; consequently, mining facilitated the direct infiltration of toxic materials into subsoil layers and groundwater as persistent contamination sources and generated long-term ecological degradation. Hg amalgamation processes represented the most hazardous aspect of gold extraction,

involving the direct manual manipulation of elemental Hg in aqueous slurries without appropriate protective equipment. This exposure posed neurological health risks to workers while also introducing Hg contamination into the local environment, raising both acute toxicity concerns and chronic bioaccumulation potential (Figure 4-32).

Figure 4-32. Gold extraction practices in ASM areas (a: miners around a small artificial pool filled with water for ore washing and panning; b: breaking rocks generates dust clouds; c: improper waste disposal; d: manual amalgamation of gold using mercury in a pan filled with muddy water).

Hydrological changes in the study area showed systematic alterations in natural drainage due to the ASM. The observed modifications in the drainage patterns and sediment transport indicated that these interventions disrupted the flow regimes and sediment deposition processes. The ASM and associated waste accumulations were located near existing drainage networks, leading to hydraulic interference that increased the sediment loading and flow-obstruction mechanisms. The spatial relationship between extraction sites and natural watercourses has resulted in conditions that accelerate sedimentation rates and reduce the channel capacity. These anthropogenic changes fundamentally influenced watershed functionality by disrupting established flow pathways and introducing artificial barriers to surface water movement (Figure 4-33).

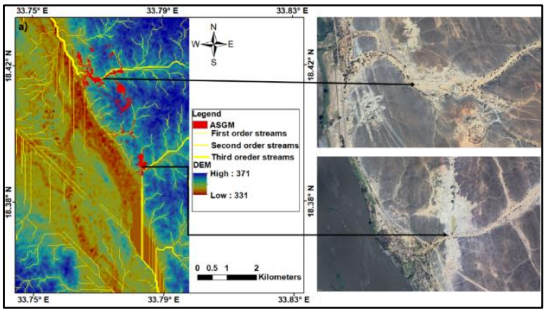


Figure 4-33. Drainage systems with potential impacts of artisanal and small-scale gold mining (ASM) operations (a: terrain height and drainage system with highlights of ASM activities; b: high-resolution Google Earth imagery with regions where intense mining activity disrupts the drainage system).

4.3.5 Discussion the Results of ASM Classification and Change Detection

ASM practices are receiving increasing focus coming from all around the world as crucial source of revenue and a major cause of adverse environmental impacts, in particular in geographically varied, vulnerable, and wealthy in minerals areas with frequently inadequate laws and regulations. In line with the results of earlier research, our outcomes showed substantial increases in ASM area during our study period 2003-2024. Similar increase in ASM tracks were reported in Brazil and Ghana by Barenblitt et al. (2021), Gallwey et al. (2020b), and Simionato et al. (2021), who connected the loss of forest, destruction of habitat, water supply issue and food availability to ASM growth. This consistency seen in the ASM in various regions emphasize how ASM serves as a universal trigger for environmental and socioeconomic transformations. According to (Ahmed et al., 2019; Fadlallah et al., 2020), the reported growth in ASM sector is coincided with wider demographic patterns. Factors such as rapid growth in population (Frisbie, 2001) around mineral rich area has turned as their basic and alternative source of income. Notably, the recent increase in gold prices (in 2006 gold was 14000 USD/kg, and from 2010 it became 38000

USD/kg, and after 2020 the price is >62000 USD/kg) (Price, 2024). The economic shock triggered by South Sudan's independence, which resulted in Sudan losing nearly 80% of its oil resources, severely disrupted the national economy. The subsequent financial crisis led to widespread unemployment, particularly in industries that were heavily dependent on oil revenues. Simultaneously, declining agricultural productivity in rural central Sudan further eroded traditional livelihoods, compelling many communities to seek alternative sources of income sectors (Ahmed et al., 2019; Fadlallah et al., 2020). ASM emerged as an attractive option due to its low entry barriers for unskilled labor. More recently, the ongoing war since 2023 has exacerbated economic instability and displacement, driving even more individuals into ASM as one of the few remaining viable means of survival, thereby accelerating its expansion. Our study area considered as a metallogenic-rich province with a high potential for future mineral exploration. However, due to the ASM increases its scientific (tectonic history) and economic potentiality are notably in danger (Abu-Fatima et al., 2021). Moreover, uncontrolled ASM operations the key geological properties (bedding plane, rock boundaries, structures, joint and orebody orientations). Besides, the obliteration or disruption of natural drainage system and potentially soil degradation and water resources (Nile) contamination.

The classification analyses from Landsat, PlanetScope, and Sentinel-2 datasets consistently revealed a substantial and progressive expansion of artisanal and ASM across the study sites during the last two decades. The binary and multiclass random RF classifications demonstrated that, despite slight variations in accuracy, both approaches effectively captured the spatial and temporal evolution of mining activities. For Site 1, binary classifications achieved high overall accuracies (OA) ranging from 0.88 (Landsat 9) to 0.91 (Landsat 8), outperforming the multiclass models (OA = 0.80–0.88). User's accuracy (UA), producer's accuracy (PA), and F1-scores for the binary approach ranged between 0.82 and 0.93, confirming its robustness in delineating ASM zones. Conversely, multiclass results displayed lower accuracies, particularly for the ASM class (0.516), while lithological classes such as marble and meta-volcanics showed higher reliability (up to 0.93). Field verification and visual comparison of the classified maps confirmed that binary RF classification produced 1–10% higher accuracy than the multiclass method. Change analysis based on Landsat imagery (2003–

2023) revealed that ASM areas expanded dramatically from 30 ha in 2003 to 992 ha in 2023, representing a 33-fold increase. The spatial trend indicated that mining initially appeared in the southeastern sector before migrating northward and westward, ultimately dominating the central zone, in close correspondence with gold-bearing lithologies. The multiclass classifications corroborated these findings and revealed the lithological transitions underpinning ASM expansion. Between 2003 and 2023, approximately 77 ha of meta-volcanic rocks and 4 ha of marble were converted to ASM, while the most substantial changes occurred between ASM and superficial deposits (501 ha), reflecting intense surface disturbance and sediment reworking. The PlanetScope dataset further substantiated this expansion trend with higher spatial detail. Binary RF classifications indicated that ASM areas increased from 50 ha in 2016 to 125 ha by 2024, while non-ASM areas correspondingly declined. The multiclass RF yielded similar temporal patterns and showed that miners predominantly targeted meta-andesite units, followed by meta-basalt. Interclass transition analysis revealed that 75 ha of meta-andesite and 21 ha of meta-basalt were converted into ASM areas during 2016–2024, underscoring the preferential exploitation of these lithologies.

Comparable findings were derived from Sentinel-2 data for Site 3. Binary RF and XGB models indicated a marked increase in mining extent between 2017 and 2024—from 34.5 ha (RF) and 28.2 ha (XGB) to 74.3 ha and 166.2 ha, respectively. Both classifiers confirmed continuous ASGM expansion toward the southeast and northwest, though the RF model demonstrated superior precision and reliability, leading to its adoption for multiclass analysis. The multiclass RF outputs reaffirmed the growth trajectory, showing an increase from 46.4 ha in 2017 to 85.2 ha in 2024. Lithological change detection indicated that 30 ha of meta-andesite and 27 ha of meta-basalt were transformed into ASM zones. Furthermore, bidirectional conversions between meta-andesite and meta-basalt (27.3 ha and 18.2 ha, respectively) reflected both the physical alteration of the geological environment by mining and spectral confusion between these closely related metavolcanic units. Minimal changes were observed between ASM and non-geological classes such as vegetation and water, indicating effective class separability for these features. Overall, the results from all sensors and classification approaches converge on a consistent conclusion: ASM has expanded

rapidly and systematically, primarily targeting metavolcanic lithologies (meta-andesite and meta-basalt), with associated degradation of superficial deposits and marble zones. The binary classification method generally outperformed the multiclass approach in accuracy metrics, while the multiclass analysis provided critical insights into lithological interactions and class transitions. This integration of multi-sensor datasets (Landsat, PlanetScope, and Sentinel-2) robustly demonstrates that ASGM activity has significantly intensified over the last two decades, transforming both the geomorphological and geological integrity of the study region.

Our analysis also reveals the paradoxical role of ASM: it is a primary agent of environmental degradation yet also acts as an inadvertent exploration tool. The stripping of deep overburden by ASM has exposed hydrothermally altered shear zones that were previously concealed, making pre-existing orogenic gold mineralization detectable. This underscores a dynamic where mechanical disturbance, rather than new mineralization, drives the visibility of alteration footprints. Based on our outcomes, we suggest a thoroughly strategy for ASM managing in Sudan, drawing from the successful elements of other countries' policies. First, there is an urgent need for awareness campaigns targeting miners and local communities to highlight the harmful effects of Hg use, land degradation, and the importance of sustainable practices. This is particularly vital given Sudan's lack of formal safety and environmental regulations, unlike countries such as Ghana and Tanzania, which mandate licenses and promote legal ASGM operations through structured frameworks such as Community Mining Schemes (CMS) and Primary Mining Licenses (PML)(Fisher, 2007). Second, Sudan needs to develop and implement coherent land management and mining policies that designate specific zones for mining, agriculture, and habitation to reduce environmental conflict, as seen in Indonesia's designated 'People's Mining Areas' under Law No. 4/2009. Integrating high-resolution satellite imagery such as PlanetScope into national ASGM monitoring systems would considerably improve the detection of mining expansion, similar to Brazil's use of remote sensing for formal oversight via the National Mining Agency (de Theije and Heemskerk, 2009). Finally, sustained collaboration among government institutions, environmental experts, and local communities is essential for enforcing regulations and achieving long-term sustainability, a model adopted in

Chile and South Africa, where government agencies engage miners through technical support and inclusive policy reforms (Fritz *et al.*, 2018; Schwartz et al., 2021). Given Sudan's current informal mining structure and fragile regulatory framework, these comparative models offer valuable policy pathways for formalization, sustainability, and environmental protection. Table 4-4. summary of ASM classification and temporal changes.

Table 4-4. Summary of ASM classification results and temporal changes

Dataset	Classifier(s)	Training Strategy	ASM Area (ha)	Observed Trend / Notes
Landsat 7 (2003)	RF	Point-based	30	ASM initially confined to SE; minimal activity.
Landsat 8 (2013)	RF	Point-based	265	8.3× increase since 2003; expansion to NE and W.
Landsat 9 (2023)	RF	Point-based	992	33× increase since 2003; dominance in central area near gold-bearing rocks.
Landsat (All)	SVM, MARS, RF	Spatially explicit (orig, ext1, ext2, ext_12, aug)	------	SVM best overall; augmented dataset improved accuracy (RF: +12%, SVM: +5%).
PRISMA (Site #1)	RF, SVM, MARS	orig, ext1, ext2, ext_12, aug	--------	ext_12 yielded best accuracy (>0.95); augmented slightly lower.
PlanetScope (2016)	RF	binary classification	50	ASM concentrated in north.
PlanetScope (2021)	RF	binary classification	90	Expansion toward central-eastern sector.
PlanetScope (2024)	RF	binary classification	125	Expansion toward south and west; meta-andesite targeted.
Sentinel-2 (2017)	RF, XGB	binary classification	RF = 37 ha; XGB = 30 ha	ASM confined to north-northeast.
Sentinel-2 (2024)	RF, XGB	binary classification	RF = 74.3 ha; XGB = 166 ha	Strong expansion toward SE and NW; RF more reliable.
Landsat (2003–2023)	RF	change analysis	30 → 992	33× expansion; major transitions from superficial deposits (501 ha) and metavolcanics (77 ha).

Dataset	Classifier(s)	Training Strategy	ASM Area (ha)	Observed Trend / Notes
PlanetScope (2016–2024)	RF	change analysis	50 → 125	+150%; conversions: meta-andesite = 75 ha, meta-basalt = 21 ha.
Sentinel-2 (2017–2024)	RF, XGB	change analysis	34.5 → 74.3 (RF), 28.2 → 166 (XGB)	Bidirectional conversion between meta-andesite and meta-basalt (27–18 ha).

4.4 Hydrothermal alteration mapping

4.3.1 Mapping hydrothermal alteration zones in ASM area #1 using Landsat 7, 8,9, Planet, Sentinel-2 and PRISMA

BR findings of Landsat/ Sentiel-2, Planet and PRISMA showed that the anomalies of alteration zones are area well mapped in the study areas1 and 2. In the study area 1 for Landsat 7 (2003), Fe-oxide (band3/band1) anomalies were detected to the north-western study area, while OH-bearing minerals (band5/band7, band5/band4) indicated signals in the central east part. Landsat 8 and Landsat 9, Fe-oxide zones mapped in the middle-north, east, and southeast, correlating, while OH-bearing minerals discovered in the central-east and later extended north and west and south (Appendix 4-9) . Sentinel-2 revealed capacity for mapping Fe-oxide (b4/b2, b6/b1) and OH-bearing (b11/b12) anomalies (Appendix 4-10). Planet data was also mapped Fe-oxide mapping (b3/b1, b6/b2, b7/b2) (Appendix 4-10)because of SWIR bands lack. Similarly, PRISMA BR and RBD demonstrated mineral alteration (alunite, jarosite, Al-sheet silicates, chlorite, epidote, Fe-bearing minerals and Oh-bearing minerals) (Figure 4-34).

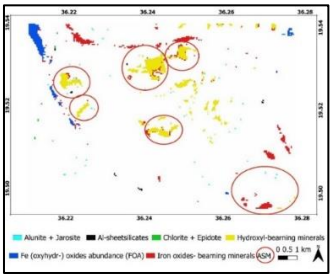


Figure 4-34. The distribution of hydrothermal alteration zones detected using BR and RBD of PRISMA data

4.3.2 Directional Principal Component Analysis

Directional Principal Component Analysis (DPCA) results from Landsat 7, 8, 9 (Figure 4-35), and sentinel-2 (Figure 4-36) revealed its effectiveness in detecting alteration zones. We obtained four PCAs corresponding to utilized bands and general for all sensors, PC1 largely represents overall brightness with uniformly positive loadings. Landsat 7 (2003) PC3 exhibits a pronounced contrast with a high positive loading for band 1 and a strong negative loading for band 3. Although the sign is arbitrary, if we invert PC3, the red band (band 3) would have a high positive contribution relative to the other bands. This contrast is particularly useful because iron oxide minerals typically manifest distinct spectral behaviour in the red region relative to other bands. Hence, PC3 is elected as the best component for enhancing Fe-oxide signatures. For hydroxyl-bearing minerals, the spectral signature of interest is the differential absorption in the SWIR region. While PC1 again mainly captures overall brightness, PC3 shows a strong negative loading on band 5 (SWIR1) and a positive loading on band 7 (SWIR2). This creates a distinct contrast that emphasizes the absorption features associated with hydroxyl groups. For iron oxide mapping using Landsat 8 (2015) and Landsat 9 (2024), PC3 was selected as the optimal component due to its distinct spectral contrast most notably, the pronounced positive loading

in band 2 (blue) that contrasts with the strong negative loading in band 4 (red). Specifically, Landsat 8's PC3 exhibits loadings of 0.763761 (band 2) and –0.593058 (band 4), while Landsat 9's PC3 shows loadings of 0.682095 (band 2) and –0.623625 (band 4).

The high positive value in band 2 enhances the contrast against the red band's negative response a diagnostic indicator of iron oxide mineralization thus effectively isolating the spectral signature of iron oxides across both sensors. For mapping hydroxyl-bearing minerals in the Landsat8 dataset, PC3 emerges as the best candidate because its strong negative loading on band 6 combined with a positive loading on band 7 effectively emphasizes the characteristic absorption features of hydroxyl minerals (Appendix 4-11). Similarly, for the Landsat9 hydroxyl dataset, PC3 stands out by delivering a distinct contrast primarily a marked negative loading on band 6 with a positive loading on band 7 that best delineates hydroxyl-bearing mineral alterations (Appendix 4-11). For Sentinel-2, our Fe-oxide DPCA results showed that PC4 loadings 0.826472 (band2) and -0.525404 band 4) provided the best spectral contrast (Appendix 4-11). While for hydroxyl-bearing minerals, PC3 (–0.6249 band 11, and 0.4174 band 12) effectively delineated the diagnostic signature through a negative response in band 11 coupled with a positive contribution in band 12 (Appendix 4-11). In contrast, the 2024 Sentinel-2 data indicated that PC4 was optimal for both targets (Appendix 4-11). They yielded loadings (–0.4239, 0.3014, –0.6546, and 0.5486 for the hydroxyl set) that maximized the spectral separability by enhancing subtle differences critical to each mineral class. DPCA approach indicated consistently capitalized on the inherent spectral contrasts within the chosen band combinations, thereby robustly mapping both Fe-oxide and hydroxyl-bearing alteration zones across different acquisition years.

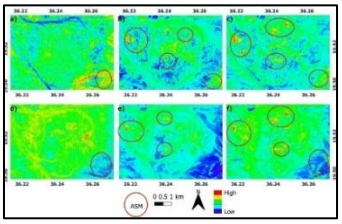


Figure 4-35. DPCA result a), b) and c) PC3 of Landsat7, 8 and 9 respectively for iron oxides. d) PC3, e) PC3 and f) PC4 of Landsat7, 8 and 9 respectively for hydroxyl bearing minerals

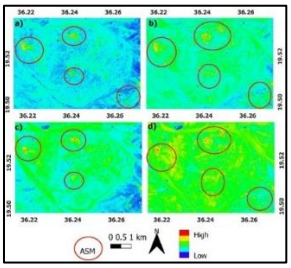


Figure 4-36. DPCA result of Sentinel-2 dataset a) PC4, b) PC3 for iron oxides in 2015 and 2024, c) PC3, d) PC4 for hydroxyl bearing minerals

4.3.3 Constrained Energy Minimization (CEM)

The result of CEM highlights the presence of hydrothermal alteration that offers valuable insights for future exploration and mining activities in the study area #1. This was consistent with prior findings of BR and density silicing, the identified alteration minerals predominantly include OH-bearing minerals such as chlorite and illite, and iron-bearing minerals like Jarosite and goethite. Their spatial distribution (Figure 4-37) clearly illustrates the extensive hydrothermal alteration affecting the study area. A spatial overlay analysis of the seven alteration minerals revealed that the area is particularly enriched in goethite, hematite, and kaolinite. These results also highlight the presence of hydrothermal

alteration along and around the extension of the ASM, which aligns with expectations and offers valuable insights for future exploration and mining activities. An important observation is the detection of alteration minerals along the streams. This is visually evident in the distinct coloration of altered superficial deposits, contrasting with unaltered deposits in regions unaffected by mining activities. This not only underscores the reliability of the CEM technique but also validates the findings through field observations, demonstrating its effectiveness in hydrothermal alteration mapping.

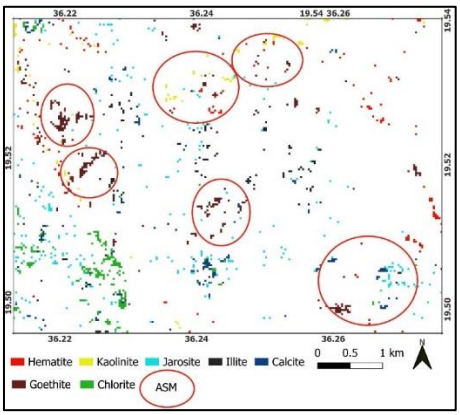


Figure 4-37. Mineral alteration mapped by CEM and their association with ASM site

4.3.4 Discussion the results of mapping hydrothermal alteration zones in ASM area using Landsat 7, 8,9, Planet, Sentinel-2 and PRISMA

The results highlighted unique spatiotemporal association between hydrothermal alterations, sensor spectral resolution and ASM development. Landsat 7 (2003) identified minor alteration minerals in the

southern part and OH-bearing in the middle, indicating alteration zones prior to ASM development. Landsat 8, Sentinel-2, and PlanetScope had detected extended Fe-bearing areas, these detected OH-bearing minerals showed association with ASM expansion into meta-volcanic rocks. PlanetScope's shortages of SWIR bands limited it to Fe-oxide identifying, highlighting the important significance of SWIR in hydroxyl mapping (van der Meer et al., 2012). PRISMA data offered unprecedented mineralogical sensitivity, revealing Fe-bearing phases and OH-bearing assemblages contained within ASM-disturbed zones. Such recently disclosed alteration zones, which were undiagnosed between 2003 and 2016, coincide with ASM's move from streams to hard-rock mining in meta-volcanics, which involving peeling eroded material to disclose underlying hydrothermal deposits. PRISMA's increased specificity in recognising mineral assemblages support hyperspectral data's superiority over multispectral sensors in complex environments (Kruse et al., 2003). Sentinel-2 and Landsat's intermediate spectral resolution improved upon Landsat 7 but lacked PRISMA's granularity, while PlanetScope's spectral limitations reaffirmed SWIR's necessity for OH-bearing mineral mapping (Abubakar et al., 2019; Chen *et al.*, 2025).

The study area's hydrothermal alteration zones predate modern ASM activities, as evidenced by regional geological models of the Arabian-Nubian Shield (ANS) (Johnson *et al.*, 2011; Abu-Alam et al., 2019; Hamimi, 2021) and localized pre-mining signatures in Landsat 7 (2003). However, the spatial correlation between ASM expansion and newly detected alteration zones highlights a critical paradox: while ASM disrupts surface geological features, it also strips deeply weathered overburden developed over 700 million years of tectonic activity, thereby exposing concealed alteration hotspots. The ANS's prolonged weathering history, coupled with intense shearing and deformation, has obscured near-surface alteration clues through supergene processes and regolith development (Butt and Zeegers, 2015; Johnson et al., 2011b).

ASM operations, particularly in meta-volcanics, mechanically remove this weathered profile, unearthing hypogene alteration assemblages that were previously undetectable by remote sensing. This duality positions ASM as both a disruptor and an unintentional exploration agent. Similar phenomena are documented in orogenic gold terrains globally, where anthropogenic disturbances expose mineralization pathways masked by deep weathering(Fisher et al., 2023; Garcia Arredondo et al., 2019). In the Red Sea Hills, where gold is hosted in hydrothermally altered shear zones (Abdelrahman et al., 2024; Abu-Fatima et al., 2021), ASM's destruction of surface features is counterbalanced by its role in revealing subsurface alteration systems a trade-off demanding innovative policies to balance environmental preservations with responsible exploration. BR, CEM and DPCA indicated hydrothermally altered rocks, consistent with orogenic gold deposit model whereby gold occurs in shear zones concealed within metavolcanic rocks and linked with late-stage alteration (Abu-Fatima et al., 2021; Groves et al., 1998).

4.3.5 PRISMA for gossan mapping in study site 2 using BR, density slicing and random forest

4.3.5.1 BR

BR results of b35/b56 (Figure 4-38/a), b108/b14 (Figure 4-38/b), b108/b42 (Figure 4-38/c) and (b35+ b108)/b14) (Figure 4-38/d) identified and mapped gossan-linked iron-bearing and clay minerals. The results disclosed gossan areas with brighter pixels corresponding to higher iron content and clay-rich zones.

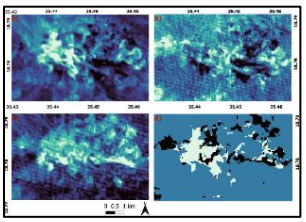


Figure 4-38. BR results. Brighter pixels indicate (gossan) Fe-rich and clay areas, blue pixels represent the host rocks, and dark areas are mining works. a) b35/b 56, b) b108/b14, c) b108/b42 and d) (b35+b108)/b14.

4.3.5.2 Density slicing

The density slicing was capable to detect the pixels corresponding to zones with the highest gossan content (Table 4-5). Accordingly, we mapped potential gossan zones along the study area (Figures 4-39). It improved the precision of gossan mapping, highlighting the benefit of integrating BR with density slicing for targeting gossan.

Table 4-5. Number of pixels representing gossan obtained by density slicing using BR results.

Density sliced BR	Number of pixels detected as anomaly per threshold		
	M +SD	M +1.5*SD	M + 2*SD
b108/b17	1283	804	492
b34/b46	1329	635	295
b108/b42	1568	819	370
(b35+b108)/b14	912	912	912

SD = Standard Deviation, Min= minimum pixel value, Max= maximum pixel value, b= band

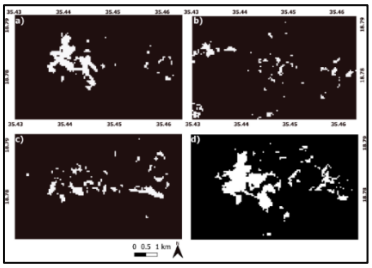


Figure 4-39. Density slicing results (white) pixels indicate Fe-rich and clay results of M +2*SD of BR and dark pixels the background. (a) b35/b56, (b) b108/b14, (c) b108/b42, and (d) (b35+ b108)/b14).

4.3.5.3 RF for hydrothermal alteration mapping and accuracy assessment

We used RF to evaluate gossan anomalies detected by BR and density slicing and in order to explore potential new gossan localities within the study area. The result showed effective gossans mapping, confirming the findings from BR and density slicing. The finindings revealed new gossan occurrences (a few pixels) in the eastern and northern parts of the study area (Figures 4-40). Furthermore, RF indictaed acceptable classification accuracy levels, with an overall accuracy of 81%, and accurcy metrices flugcuated from F-score range between 0.82-0.83, user's accuracy 0.76-0.87, and producer's accuracy 0.90 -0.77 (Table 4-6). The results of RF, BR and density slicing were integrated using intersection analysis to generate the final gossan map, which is presented in (Figures 4-41). The spectral characteristics extracted from PRISMA data, where the gossan pixel's spectral signature matched the known reflectance profile of jarosite, further validating the remote sensing interpretation (Appendix 4-12).

Table 4-6. Accuracy measures and confusion matrices for PRISMA data binary classification (UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy)

Binary RF		Reference					
	Predicted	Non-gossan	gossan	Total	UA	PA	F1-Score
	Non-gossan	20	6	26	0.7692	0.9091	0.8333
	Gossan	3	21	24	0.8750	0.7778	0.8235
	Total	23	27	50			
						OA: 0.81	

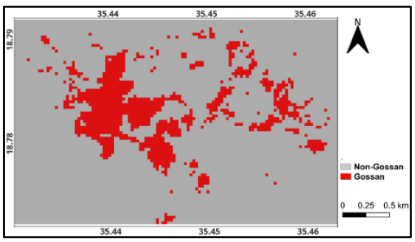


Figure 4-40: Gossan map derived from RF classification, emphasising newly identified occurrences in the northern and eastern study area.

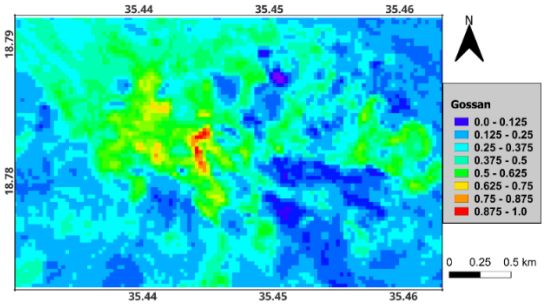


Figure 4-41. Final gossan map generated through the intersection of BR, density slicing, classification results, showing excellent agreement with field-observed gossan occurrences.

4.3.5.4 Field verification of gossan and alteration zones

Field validation revealed diverse surface expressions of gossans and associated alteration features in the Ariab region. Observed outcrops included massive primary sulphides and transitional zones where oxidation and leaching of sulphides produced characteristic gossan textures and colors These transitions were marked by changes from dark metallic phases to reddish-brown and yellowish hues, indicative of iron oxide formation and supergene alteration. Foliated gossans were often spatially associated with bleached acidic volcanic rocks and powdery silica-barite residues (SBR), reflecting intense hydrothermal activity (Figure 4-42) Brecciated siliceous gossans, often cemented by a matrix of hematite, limonite, and silica, showed strong leaching features, while other exposures revealed limonitic zones with distinct dark yellowish-orange and reddish-brown coloration (Figure 4-43). Jarosite mineralisation was also observed, typically within leached rhyolitic rocks, confirming acid-sulphate alteration under oxidising conditions. Side view of VMS and gossan. These field observations are consistent with the spectral characteristics extracted from PRISMA data, where the gossan pixel's spectral signature matched the known reflectance profile of jarosite, further validating the remote sensing interpretation.

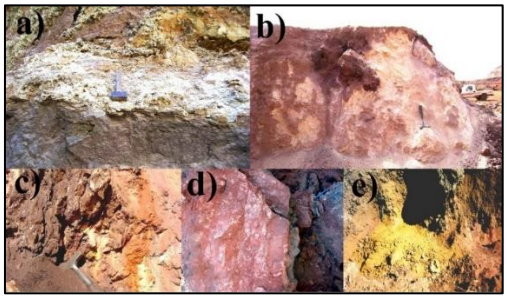


Figure 4-42. Field examples of gossans and related alteration. (a) Massive primary sulphide. (b and c) Transition from primary sulphide to gossan, showing evidence of leaching. (d) Foliated gossan associated with bleached acid volcanic rock and powdery/sugary SBR. (e) Cross-sectional view of primary sulphide. (f) Brecciated siliceous gossan with a matrix of silica, hematite, and limonite, exhibiting signs of leaching. (g) Iron oxides (limonite) displaying characteristic dark yellowish-orange and moderate reddish-brown colors. (h) Jarosite $KFe_3(SO_4)_2(OH)_6$, light green) within leached rhyolite and quartz rhyolite. (i) Altered meta-rhyolite showing iron oxide staining, kaolinisation, illitisation, and possibly silicification.

Figure 4-43. Side view VMS and gossan in the study area

4.3.5.4 PRISMA potentiality for mapping alteration zones linked to gossan in study site #2 using BR, density slicing and RF

PRISMA hyperspectral data used previously (Shebl *et al.*, 2023; Elrasheed and Szabó, 2024) to study lithological mapping using ML techniques to classify broader rock unit categories. Regarding Sudan, the application of PRISMA is still in its nascent stages but shows significant promise. Recent studies have begun to demonstrate its utility, such as for lithological mapping in the Red Sea Hills (Daoud *et al.*, 2025). While such studies provide valuable geological insights, they do not directly address mineral exploration related to gossan and VMS deposits. In contrast, our study specifically targets the mapping of gossan, which represents the surficial footprint of several potential mineralisation, primarily VMS deposits and gold. We applied color composite (CC), BR, density slicing, and RF to identify gossan using relatively coarse spatial resolution PRISMA data. CC results; bands b33, b45, and b56, and bands b108, b35, and b14 in RGB successfully highlighted the gossan and other variations across the study area. However, mapping gossan by visual inspection based solely on these CC was challenging. The BR highlighted iron-rich areas (brighter pixels), followed by density slicing, which further mapped only the highest iron-rich pixels (gossan in our study area). The application of the RF algorithm to PRISMA for gossan detection in the Red Sea region proved promising results, with an acceptable overall accuracy of 81%, F-score > 82%, UA> 76% and PA > 77%. The model successfully identified new gossan occurrences, particularly in the eastern and north-to-northwestern parts of the study area. However, some degree of misclassification was anticipated due to the complexity and inhomogeneity of lithology, the presence of alteration zones, and the geological similarities within the region. Additionally, the limited size of the study area restricted the availability of training and testing data, which may have increased the likelihood of misclassification.

Despite these challenges, the study demonstrates the potential of ML approaches for the automated detection of gossan and VMS deposits, especially in large, remote, and environmentally harsh areas. Accordingly, we recommend PRISMA data as essential for gossan detection and Au-bearing VMS in arid and semiarid regions because the scene coverage area of PRISMA is large (semi-major axis 6992.935 km),

allowing large areas to be explored and PRISMA data have spectral resolution (hyperspectral) better than the frequently used Sentinel-2, ASTER or Landsat. Allowing detailed mineral alteration mapping. Additionally, PRISMA data are freely available; therefore, they are affordable by developing countries in arid regions. However, we believe that using hyperspectral remote sensing data with spatial resolution better than PRISMA data could be useful for identifying gossans in the AN S and other arid regions. While RF provided reliable classification performance in our study, it is important to situate these results within the broader landscape of ML algorithms commonly used for hyperspectral mineral mapping. SVM have been widely adopted for lithological classification due to their ability to separate classes in high-dimensional spaces, often outperforming RF when training data are limited (Pal, 2005; Waske et al., 2007). Artificial Neural Networks (ANN), by contrast, excel at capturing complex non-linear relationships between spectral features, though they typically require larger datasets and longer training times (Lee and Oh, 2011; Saljoughi and Hezarkhani, 2018; Zareef et al., 2020). More recent approaches, such as Convolutional Neural Networks (CNN), have demonstrated superior performance in hyperspectral classification, but their computational cost and data requirements remain substantial(Zhang and Goh, 2016). Compared to these alternatives, RF offers a strong balance between accuracy, interpretability, and robustness to noise, making it particularly well-suited to the geological complexity and limited ground-truth data conditions of the ANS.

A common challenge in gossan detection using remote sensing is the presence of lateritic Fe-rich and clay-rich palaeosoils, which can be superficially similar to gossans and complicate mineral exploration efforts. These Early Tertiary laterites, formed under humid conditions, are often associated with exhumed paleosurfaces and ferricrete horizons, particularly in regions such as the southern ANS. However, in our study area, this issue is not relevant. There is no evidence of exhumed palaeosurfaces or lateritic soils contributing to the Fe and clay alteration

zones. Furthermore, the small size of the study area and the extensive basement exposure minimise the possibility of lateritic interference. These factors reinforce the validity of our remote sensing approach and its focus on true gossan detection. This conclusion is further supported by the dominance of jarosite in the gossans detected in our study area, as confirmed by the spectral signatures extracted from the pixels with the highest gossan probability (1.0). The presence of jarosite is a strong indicator of the past or ongoing oxidation of sulphides, which is characteristic of true gossans. This mineralogical evidence aligns with the geological context of the area and strengthens the reliability of our remote sensing-based detection approach, further confirming that lateritic interference is not a complicating factor in our study.

Detailed spectral analysis was performed again using spectral curves of pixels identified as gossan, along with the spectral characteristics of typical gossan minerals (e.g., jarosite). The latter clearly illustrates an overall match between the gossan pixels extracted from PRISMA data (red) and the reference jarosite curve derived from the USGS spectral library. A detailed investigation of both curves reveals a general coincidence in the main absorption features and peaks, as highlighted by the green markers in the SWIR region. However, the curves are not identical due to differences in measurement conditions. For instance, the jarosite reference curve represents the reflectance of a pure jarosite sample, which is nearly impossible to achieve from remote sensing data at this spatial resolution. In particular, within the study area, it is unlikely to find a pure pixel (30 m) composed solely of jarosite, as it may contain clay minerals, weathering products, remnants of the host rock, and even spectral influences from mining activities. Despite these factors, the overall alignment of the main peaks and troughs remains reasonable. This kind of spectral analysis is considered reliable, as documented by several studies (Daoud *et al.*, 2025). Here, it serves to validate the findings obtained through remote sensing image analysis, density slicing, and RF classification.

CHAPTER 5. CONCLUSIONS

Several conclusions were drawn in this study using over six remote sensing data, field observations, and petrographic studies. We used various ML models to deliver a modified lithological map of the study area and ASM mapping and monitoring. By revealing the value of employing Landsat and Sentinel-2 data as the most suitable satellite data sources for lithological and ASM mapping, our research paves the way for using ML and remote sensing data for ASM and lithological mapping research in the future in remote, inaccessible, and challenging environments with complex geology and intensified ASM activities. Furthermore, our research explored data-driven strategies to address reference data limitations in ASM and lithological mapping fields. Segmentation-based data augmentation and PL were helpful in improving model accuracy and spatial consistency with limited training samples. The RF algorithm remained accurate, whereas MARS benefited considerably from PL, resulting in improved class-level accuracies. These outcomes reveal that adaptive data augmentation can enhance ASM and lithological mapping, even with limited ground-truth data. The results also highlight the value of combining algorithmic advances with reliable visual inspection and validation to esure geologically precise findings.

Moreover, our study used Planet very high spatial resolution and PRISMA hyperspectral data to validate their potential for ASM detection when combined with ML. We conducted several image enhancements, such as CC, BR, CEM, PCA, and DPCA, for the different satellite datasets before using them as input data for model building. These image enhancements facilitate the process of reference data collection to train and test our models. Several rock types, including ophiolite, marble, metavolcanic, granitoids, and superficial deposits, as well as ASM areas, were identified. The occurrence of gossan was detected in study site 2 by the inspection of spectral bands of PRISMA that indicated positive anomalies of iron and OH-bearing minerals as a surface indication of deeply rooted VMS.

Regarding ASM mapping and monitoring, we introduced Planet and PRISMA data in ASM studies, in addition to Landsat and Sentinel-2, for mapping ASM development over two study areas using binary and multiclass ML classifiers. All results obtained from the different data

sources showed significant increases during the study period. This ASM expanse has significantly impacted geological features, including the obliteration of orebody continuity, land degradation, and drainage system disruption, which potentially pose an environmental hazards, unsustainable resource exploitation, and potential River Nile water pollution. It also showed that the binary method outperformed (1-10%) the multiclass classification approach; in class-level metrics, multiclass indicated the same levels. Although the multiclass approach was less accurate, it offered valuable geological information on rock units linked to gold mining and was used more frequently over time.

The study also found that the area witnessed a substantial hydrothermal alteration process which resulted in the area being reached with hydrothermal alteration minerals, mainly Fe-bearing minerals, including different iron oxide minerals, and OH-bearing minerals, and clay minerals such as kaolinite. This finding is supported by field observations and petrographic studies, which have proven the presence of a hydrothermal alteration assembly, including carbonatisaion (calcite), ferrugination (hematite), silicification (quartz), sericitisation (sericite), and chloritisation (chlorite and epidotes). Furthermore, ASM sites also represent strong prospecting criteria as the miners are directly targeting and producing gold in the area, and the activities expose the hidden alteration zones, which are due to deep weathering processes and were observed during the fieldwork.

It is important to note that our findings are primarily relevant to environments similar to the study area, where arid to semi-arid conditions are dominant and must therefore be perceived consequently. Remarkably, the outcomes achieved by combining of various methods showed an outstanding level of reliability at various stages. For instance, detected hydrothermal alteration zones are closely associated with areas where ASM activities are developing and gold is being produced. Likewise, the miners in the ASM sites targeted metavolcanic rocks, which are known as the source of the orogenic gold deposit in the region, and the obliteration of geological features therefore mostly affected the metavolcanic rocks, as observed during fieldwork and supported by petrographic investigations.

Summary

This dissertation investigated the integration of multi-sensor remote sensing and machine learning (ML) techniques for lithological mapping, hydrothermal alteration detection, and artisanal and small-scale mining (ASM) monitoring in the Red Sea Hills (RSH) region of northeast Sudan. This study was motivated by the need to develop efficient, low-cost, and replicable geospatial approaches for geological and environmental assessments in remote, data-scarce, and geologically complex terrains. The Red Sea Hills, part of the Arabian–Nubian Shield (ANS), host diverse lithological assemblages, significant mineralization zones, and rapidly expanding ASM activities, yet remain underexplored using advanced earth observation methods. This dissertation thus aimed to enhance understanding of the region's geological evolution and anthropogenic impacts through the application of machine learning algorithms to multi-sensor optical and hyperspectral datasets.

The literature review highlighted the evolution of remote sensing techniques for lithological and mineral exploration, emphasizing the role of multispectral and hyperspectral data in discriminating lithological units, detecting alteration zones, and monitoring ASM dynamics in the study area. Previous studies on the ANS and RSH primarily focused on regional tectonics and conventional geological mapping using medium-resolution satellite data. However, the application of high-resolution sensors and advanced ML algorithms for integrated lithological and ASM mapping remains limited. This review synthesizes findings from prior work on supervised and ensemble classifiers, particularly Random Forest (RF), Support Vector Machine (SVM), and Multivariate Adaptive Regression Splines (MARS), and discusses challenges related to spectral similarity, limited reference data, and validation accuracy. Emerging methods, such as pseudo-labelling (PL) and data augmentation have been identified as promising strategies for addressing data scarcity. The review also contextualizes ASM within the broader socioeconomic and environmental landscape, emphasizing its dual role as both an economic lifeline and a source of significant ecological degradation and geological disturbance.

This study utilized an array of remote sensing datasets including Landsat-8, Landsat-9, Sentinel-2, PlanetScope, PRISMA, and EnMAP imagery, complemented by field observations, petrographic analysis, and

existing geological maps. The study area encompassed representative sites within the RSHs, characterized by diverse lithological units, including ophiolites, metavolcanics, granitoids, marble, altered rocks, superficial deposits, and active ASM operations. Pre-processing involved radiometric and atmospheric correction, followed by image enhancement techniques such as false colour composites (FCC), Band Ratios (BR), and Principal Component Analysis (PCA) to optimize spectral separability.

Supervised machine learning classifiers (RF, SVM, Naïve Bayes, and MARS) were applied to generate lithological and ASM maps, with model optimization achieved through hyperparameter tuning, oversampling, and cross-validation. Pseudo-labelling was employed to augment training data by incorporating high-probability samples, and a region-growing segmentation algorithm was developed to further enhance training datasets. Accuracy assessment was conducted using both point-based and area-based approaches, computing metrics such as overall accuracy (OA), user' and producer' accuracy, F1-score, and Kappa coefficient (Foody, 2002). This study also introduced a spatially explicit evaluation framework to assess the extent and spatial consistency of ASM expansion.

The integration of multispectral (Landsat-8/9, and Sentinel-2) and hyperspectral (PRISMA) datasets demonstrated that both RF and SVM classifiers are highly effective in lithological discrimination, achieving OAs between 0.90 and 0.96. Sentinel-2 outperformed Landsat data in term of spectral separability, particularly in distinguishing metavolcanic, marble, and altered rock units. PRISMA hyperspectral imagery produced the most accurate classification (OA = 0.96; $\kappa = 0.95$) when combined with RF, followed by Naïve Bayes (OA = 0.92; $\kappa = 0.90$). The results are consistent with field observations and petrographic analyses. Misclassifications primarily occurred between granitoids, altered rocks, and superficial deposits owing to their mineralogical similarity and weathering-induced spectral overlap. The study confirmed that ensemble algorithms, such as RF, provide stable and robust outputs, even with heterogeneous training data.

The detection of hydrothermal alteration zones using PRISMA data successfully identified gossan occurrences associated with volcanogenic massive sulphide (VMS) mineralization. Spectral analysis and image processing revealed diagnostic signatures of Fe- and OH-bearing

minerals, including hematite, goethite, kaolinite, and sericite, which were validated by field and petrographic evidence. Alteration assemblages, such as ferrugination, silicification, sericitization, chloritization, and carbonatization, were spatially correlated with structural lineaments and ASM sites, demonstrating that artisanal mining often exposes and exploits hydrothermally altered rocks. These findings highlight the potential of hyperspectral remote sensing for mapping mineralization indicators and guiding exploration in arid environments in the future.

PlanetScope, Sentinel-2, and Landsat datasets revealed substantial temporal expansion of ASM activities between 2003 and 2024. Binary and multiclass RF classifications consistently indicated that the ASM areas increased by 150–300% across the monitored sites. For example, at Site 3, PlanetScope data showed ASM growth from 50 ha in 2016 to 125 ha in 2024, while Sentinel-2 data recorded an increase from 46.4 ha in 2017 to 85.2 ha in 2024. Change detection analysis confirmed that the expansion primarily occurred over the meta-andesite and, to a lesser extent, the meta-basalt units. These lithologies, rich in iron and magnesium, are closely associated with gold mineralization, which explains the miners' spatial targeting patterns. The findings also demonstrated that ASM expansion led to significant geomorphological and environmental impacts, including drainage alteration, surface degradation, and destruction of structural features such as faults and orebody continuity.

Experiments on pseudo-labelling demonstrated that this technique can improve model stability and reliability when ground-truth data are limited. However, improvements in accuracy were not consistent across classifiers or lithologies. The RF model benefited moderately from PL, whereas MARS displayed over-optimistic probability outputs that did not correspond to the true accuracy levels. The region-growing segmentation method, developed to augment the training data, improved the classification accuracy for ASM mapping, particularly when small but representative additional samples were used. These results highlight the importance of expert-guided validation and adaptive data augmentation in geological remote sensing.

This dissertation contributes to the fields of remote sensing and geological mapping in several ways.

-Methodological Advancement: It integrates multi-sensor optical and hyperspectral data with machine learning and data augmentation to

overcome challenges of limited reference data in arid, inaccessible regions.
-Novel Application: This study represents one of the first comprehensive applications of PRISMA hyperspectral and Planet data for ASM and lithological mapping in the RSH.
-Spatial Insight: It provides the first detailed spatiotemporal assessment of ASM expansion in northeast Sudan, linking mining activity to specific lithological and alteration zones.
-Validation Framework: It combined point- and area-based accuracy assessment approach was introduced, improving the spatial interpretation of the classification performance.
-Geological Understanding: It enhances the understanding of the relationship between hydrothermal alteration, ASM activity, and geology within the RSHs, supporting more sustainable mineral exploration and land-use planning.

The findings confirm that the integration of multi-sensor remote sensing and machine learning offers a powerful and cost-effective framework for lithological mapping and ASM monitoring in complex and data-scarce environments. Hyperspectral data, particularly PRISMA data, provide superior performance in detecting subtle mineralogical variations, whereas ensemble learning algorithms such as RF offer robust and interpretable results. The study also underscores the potential of adaptive data augmentation (pseudo-labelling and region-growing) to improve classification accuracy under reference data limitations.

Future research should expand the application of advanced hyperspectral missions (e.g., EnMAP and CHIME) for mineral exploration and investigate the use of deep learning architectures (CNNs, transformers) to improve spatial feature extraction. The integratation of InSAR and LiDAR data can further enhance the detection of geomorphological and structural deformations associated with mining. Finally, the regional-scale implementation of the developed methodology could support the sustainable management of ASM activities and guide mineral resource policy in Sudan and other parts of the Arabian–Nubian Shield.

Acknowledgements

First and foremost, I would like to express my profound gratitude to Allah, the Almighty, for granting me the strength, patience, and perseverance to complete this PhD journey. Without His endless mercy and guidance, none of this would have been possible.

To my beloved mother, your unconditional love, prayers, and unwavering faith in me have been my greatest source of strength and inspiration. May Allah bless you abundantly, protect you always, and reward you for every sacrifice you have made on my behalf.

To my dear family, I am deeply thankful for your constant encouragement, understanding, and support throughout this long and challenging path. Your belief in me has sustained me during the most difficult moments of this academic pursuit.

My deepest appreciation goes to my supervisor, Dr. Szabó Szilárd, whose exceptional guidance, patience, and insightful advice have shaped not only this dissertation but also my academic and personal growth. You have been more than a supervisor a true mentor, honest advisor, and trusted friend, always generous with your time and support in overcoming every challenge I faced.

I am sincerely grateful to David, Zyenap, and Yusif Yassin for their enduring support, encouragement, and friendship throughout this journey. Your words of motivation and acts of kindness have meant more to me than I can express. A special word of thanks goes to Vera for her continuous assistance and kindness, especially during the demanding days leading up to my pre-defense. Your help and positive spirit were invaluable, and I deeply appreciate your generosity and understanding.

I would also like to extend my sincere thanks to all my professors, colleagues, and friends who contributed to this work, whether through academic discussions, moral support, or assistance during fieldwork. Your cooperation and companionship have made this experience truly rewarding.

Lastly, to all those who, in one way or another, supported, inspired, or encouraged me throughout this academic and personal journey. I express my heartfelt gratitude. This accomplishment is as much yours as it is mine.

The Author was supported by the Stipendium Hungaricum Scholarship of Hungary, and the NKFI K 138079 , and the RRF-2.3.1-21-2022-00008 by the 'Széchenyi Terv Plusz program' projects.

References

Abbaszadeh, M., Soltani-Mohammadi, S., Ahmed, A.N., 2022. Optimization of support vector machine parameters in modeling of Iju deposit mineralization and alteration zones using particle swarm optimization algorithm and grid search method. Comput. Geosci. 165, 105140. https://doi.org/10.1016/j.cageo.2022.105140

Abdelkareem, M., Hamimi, Z., El-Bialy, M.Z., Khamis, H., Abdel Wahed, S.A., 2021. Integration of remote-sensing data for mapping lithological and structural features in the Esh El-Mallaha area, west Gulf of Suez, Egypt. Arab. J. Geosci. 14, 497. https://doi.org/10.1007/s12517-021-06791-3

Abdelrahman, S., Ibrahim, M.A.E., Li, H., Abdel Rahman, E.M., Faisal, M., 2024. Geochemical characteristics of Neoproterozoic metavolcanic rocks of Ariab Auriferous Volcanogenic Massive Sulfide deposit, Red Sea hills, North-East Sudan. J. Afr. Earth Sci. 216, 105305. https://doi.org/10.1016/j.jafrearsci.2024.105305

ABDELSALAM, M.G., 2000. Mapping gossans in arid regions with Landsat TM and SIR-C images: the Beddaho Alteration Zone in northern Eritrea.

Abdelsalam, M.G., Stern, R.J., 1993. Tectonic evolution of the Nakasib suture, Red Sea Hills, Sudan: evidence

for a late Precambrian Wilson Cycle. J. Geol. Soc. 150, 393–404. https://doi.org/10.1144/gsjgs.150.2.0393

Abdelsalam, M.G., Stern, R.J., Berhane, W.G., 2000. Mapping gossans in arid regions with Landsat TM and SIR-C images: the Beddaho Alteration Zone in northern Eritrea. Journal of African Earth Sciences 30(4), 903-916. https://doi.org/. https://doi.org/10.1016/S0899-5362(00)00059-2

Abdoulatif, A.S., Mahamadou, S., Mohamed, M., 2025. Mapping of Hydrothermal Alteration Minerals Related to Tongue Tonguey (Sirba, Niger) Gold Mineralization Using Landsat 8 OLI Data and Remote Sensing. Asian J. Phys. Chem. Sci. 13, 1–14. https://doi.org/10.9734/ajopacs/2025/v13i1237

Abrams, M., Yamaguchi, Y., 2019. Twenty Years of ASTER Contributions to Lithologic Mapping and Mineral Exploration. Remote Sens. 11, 1394. https://doi.org/10.3390/rs11111394

Abrams, M.J., Brown, D., Lepley, L., Sadowski, R., 1983. Remote sensing for porphyry copper deposits in southern Arizona. Econ. Geol. 78, 591–604. https://doi.org/10.2113/GSECONGEO.78.4.591

Abu-Alam, T., El Monsef, M.A., Grosch, E., 2019. Shear-zone hosted gold mineralization of the arabian-nubian shield: Devolatilization processes across the greenschist-amphibolite-facies transition. Geol. Soc. Spec. Publ. 478, 287–313. https://doi.org/10.1144/SP478.13

Abubakar, A.J., Hashim, M., Pour, A.B., 2019. Identification of hydrothermal alteration minerals associated with geothermal system using ASTER and Hyperion satellite data: a case study from Yankari Park, NE Nigeria. Geocarto Int. 34, 597–625. https://doi.org/10.1080/10106049.2017.1421716

Abu-Fatima, M., Marignac, C., Cathelineau, M., Boiron, M.-C., 2021. Metallogeny of a Pan-African oceanic arc: VHMS and gold deposits in the Ariab-Arbaat belt, Haya terrane, Red Sea Hills (Sudan). Gondwana Res. 98, 76–106. https://doi.org/10.1016/j.gr.2021.06.001

Achuta Rao, S.V., Kondaiah, K., Rajesh Chandra, .G., Kiran Kumar, K., 2017. A Survey on Machine Learning: Concept, Algorithms and Applications.

Adam, M.S.M., Babikir, K.A., Kim, Y., Park, M.-E., F.H. Orellana, J., Kim, Y.-S., 2024. Geology and structural setting of the undiscovered Amgififa volcanogenic massive sulfide (VMS) deposit in the southwestern part of the Haya Terrane, Arabian-Nubian Shield, Sudan. Ore Geol. Rev. 167, 105969. https://doi.org/10.1016/j.oregeorev.2024.105969

Ahmed, A., Purwanto, P., Sunoko, H., 2019. Consequences of Mercury Used by Artisanal and Small-Scale Gold Mining Processes – A Case of River Nile State Sudan. J. Ecol. Eng. 20, 106–115. https://doi.org/10.12911/22998993/96275

Ahmed El Tohami, A.B.E.S., 2018. Smart Artisanal Gold Mining from a Sudanese Perspective. Biomed. J. Sci.

Tech. Res. 8. https://doi.org/10.26717/BJSTR.2018.08.001704

Alessi, M.A., Chirico, P.G., Millones, M., 2023. Artisanal Mining River Dredge Detection Using SAR: A Method Comparison. Remote Sens. 15, 5701. https://doi.org/10.3390/rs15245701

Ali, S., Li, H., Ali, A., Hassan, J.I., 2024. Lithological Discrimination of Khyber Range Using Remote Sensing and Machine Learning Algorithms. Appl. Sci. 14, 5064. https://doi.org/10.3390/app14125064

Almond, D.C., Ahmed, F., 1987. Ductile shear zones in the northern Red Sea Hills, Sudan and their implication for crustal collision. Geol. J. 22, 175–184. https://doi.org/10.1002/gj.3350220614

Al-Mufti, O.N., 2022. Discussion of Li *et al* . (2021) – On the origin and significance of composite particles in mudstones: Examples from the Cenomanian Dunvegan Formation, *Sedimentology* , 68, 737–754. Sedimentology 69, 2676–2681. https://doi.org/10.1111/sed.13001

Alvioli, M., Guzzetti, F., Marchesini, I., 2020. Parameter-free delineation of slope units and terrain subdivision of Italy. Geomorphology 358, 107124. https://doi.org/10.1016/j.geomorph.2020.107124

Alvioli, M., Santangelo, M., Fiorucci, F., Cardinali, M., Marchesini, I., Reichenbach, P., Rossi, M., Guzzetti, F., Peruccacci, S., 2021. Rockfall susceptibility and network-ranked susceptibility along the Italian railway. Eng. Geol. 293, 106301. https://doi.org/10.1016/j.enggeo.2021.106301

Amer, R., Kusky, T., Ghulam, A., 2010. Lithological mapping in the Central Eastern Desert of Egypt using ASTER data. J. Afr. Earth Sci. 56, 75–82. https://doi.org/10.1016/j.jafrearsci.2009.06.004

Amri, I., Saringatin, S., Ruslanjari, D., 2023. Detection and mapping abandoned areas of artisanal and small-scale gold mining (ASGM) using multi-sensor data on Google Earth Engine: A case study of Kuantan Singingi, Riau. E3S Web Conf. 468, 03003. https://doi.org/10.1051/e3sconf/202346803003

Amusuk, D.J., Hashim, M., Beiranvand Pour, A., Habashi, J., 2024. Regional Mapping of Basement Lithologies Using Geospatial Data in Semi-Arid Regions: Techniques, Advancements and Applications. Adv. Geol. Geotech. Eng. Res. 6, 12–40. https://doi.org/10.30564/agger.v6i2.6130

Ananthkumar, A., 2023. Using Machine Learning to Predict Lithostratigraphic Facies. J. Stud. Res. 12. https://doi.org/10.47611/jsrhs.v12i4.5150

Anifadi, A., Sykioti, O., National, E.V., Geology, F., 2019. Detection of chromite minerals using Spectral Linear Unmixing on Sentinel-2 imagery . Case study : Ingessana Hills , Blue Nile Province , Sudan . 2002–2003.

Arif Ali, Z., H. Abduljabbar, Z., A. Tahir, H., Bibo Sallow, A., Almufti, S.M., 2023. eXtreme Gradient Boosting Algorithm with Machine Learning: a Review. Acad. J. Nawroz Univ. 12, 320–334. https://doi.org/10.25007/ajnu.v12n2a1612

Aslam, F., 2023. Advancing Intelligence: Unveiling the Power of Advanced Machine Learning Algorithms for Real-World Applications. J. Eng. Res. Rep. 25, 159–165. https://doi.org/10.9734/jerr/2023/v25i7949

Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A.C., Benbouziane, A., 2019. Machine Learning Algorithms for Automatic Lithological Mapping Using Remote Sensing Data: A Case Study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. ISPRS Int. J. Geo-Inf. 8, 248. https://doi.org/10.3390/ijgi8060248

Bahrami, H., Esmaeili, P., Homayouni, S., Pour, A.B., Chokmani, K., Bahroudi, A., 2024. Machine Learning-Based Lithological Mapping from ASTER Remote-Sensing Imagery. Minerals 14, 202. https://doi.org/10.3390/min14020202

Bai, B., Tan, Y., Donchyts, G., Haag, A., Xu, B., Chen, G., Weerts, A.H., 2023. Naive Bayes classification-based surface water gap-filling from partially contaminated optical remote sensing image. J. Hydrol. 616, 128791. https://doi.org/10.1016/j.jhydrol.2022.128791

Bajwa, R.S., Ahsan, N., Ahmad, S.R., 2020. A Review of Landsat False Color Composite Images for Lithological Mapping of Pre-Cambrian to Recent Rocks: A Case Study of Pail/Padhrar Area in Punjab Province, Pakistan. J. Indian Soc. Remote Sens. 48, 721–728. https://doi.org/10.1007/s12524-019-01090-7

Bakheit, A.K., Matheis, G., 2017. Gold-productive volcanogenic sulphide mineralization in the Ariab Belt, Red Sea Hills/Sudan: Evidences for Late Proterozoic seafloor hydrothermal systems, in: Thorweihe, U., Schandelmeier, H. (Eds.), Geoscientific Research in Northeast Africa. CRC Press, pp. 533–540. https://doi.org/10.1201/9780203753392-97

Barenblitt, A., Payton, A., Lagomasino, D., Fatoyinbo, L., Asare, K., Aidoo, K., Pigott, H., Som, C.K., Smeets, L., Seidu, O., Wood, D., 2021. The large footprint of small-scale artisanal gold mining in Ghana. Sci. Total Environ. 781, 146644. https://doi.org/10.1016/j.scitotenv.2021.146644

Barsi, Á., Kugler, Zs., László, I., Szabó, Gy., Abdulmutalib, H.M., 2018. ACCURACY DIMENSIONS IN REMOTE SENSING. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII–3, 61–67. https://doi.org/10.5194/isprs-archives-XLII-3-61-2018

Be'eri-Shlevin, Y., Katzir, Y., Whitehouse, M.J., Kleinhanns, I.C., 2009. Contribution of pre Pan-African crust to formation of the Arabian Nubian Shield: New secondary ionization mass spectrometry U-Pb and O studies of zircon. Geology 37, 899–902. https://doi.org/10.1130/g30206a.1

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114,

24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bhadula, S., Adnan, M.M., Kumar, R., Rana, A., Kaliyaperumal, G., Rao, B.D., Boob, N.S., 2024. Optimizing Random Forest Algorithms for LargeScale Data Analysis, in: 2024 7th International Conference on Contemporary Computing and Informatics (IC3I). Presented at the 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Greater Noida, India, pp. 1673–1678. https://doi.org/10.1109/IC3I61595.2024.10829145

Bhagwat, R.U., Uma Shankar, B., 2019. A novel multilabel classification of remote sensing images using XGBoost, in: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). Presented at the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), IEEE, Bombay, India, pp. 1–5. https://doi.org/10.1109/I2CT45611.2019.9033768

Bierlein, F.P., McKeag, S., Reynolds, N., Bargmann, C.J., Bullen, W., Murphy, F.C., Al-Athbah, H., Brauhart, C., Potma, W., Meffre, S., McKnight, S., 2016. The Jebel Ohier deposit—a newly discovered porphyry copper–gold system in the Neoproterozoic Arabian–Nubian Shield, Red Sea Hills, NE Sudan. Miner. Deposita 51, 713–724. https://doi.org/10.1007/s00126-015-0630-z

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

https://doi.org/10.1023/A:1010933404324
Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4
Bucci, F., Santangelo, M., Fongo, L., Alvioli, M., Cardinali, M., Melelli, L., Marchesini, I., 2022. A new digital lithological map of Italy at the 1:100 000 scale for geomechanical modelling. Earth Syst. Sci. Data 14, 4129–4151. https://doi.org/10.5194/essd-14-4129-2022
Bui, D.H., Mucsi, L., 2022. Predicting the future land-use change and evaluating the change in landscape pattern in Binh Duong province, Vietnam. Hung. Geogr. Bull. 71, 349–364. https://doi.org/10.15201/hungeobull.71.4.3
Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167.
Butler, R.W.H., Torvela, T., Williams, L., 2024. An introduction to geological mapping of our world and others. Geol. Soc. Lond. Spec. Publ. 541, 1–17. https://doi.org/10.1144/SP541-2023-201
Butt, C.R.M., Zeegers, H., 2015. Regolith Exploration Geochemistry in Tropical and Subtropical Terrains.

Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 43, 1351–1362. https://doi.org/10.1109/TGRS.2005.846154

Cao, C., Chicco, D., Hoffman, M.M., 2020. The MCC-F1 curve: a performance evaluation technique for binary classification. https://doi.org/10.48550/arXiv.2006.11278

Chen, Q., Cai, D., Xia, J., Zeng, M., Yang, H., Zhang, R., He, Y., Zhang, X., Chen, Y., Xu, X., Zhao, Z., 2025. Remote sensing identification of hydrothermal alteration minerals in the Duobuza porphyry copper mining area in Tibet using WorldView-3 and GF-5 data: The impact of spatial and spectral resolution. Ore Geol. Rev. 106573. https://doi.org/10.1016/j.oregeorev.2025.106573

Chen, S.-Y., Lin, C., Chuang, S.-J., Kao, Z.-Y., 2019. Weighted Background Suppression Target Detection Using Sparse Image Enhancement Technique for Newly Grown Tree Leaves. Remote Sens. 11, 1081. https://doi.org/10.3390/rs11091081

Chen, Y., Sui, Y., Shayilan, A., 2023. Constructing a high-performance self-training model based on support vector classifiers to detect gold mineralization-related geochemical anomalies for gold exploration targeting. Ore Geol. Rev. 153, 105265. https://doi.org/10.1016/j.oregeorev.2022.105265

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification

evaluation. BMC Genomics 21, 6. https://doi.org/10.1186/s12864-019-6413-7

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Routledge, New York. https://doi.org/10.4324/9780203771587

Collins, L., McCarthy, G., Mellor, A., Newell, G., Smith, L., 2020. Training data requirements for fire severity mapping using Landsat imagery and random forest. Remote Sens. Environ. 245, 111839. https://doi.org/10.1016/j.rse.2020.111839

Computer Science & Engineering &GZSCCET Bhatinda, Punjab, India, Goel, E., Abhilasha, Er., Computer Science & Engineering &GZSCCET Bhatinda, Punjab, India, 2017. Random Forest: A Review. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 7, 251–257. https://doi.org/10.23956/ijarcsse/V7I1/01113

Congalton, R.G., 2001. Accuracy assessment and validation of remotely sensed and other spatial information. Int. J. Wildland Fire 10, 321. https://doi.org/10.1071/WF01031

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Coulthard, T.J., 2001. Landscape evolution models: a software review. Hydrol. Process. 15, 165–173. https://doi.org/10.1002/hyp.426

Crosta, A.P., 1989. Enhancement of Landsat Thematic Mapper imagery for residual soil mapping in SW Minas Gerais State Brazil, a prospecting case history

in greenstone belt terrain. Presented at the Proceedings of the 7^< th> ERIM Thematic Conference on Remote Sensing for Exploration Geology, 1989.

Cruz, C., McGuinness ,Kevin, Perrin ,Philip M., O'Connell ,Jerome, Martin ,James R., and Connolly, J., 2023. Improving the mapping of coastal invasive species using UAV imagery and deep learning. Int. J. Remote Sens. 44, 5713–5735. https://doi.org/10.1080/01431161.2023.2251186

Curran, J., Hersh, T., 2012. Hotelling: Hotelling's T^2 Test and Variants. https://doi.org/10.32614/CRAN.package.Hotelling

Daoud, A.M.A., Shebl, A., Abdelkader, M.M., Mohieldain, A.A., Csámer, Á., Satti, A.M.N., Rózsa, P., 2025. Remote sensing and gravity investigations for barite detection in Neoproterozoic rocks in the Ariab area, Red Sea Hills, Sudan. Remote Sens. Appl. Soc. Environ. 37, 101416. https://doi.org/10.1016/j.rsase.2024.101416

Dayley, B., 2009. Python phrasebook: essential code and commands, 5. print. ed, Developer's library. Sams, Indianapolis, Ind.

De Graaf, I.E.M., Van Beek, R.L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E.H., Bierkens, M.F.P., 2017. A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. Adv. Water Resour. 102, 53–67. https://doi.org/10.1016/j.advwatres.2017.01.011

de Theije, M., Heemskerk, M., 2009. Moving Frontiers in the Amazon: Brazilian Small-Scale Gold Miners in Suriname. Eur. Rev. Lat. Am. Caribb. Stud. Rev. Eur. Estud. Latinoam. Caribe 5–25.

Diwyacitta, K., Gustiandi, B., Indradjad, A., 2022. Automatic True Color Composites Generation Based on NOAA JPSS Satellites Data, in: Yulihastin, E., Abadi, P., Sitompul, P., Harjupa, W. (Eds.), Proceedings of the International Conference on Radioscience, Equatorial Atmospheric Science and Environment and Humanosphere Science, 2021, Springer Proceedings in Physics. Springer Nature Singapore, Singapore, pp. 637–649. https://doi.org/10.1007/978-981-19-0308-3_51

Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian classifier under zero-one loss. Mach. Learn. 29, 103–130.

Donnini, M., Marchesini, I., Zucchini, A., 2020. A new Alpine geo-lithological map (Alpine-Geo-LiM) and global carbon cycle implications. GSA Bull. 132, 2004–2022. https://doi.org/10.1130/b35236.1

Dorner, U., 2012. Artisanal and Small-Scale Mining (ASM).

Dou, B., Zhu, Z., Merkurjev, E., Ke, L., Chen, L., Jiang, J., Zhu, Y., Liu, J., Zhang, B., Wei, G.-W., 2023. Machine Learning Methods for Small Data Challenges in Molecular Science. Chem. Rev. 123, 8736–8780. https://doi.org/10.1021/acs.chemrev.3c00189

Drury, S.A., 2001. Image Interpretation in Geology, 3rd ed. Nelson Thornes, Cheltenham.

Elahi, F., Muhammad, K., Din, S.U., Khan, M.F.A., Bashir, S., Hanif, M., 2022. Lithological Mapping of Kohat Basin in Pakistan Using Multispectral Remote Sensing Data: A Comparison of Support Vector Machine (SVM) and Artificial Neural Network (ANN). Appl. Sci. 12, 12147. https://doi.org/10.3390/app122312147

El-Nadi, A.H., 1989. Late precambrian volcanism in NE Sudan and the evolution of the Nubian shield. J. Afr. Earth Sci. 9, 467–480. https://doi.org/10.1016/0899-5362(89)90031-6

El-Omairi, M.A., El Garouani, A., 2024. Lithological Mapping using Artificial Intelligence and Remote Sensing data: A Case Study of Bab Boudir region, Morocco. BIO Web Conf. 115, 01005. https://doi.org/10.1051/bioconf/202411501005

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

https://doi.org/10.1016/j.heliyon.2023.e20168

Elrasheed, A.A., Obaid, Y.Y., Szabó, S., 2025. Spatial expansion of artisanal and small-scale gold mining nearby the Nile River, Sudan and its potential environmental impacts: Insights from Planetscope data and machine learning. Environ. Chall. 20, 101278. https://doi.org/10.1016/j.envc.2025.101278

Elrasheed, A.A., Szabó, S., 2024. Comparing the Capability of Multi- and Hyperspectral Remote Sensing Data in

Lithological Mapping Using Machine Learning Algorithms: A Case Study from Sudan. https://doi.org/10.5194/egusphere-egu24-5443

Elsheikh, A.E.M., Zeinelabdein, K.A.E., Rahman, E.M.A., Eljah, M.A., 2015. The Structural Evolution of the Hamisana Geodynamic Zone , Red Sea Region , NE Sudan 2, 52–60.

Elwaleed, A., Jeong, H., Abdelbagi, A.H., Thi Quynh, N., Nugraha, W.C., Agusa, T., Ishibashi, Y., Arizono, K., 2024. Assessment of Mercury Contamination in Water and Soil from Informal Artisanal Gold Mining: Implications for Environmental and Human Health in Darmali Area, Sudan. Sustainability 16, 3931. https://doi.org/10.3390/su16103931

Evans, J.S., Murphy, M.A., Ram, K., 2023. spatialEco: Spatial Analysis and Modelling Utilities.

Evuk, D., 2013. Geodynamic evolution of the central-eastern Bayuda Desert Basement, Sudan.

Fadlallah, M.A., Pal, I., Hoe, V.C., 2020. Determinants of perceived risk among artisanal gold miners: A case study of Berber locality, Sudan. Extr. Ind. Soc. 7, 748–757. https://doi.org/10.1016/j.exis.2020.03.006

Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10, 175–189. https://doi.org/10.1007/s10346-012-0320-1

Feng, M., Wang, X., Zhao, Z., Jiang, C., Xiong, J., Zhang, N., 2024. Enhanced Heart Attack Prediction Using eXtreme Gradient Boosting. J. Theory Pract. Eng. Sci. 4, 9–16. https://doi.org/10.53469/jtpes.2024.04(04).02

Fisher, B.A., Yoo, K., Aufdenkampe, A.K., Nater, E.A., Feinberg, J.M., Nyquist, J.E., 2023. Mineral surface area in deep weathering profiles reveals the interrelationship of iron oxidation and silicate weathering. Earth Surf. Dyn. 11, 51–69. https://doi.org/10.5194/esurf-11-51-2023

Fisher, E., 2007. Occupying the Margins: Labour Integration and Social Exclusion in Artisanal Mining in Tanzania. Dev. Change 38, 735–760. https://doi.org/10.1111/j.1467-7660.2007.00431.x

Foody, G.M., 2009a. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. Remote Sens. Environ. 113, 1658–1663. https://doi.org/10.1016/j.rse.2009.03.014

Foody, G.M., 2009b. Sample size determination for image classification accuracy assessment and comparison. Int. J. Remote Sens. 30, 5273–5291. https://doi.org/10.1080/01431160903130937

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector

machines. IEEE Trans. Geosci. Remote Sens. 42, 1335–1343. https://doi.org/10.1109/TGRS.2004.827257

Friedl, M.A., McGwire, K.C., McIver, D.K., 2001. An overview of uncertainty in optical remotely sensed data for ecological applications. Spat. Uncertain. Ecol. Implic. Remote Sens. GIS Appl. 258–283.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–67.

Friedman, J.H., Steppel, S., 1974. A NONPARAMETRIC PROCEDURE FOR COMPARING MULTIVARIATE POINT SETS.

Frisbie, P., 2001. Human Ecology: Insights on Demographic Behavior, in: International Encyclopedia of the Social & Behavioral Sciences. Elsevier, pp. 6978–6984. https://doi.org/10.1016/B0-08-043076-7/02091-X

Fritz, M.M.C., Mcquilken, J., Collins, N., Weldegiorgis, F., 2018. Global Trends in Artisanal and Small-Scale Mining (ASM): A review of key numbers and issues. Work. Pap., Working Papers.

Gad, S., Kusky, T., 2007. ASTER spectral ratioing for lithological mapping in the Arabian–Nubian shield, the Neoproterozoic Wadi Kid area, Sinai, Egypt. Gondwana Res. 11, 326–335. https://doi.org/10.1016/j.gr.2006.02.010

Gahlan, H., Ghrefat, H., 2018. Detection of Gossan Zones in Arid Regions Using Landsat 8 OLI Data: Implication for Mineral Exploration in the Eastern Arabian Shield, Saudi Arabia. Nat. Resour. Res. 27, 109–124. https://doi.org/10.1007/s11053-017-9341-8

Gallwey, J., Robiati, C., Coggan, J., Vogt, D., Eyre, M., 2020. A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. Remote Sens. Environ. 248, 111970. https://doi.org/10.1016/j.rse.2020.111970

Garcia Arredondo, M., Lawrence, C.R., Schulz, M.S., Tfaily, M.M., Kukkadapu, R., Jones, M.E., Boye, K., Keiluweit, M., 2019. Root-driven weathering impacts on mineral-organic associations in deep soils over pedogenic time scales. Geochim. Cosmochim. Acta 263, 68–84. https://doi.org/10.1016/j.gca.2019.07.030

Gasmi, A., Gomez, C., Zouari, H., Masse, A., Ducrot, D., 2016. PCA and SVM as geo-computational methods for geological mapping in the southern of Tunisia, using ASTER remote sensing data set. Arab. J. Geosci. 9, 753. https://doi.org/10.1007/s12517-016-2791-1

Gass, I.G., 1977. The evolution of the Pan African crystalline basement in NE Africa and Arabia. J. Geol. Soc. 134, 129–138. https://doi.org/10.1144/gsjgs.134.2.0129

Genna, A., Nehlig, P., Le Goff, E., Guerrot, C., Shanti, M., 2002. Proterozoic tectonism of the Arabian Shield. Precambrian Res. 117, 21–40. https://doi.org/10.1016/S0301-9268(02)00061-X

Ghosh, A., Trivedi, P.P., Timbalia, S.A., Griffin, A.T., Rahn, J.J., Chan, S.S.L., Gohil, V.M., 2014. Copper supplementation restores cytochrome c oxidase assembly defect in a mitochondrial disease model of

COA6 deficiency. Hum. Mol. Genet. 23, 3596–3606. https://doi.org/10.1093/hmg/ddu069

Gibb, H., O'Leary, K.G., 2014. Mercury Exposure and Health Impacts among Individuals in the Artisanal and Small-Scale Gold Mining Community: A Comprehensive Review. Environ. Health Perspect. 122, 667–672. https://doi.org/10.1289/ehp.1307864

Gibbs, M.T., Kump, L.R., 1994. Global chemical erosion during the Last Glacial Maximum and the present: Sensitivity to changes in lithology and hydrology. Paleoceanography 9, 529–543. https://doi.org/10.1029/94pa01009

Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H.H., Manning, A.H., Van Beek, L.P.H., Jellinek, A.M., 2011. Mapping permeability over the surface of the Earth: MAPPING GLOBAL PERMEABILITY. Geophys. Res. Lett. 38, n/a-n/a. https://doi.org/10.1029/2010gl045565

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for Multi-Class Classification: an Overview. https://doi.org/10.48550/arXiv.2008.05756

Groves, D.I., Goldfarb, R.J., Gebre-Mariam, M., Hagemann, S.G., Robert, F., 1998. Orogenic gold deposits: A proposed classification in the context of their crustal distribution and relationship to other gold deposit types. Ore Geol. Rev. 13, 7–27. https://doi.org/10.1016/S0169-1368(97)00012-7

Gupta, M. (Ed.), 2025. Remote sensing for geophysicists, First edition. ed. CRC Press, Boca Raton, FL.

H., E., Ibouh, H., Bachnou, A., Ait Babram, M., El Harti, A., 2016. Mapping and analysis of geological fractures extracted by remote sensing on Landsat TM images, example of the Imilchil-Tounfite area (Central High Atlas, Morocco). Estud. Geológicos 72–2), 1–12. https://doi.org/10.3989/egeol.42328.394

Hamandawana, H., Eckardt, F., Ringrose, S., 2006. The use of step-wise density slicing in classifying high-resolution panchromatic photographs. Int. J. Remote Sens. 27, 4923–4942. https://doi.org/10.1080/01431160600857436

Hamimi, Z., 2021. The geology of the Arabian-Nubian Shield, 1st ed. ed, Regional geology reviews. Springer international publishing, Cham.

Hamimi, Z., El-Kazzaz, Y., Fawzy, K., Abdelrahman, E., El-Shafei, M., Elfakharani, A., 2014. Editorial: Geology and Tectonic Setting of the Arabian-Nubian Shield. Open Geol. J. 8, 1–2. https://doi.org/10.2174/1874262901408010001

Han, L., Fuqiang, L., Zheng, D., Weixu, X., 2018. A lithology identification method for continental shale oil reservoir based on BP neural network. J. Geophys. Eng. 15, 895–908. https://doi.org/10.1088/1742-2140/aaa4db

Han, W., Zhang, Xiaohan, Wang, Yi, Wang, L., Huang, X., Li, J., Wang, S., Chen, W., Li, X., Feng, R., Fan, R., Zhang, Xinyu, Wang, Yuewei, 2023. A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. ISPRS J. Photogramm. Remote Sens.

202, 87–113. https://doi.org/10.1016/j.isprsjprs.2023.05.032

Han, Y., Yu, J., Zhang, N., Meng, C., Ma, P., Zhong, W., Zou, C., 2025. Leverage Classifier: Another Look at Support Vector Machine. Stat. Sin. https://doi.org/10.5705/ss.202023.0124

HASSAN, M.A.A., KOTEL'NIKOV, A.E., 2020. The geological and structural controls of gold mineralization in Qala en Nahal-Um Sagata region, South Gedarif, Sudan. NEWS Ural State Min. Univ. 59, 19–26. https://doi.org/10.21440/2307-2091-2020-3-19-26

Hdeid, O.M., Morsli, Y., Raji, M., Baroudi, Z., Adjour, M., Nebagha, K.C., Arby, Z.E., Moktar, V.M.E., Vall, I.B., 2024. Application of Remote Sensing and GIS in Mineral Alteration Mapping and Lineament Extraction Case of Oudiane Elkharoub (Requibat Shield, Northern of Mauritania). Open J. Geol. 14, 823–854. https://doi.org/10.4236/ojg.2024.149036

Hebri, D., Kirubakaran, S., Jasmine, R.R., Poosamooper, A., Devi, S.R., Joseph, D.R., 2024. Analysis of Satellite Images for Historical Maps Classification using Machine Learning Algorithms, in: 2024 International Conference on Expert Clouds and Applications (ICOECA). Presented at the 2024 International Conference on Expert Clouds and Applications (ICOECA), IEEE, Bengaluru, India, pp. 632–637. https://doi.org/10.1109/ICOECA62351.2024.00115

Hentschel, T.H., Felix Hruschka, Michael Priester, 2002. Global Report on Artisanal & Small-Scale Mining (No.

70). the International Institute for Environment and Development (IIED).
Hesselbarth, M.H.K., Sciaini, M., Nowosad, J., Hanss, S., 2025. landscapemetrics: Landscape Metrics for Categorical Map Patterns. https://doi.org/10.32614/CRAN.package.landscapemetrics
Hills, R.S., 2014. Ratio Image Processing Techniques : A Prospecting Tool for Mineral Deposits , RATIO IMAGE PROCESSING TECHNIQUES : A PROSPECTING TOOL FOR. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXVII, 1981–1984.
Hilson, G., Maconachie, R., 2020. Artisanal and small-scale mining and the Sustainable Development Goals: Opportunities and new directions for sub-Saharan Africa. Geoforum 111, 125–141. https://doi.org/10.1016/j.geoforum.2019.09.006
Hilson, G., McQuilken, J., 2014. Four decades of support for artisanal and small-scale mining in sub-Saharan Africa: A critical review. Extr. Ind. Soc. 1, 104–118. https://doi.org/10.1016/j.exis.2014.01.002
Hinton, 2003. Communities and Small-Scale Mining (CASM) Initiative.
Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441. https://doi.org/10.1037/h0071325
Huber, F., Yushchenko, A., Stratmann, B., Steinhage, V., 2022. Extreme Gradient Boosting for Yield Estimation compared with Deep Learning

Approaches. https://doi.org/10.48550/ARXIV.2208.12633

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 14, 55–63. https://doi.org/10.1109/TIT.1968.1054102

Ibrahim, W.S., Watanabe, K., Yonezu, K., 2016. Structural and litho-tectonic controls on Neoproterozoic base metal sulfide and gold mineralization in North Hamisana shear zone, South Eastern Desert, Egypt: The integrated field, structural, Landsat 7 ETM + and ASTER data approach. Ore Geol. Rev. 79, 62–77. https://doi.org/10.1016/j.oregeorev.2016.05.012

ICMM, 2021. Social Progress in Mining-Dependent Countries: Analysing the Role of Resource Governance in Delivering the UN Sustainable Development Goals [WWW Document]. URL https://www.icmm.com/en-gb/research/social-performance/2021/social-progress (accessed 2.9.25).

Jaiswal, I., Bharadwaj, A., Kumari, K., Agarwal, N., 2024. Credit Card Deception Recognition Using Random Forest Machine Learning Algorithm. EAI Endorsed Trans. Internet Things 10. https://doi.org/10.4108/eetiot.5347

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R, 1st ed. 2013, Corr. 7th printing 2017 edition. ed. Springer, New York.

Javidan, N., Kavian, A., Pourghasemi, H.R., Conoscenti, C., Jafarian, Z., 2019. Gully erosion susceptibility

mapping using multivariate adaptive regression splines—replications and sample size scenarios. Water 11, 2319.

Jensen, J.R., 2016. Introductory digital image processing: a remote sensing perspective, Pearson series in geographic information science. Pearson Education, Inc, Glenview, IL.

Johnson, P.R., Andresen, A., Collins, A.S., Fowler, A.R., Fritz, H., Ghebreab, W., Kusky, T., Stern, R.J., 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. J. Afr. Earth Sci. 61, 167–232. https://doi.org/10.1016/j.jafrearsci.2011.07.003

Johnson, P.R., Woldehaimanot, B., 2003. Development of the Arabian-Nubian Shield: perspectives on accretion and deformation in the northern East African Orogen and the assembly of Gondwana. Geol. Soc. Lond. Spec. Publ. 206, 289–325. https://doi.org/10.1144/GSL.SP.2003.206.01.15

Johnson, P.R., Zoheir, B.A., Ghebreab, W., Stern, R.J., Barrie, C.T., Hamer, R.D., 2017. Gold-bearing volcanogenic massive sulfides and orogenic-gold deposits in the Nubian Shield. South Afr. J. Geol. 120, 63–76. https://doi.org/10.25131/gssajg.120.1.63

Kao, Y.-H., Van Roy, B., 2014. Directed Principal Component Analysis. Oper. Res. 62, 957–972. https://doi.org/10.1287/opre.2014.1290

Kar, S., Singh, P.K., 1979. Effect of nutrients on the toxicity of pesticides carbofuran and hexachlorocyclohexane to blue-green alga Nostoc muscorum. Z. Allg. Mikrobiol. 19, 467–472. https://doi.org/10.1002/jobm.3630190703

Karamizadeh, S., Abdullah, S.M., Manaf, A.A., Zamani, M., Hooman, A., 2013. An Overview of Principal Component Analysis. J. Signal Inf. Process. 04, 173–175. https://doi.org/10.4236/jsip.2013.43B031

Kelley, M.E., 2010. Principal Component Analysis, in: Weiner, I.B., Craighead, W.E. (Eds.), The Corsini Encyclopedia of Psychology. Wiley, pp. 1–1. https://doi.org/10.1002/9780470479216.corpsy0709

Kenea, N.H., 1997. Digital enhancement of landsat data, spectral analysis and GIS data integration for geological studies of the Derudeb Area, Southern Red Sea Hills, NE Sudan. Berliner geowissenschaftliche Abhandlungen Reihe D, Geoinformatik. Fachbereich Geowiss., FU Berlin, Berlin.

Kim, E., Moon, J., Shim, J., Hwang, E., 2024. Predicting invasive species distributions using incremental ensemble-based pseudo-labeling. Ecol. Inform. 79, 102407. https://doi.org/10.1016/j.ecoinf.2023.102407

Klemenic, P.M., Poole, S., 1988. The geology and geochemistry of Upper Proterozoic granitoids from the Red Sea Hills, Sudan. J. Geol. Soc. 145, 635–643. https://doi.org/10.1144/gsjgs.145.4.0635

Kooperberg, C., 2012. Multivariate Adaptive Regression Splines, in: El-Shaarawi, A.H., Piegorsch, W.W. (Eds.), Encyclopedia of Environmetrics. Wiley. https://doi.org/10.1002/9780470057339.vaa008

Kröner, A., 1985. Ophiolites and the evolution of tectonic boundaries in the late proterozoic Arabian—Nubian shield of northeast Africa and Arabia. Precambrian Res. 27, 277–300. https://doi.org/10.1016/0301-9268(85)90016-6

Kröner, A., Grieling, R., Reischmann, T., Hussein, I.M., Stern, R.J., Dürr, S., Krüger, J., Zimmer, M., 1987. Pan-African crustal evolution in the Nubian segment of northeast Africa, in: Kröner, A. (Ed.), Geodynamics Series. American Geophysical Union, Washington, D. C., pp. 235–257. https://doi.org/10.1029/GD017p0235

Kruse, F.A., Boardman, J.W., Huntington, J.F., 2003. Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping. IEEE Trans. Geosci. Remote Sens. 41, 1388–1400. https://doi.org/10.1109/TGRS.2003.812908

Kulkarni, V., Sinha, P.K., 2014. Effective Learning and Classification using Random Forest Algorithm.

Kumar, R., Ruhel, R., Van Wijnen, A.J., 2024. Unlocking biological complexity: the role of machine learning in integrative multi-omics. Acad. Biol. 2. https://doi.org/10.20935/AcadBiol7428

Küster, D., Jean-Paul, L., 2001. Sr, Nd isotopes and geochemistry of the Bayuda Desert high-grade metamorphic basement (Sudan): an early Pan-

African oceanic convergent margin, not the edge of the East Saharan ghost craton? Precambrian Res. 109, 1–23. https://doi.org/10.1016/S0301-9268(00)00147-9

Kwang, C., Afele, I., 2024. Exploring Different Machine Learning Algorithms for Evaluating Illegal Small-Scale Mining (Galamsey) Impacts on Vegetation: A Comparative Analysis in Ghana's High Forest Zone. https://doi.org/10.2139/ssrn.4690577

Labdaoui, B., Benali, H., Boughacha, A., Moussaoui, K., 2023. Mapping hydrothermal alterations and lineaments associated with epithermal and massive sulphides deposits of Tifraouine (northwest Algerian coast): Use of Landsat 8 OLI data and remote sensing. Rev. Soc. Geológica Esp. 36, 3–15. https://doi.org/10.55407/rsge.96703

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Lee, S., Oh, H.J., 2011. Application of artificial neural network for mineral potential mapping. Artif. Neural Netw.-Appl. Shanghai Tech 67–104.

Li, C., Wang, J., Wang, L., Hu, L., Gong, P., 2014. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat

Thematic Mapper Imagery. Remote Sens. 6, 964–983. https://doi.org/10.3390/rs6020964

Likó, S.B., Holb, I.J., Oláh, V., Burai, P., Szabó, S., 2024. Deep learning-based training data augmentation combined with post-classification improves the classification accuracy for dominant and scattered invasive forest tree species. Remote Sens. Ecol. Conserv. 10, 203–219. https://doi.org/10.1002/rse2.365

Lillesand, T., Kiefer, R., 2015. Chipman., JW (2015). Remote Sens. Image Interpret. 7th Ed. Photogramm. Eng. Remote Sens. 7th Edn Vol 81 Wiley Httpsdoi Org1014358pers 81, 615.

Lin, N., Fu, J., Jiang, R., Li, G., Yang, Q., 2023. Lithological Classification by Hyperspectral Images Based on a Two-Layer XGBoost Model, Combined with a Greedy Algorithm. Remote Sens. 15, 3764. https://doi.org/10.3390/rs15153764

Ling, Q., 2023. Machine learning algorithms review. Appl. Comput. Eng. 4, 91–98. https://doi.org/10.54254/2755-2721/4/20230355

Liu, L., Zhuang, D.-F., Zhou, J., Qiu, D.-S., 2011. Alteration mineral mapping using masking and Crosta technique for mineral exploration in mid-vegetated areas: a case study in Areletuobie, Xinjiang (China). Int. J. Remote Sens. 32, 1931–1944. https://doi.org/10.1080/01431161003639678

Loughlin, W.P., 1991. PRINCIPAL COMPONENT ANALYSIS FOR ALTERATION MAPPING. Photogramm. Eng. Remote Sens. 57, 1163–1169.

Madani, A.A., 2009. Utilization of Landsat ETM+ data for mapping gossans and iron rich zones exposed at Bahrah area, Western Arabian Shield, Saudi Arabia. J. King Abdulaziz Univ. Earth Sci. 20, 25–49.

Mahmood, T.H., Hasan, K., Akhter, S.H., 2019. Lithologic mapping of a forested montane terrain from Landsat 5 TM image. Geocarto Int. 34, 750–768. https://doi.org/10.1080/10106049.2018.1434688

Maindola, M., Al-Fatlawy, R.R., Kumar, R., Boob, N.S., Sreeja, S.P., Sirisha, N., Srivastava, A., 2024. Utilizing Random Forests for High-Accuracy Classification in Medical Diagnostics, in: 2024 7th International Conference on Contemporary Computing and Informatics (IC3I). Presented at the 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Greater Noida, India, pp. 1679–1685. https://doi.org/10.1109/IC3I61595.2024.10828609

Mangeruga, M., Casavola, A., Pupo, F., Bruno, F., 2020. An Underwater Pathfinding Algorithm for Optimised Planning of Survey Dives. Remote Sens. 12, 3974. https://doi.org/10.3390/rs12233974

Maron, M.E., 1961. Automatic indexing: an experimental inquiry. J. ACM JACM 8, 404–417.

Matcharashvili, T., Czechowski, Z., Zhukova, N., 2019. Mahalanobis distance-based recognition of changes in the dynamics of a seismic process. Nonlinear Process. Geophys. 26, 291–305. https://doi.org/10.5194/npg-26-291-2019

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Maxwell, B.A., Gwon, Y., Mishra, A., Peng, J., Nakamura, H., Zhang, K., Kim, H.J., Taylor, J.P., 2021. Ubiquitination is essential for recovery of cellular activities after heat shock. Science 372, eabc3593. https://doi.org/10.1126/science.abc3593

Mead, R.A., Sharik ,Terry L., Prisley ,Stephen P., and Heinen, J.T., 1981. A Computerized Spatial Analysis System for Assessing Wildlife Habitat from Vegetation Maps. Can. J. Remote Sens. 7, 34–40. https://doi.org/10.1080/07038992.1981.10855007

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790.

Mesa, F., Velez, G.C., Cardenas Alzate, P.P., 2018. Analysis of economic and social indicators through the principal components analysis. Contemp. Eng. Sci. 11, 763–769. https://doi.org/10.12988/ces.2018.8120

Mhangara, P., Tsoeleng, L.T., Mapurisa, W., 2020. Monitoring the development of artisanal mines in South Africa. J. South. Afr. Inst. Min. Metall. 120, 299–306. https://doi.org/10.17159/2411-9717/938/2020

Mielke, C., Boesche, N., Rogass, C., Kaufmann, H., Gauert, C., De Wit, M., 2014. Spaceborne Mine Waste Mineralogy Monitoring in South Africa, Applications for Modern Push-Broom Missions: Hyperion/OLI and EnMAP/Sentinel-2. Remote Sens. 6, 6790–6816. https://doi.org/10.3390/rs6086790

Miller, N.R., Stern, R.J., 2021. Evolution of the Arabian Nubian Shield and Snowball Earth, in: Hamimi, Z., Fowler, A.-R., Liégeois, J.-P., Collins, A., Abdelsalam, M.G., Abd EI-Wahed, M. (Eds.), The Geology of the Arabian-Nubian Shield, Regional Geology Reviews. Springer International Publishing, Cham, pp. 153–194. https://doi.org/10.1007/978-3-030-72995-0_7

Minamata Convention: Initial assessment in the Republic of Sudan [WWW Document], 2015. . Glob. Environ. Facil. URL https://www.thegef.org/projects-operations/projects/9345 (accessed 5.29.25).

Moeck, I.S., 2014. Catalog of geothermal play types based on geologic controls. Renew. Sustain. Energy Rev. 37, 867–882. https://doi.org/10.1016/j.rser.2014.05.032

Mohamed Taha, M., Xi, Y., He, Q., Hu, A., Wang, S., Liu, X., 2022. Investigating the Capabilities of Various Multispectral Remote Sensors Data to Map Mineral Prospectivity Based on Random Forest Predictive Model: A Case Study for Gold Deposits in Hamissana Area, NE Sudan. Minerals 13, 49. https://doi.org/10.3390/min13010049

Mohammed, S.B., Khalid, A., Osman, S.E.F., Helali, R.G.M., 2016. Usage of Principal Component Analysis (PCA) in AI Applications. Int. J. Eng. Res. 5.

Moomen, A.-W., Lacroix, P., Benvenuti, A., Planque, M., Piller, T., Davis, K., Miranda, M., Ibrahim, E., Giuliani, G., 2022. Assessing the Applications of Earth Observation Data for Monitoring Artisanal and Small-Scale Gold Mining (ASGM) in Developing Countries. Remote Sens. 14, 2971. https://doi.org/10.3390/rs14132971

Mori, F., Mendicelli, A., Moscatelli, M., Romagnoli, G., Peronace, E., Naso, G., 2020. A new Vs30 map for Italy based on the seismic microzonation dataset. Eng. Geol. 275, 105745. https://doi.org/10.1016/j.enggeo.2020.105745

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

Mrinmoy Dhara, Sengar, V.K., Shovan L. Chattoraj, Soumiya Bhattacharjee, 2017. Mapping Of Alteration Zones In Mineral Rich Belt Of South-East Rajasthan Using Remote Sensing Techniques. https://doi.org/10.5281/ZENODO.1128817

Muavhi, N., 2022. Application of per-pixel and sub-pixel unmixing methods on ASTER images to map hydrothermal alterations in a highly metamorphosed terrain: a case of the Musina copper deposit field. Geocarto Int. 37, 13579–13595. https://doi.org/10.1080/10106049.2022.2082547

Mwaniki, M.W., Moeller, M.S., Schellmann, G., 2015. A comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) in mapping geology and visualising lineaments: A case study of central region Kenya. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-7/W3, 897–903. https://doi.org/10.5194/isprsarchives-XL-7-W3-897-2015

Nagpal, S.J.S., Chari, S.T., 2019. Immunoglobulin G4 Levels. JAMA 321, 202. https://doi.org/10.1001/jama.2018.16665

Nair, P., Srivastava, D.K., Bhatnagar, R., 2023. Application of Machine Learning in Mineral Mapping Using Remote Sensing, in: Choudrie, J., Mahalle, P., Perumal, T., Joshi, A. (Eds.), IOT with Smart Systems. Springer Nature, Singapore, pp. 27–35. https://doi.org/10.1007/978-981-19-3575-6_4

Ngom, N.M., Baratoux, D., Bolay, M., Dessertine, A., Abass Saley, A., Baratoux, L., Mbaye, M., Faye, G., Yao, A.K., Kouamé, K.J., 2023. Artisanal Exploitation of Mineral Resources: Remote Sensing Observations of Environmental Consequences, Social and Ethical Aspects. Surv. Geophys. 44, 225–247. https://doi.org/10.1007/s10712-022-09740-1

Noetstaller, R., 1987. Small-Scale Mining A Review of the Issues (No. 23). Washington, D.C.

Novaes, R.C., Souza, L.E.D., Temporim, F.A., Abichequer, L.A., 2024. Hydrothermal Alteration Zones Mapping Using Sentinel 2 MSI and Terra Aster Data in The Uruguay Mine, Minas do Camaquã District. Rev.

Gest. Soc. E Ambient. 18, e6142. https://doi.org/10.24857/rgsa.v18n8-085

Nugroho, H., Wikantika, K., Bijaksana, S., Saepuloh, A., 2023. Integration of remote sensing and geophysical data to enhance lithological mapping utilizing the Random Forest classifier: a case study from Komopa, Papua Province, Indonesia. J. Degraded Min. Lands Manag. 10, 4417. https://doi.org/10.15243/jdmlm.2023.103.4417

Nursamsi, I., Phinn, S.R., Levin, N., Luskin, M.S., Sonter, L.J., 2024. Remote sensing of artisanal and small-scale mining: A review of scalable mapping approaches. Sci. Total Environ. 951, 175761. https://doi.org/10.1016/j.scitotenv.2024.175761

Nyamekye, C., Ghansah, B., Agyapong, E., Kwofie, S., 2021. Mapping changes in artisanal and small-scale mining (ASM) landscape using machine and deep learning algorithms. - a proxy evaluation of the 2017 ban on ASM in Ghana. Environ. Chall. 3, 100053. https://doi.org/10.1016/j.envc.2021.100053

OECD, 2017. Supply Chain Due Diligence. Sphera. URL https://sphera.com/solutions/supply-chain-transparency/supply-chain-sustainability-solution/supply-chain-due-diligence/ (accessed 2.9.25).

Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, 2011. Scikit-learn: Machine Learning in Python.

Othman, A.A., Gloaguen, R., 2017. Integration of spectral, spatial and morphometric data into lithological

mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq. J. Asian Earth Sci. 146, 90–102.

Othman, A.A., Gloaguen, R., 2014. Improving lithological mapping by SVM classification of spectral and morphological features: The discovery of a new chromite body in the Mawat ophiolite complex (Kurdistan, NE Iraq). Remote Sens. 6, 6867–6896.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26, 217–222. https://doi.org/10.1080/01431160412331269698

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011. https://doi.org/10.1080/01431160512331314083

Park, S., Hamm, S.-Y., Jeon, H.-T., Kim, J., 2017. Evaluation of Logistic Regression and Multivariate Adaptive Regression Spline Models for Groundwater Potential Mapping Using R and GIS. Sustainability 9, 1157. https://doi.org/10.3390/su9071157

Pearson, K., 1901. LIII. *On lines and planes of closest fit to systems of points in space*. Lond. Edinb. Dublin Philos. Mag. J. Sci. 2, 559–572. https://doi.org/10.1080/14786440109462720

Pedregosa, F., Varoquaux, G., Gramfort, A., 2011. scikit-learn.

Pelletier, L., Shanmugasegaram, S., Patten, S.B., Demers, A., 2017. Self-management of mood and/or anxiety disorders through physical activity/exercise. Health

Promot. Chronic Dis. Prev. Can. 37, 149–159. https://doi.org/10.24095/hpcdp.37.5.03

Peng, Z., Zhang, Y., 2018. Human motion classification based on inertial sensors with extreme gradient boosting, in: Su, R. (Ed.), 2018 International Conference on Image and Video Processing, and Artificial Intelligence. Presented at the 2018 International Conference on Image, Video Processing and Artificial Intelligence, SPIE, Shanghai, China, p. 99. https://doi.org/10.1117/12.2514563

Pour, A.B., Hashim, M., Park, Y., Hong, J.K., 2018. Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. Geocarto Int. 33, 1281–1306. https://doi.org/10.1080/10106049.2017.1347207

Pour, Park, Park, Hong, Muslim, Läufer, Crispini, Pradhan, Zoheir, Rahmani, Hashim, Hossain, 2019. Landsat-8, Advanced Spaceborne Thermal Emission and Reflection Radiometer, and WorldView-3 Multispectral Satellite Imagery for Prospecting Copper-Gold Mineralization in the Northeastern Inglefield Mobile Belt (IMB), Northwest Greenland. Remote Sens. 11, 2430. https://doi.org/10.3390/rs11202430

Price, G., 2024. Gold Price History [WWW Document]. Gold Price. URL https://goldprice.org/gold-price-history.html (accessed 1.6.25).

Pristyanto, Y., Mukarabiman, Z., Nugraha, A.F., 2023. Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass

Imbalanced Dataset. JOIV Int. J. Inform. Vis. 7, 710–715. https://doi.org/10.30630/joiv.7.3.1102

Qiu, F., Abdelsalam, M., Thakkar, P., 2006. Spectral analysis of ASTER data covering part of the Neoproterozoic Allaqi-Heiani suture, Southern Egypt. J. Afr. Earth Sci. 44, 169–180. https://doi.org/10.1016/j.jafrearsci.2005.10.009

Quinn, D.P., Ehlmann, B.L., 2019. A PCA-Based Framework for Determining Remotely Sensed Geological Surface Orientations and Their Statistical Quality. Earth Space Sci. 6, 1378–1408. https://doi.org/10.1029/2018EA000416

Qureshi, M.U., Mahmood, Z., Rasool, A.M., 2022. Using multivariate adaptive regression splines to develop relationship between rock quality designation and permeability. J. Rock Mech. Geotech. Eng. 14, 1180–1187.

R Core Team, R., 2024. R: The R Project for Statistical Computing [WWW Document]. URL https://www.r-project.org/ (accessed 6.4.25).

Rajan Girija, R., Mayappan, S., 2019. Mapping of mineral resources and lithological units: a review of remote sensing techniques. Int. J. Image Data Fusion 10, 79–106. https://doi.org/10.1080/19479832.2019.1589585

Ramos, M.M., Bijani, R., Santos, F.V., Lupinacci, W.M., Freire, A.F.M., 2023. Analysis of alternative strategies applied to Naïve-Bayes classifier into the recognition of electrofacies: Application in well-log data at Recôncavo Basin, North-East Brazil.

Geoenergy Sci. Eng. 227, 211889. https://doi.org/10.1016/j.geoen.2023.211889

Ren, Q., Zhang, H., Zhang, D., Zhao, X., Yan, L., Rui, J., Zeng, F., Zhu, X., 2022. A framework of active learning and semi-supervised learning for lithology identification based on improved naive Bayes. Expert Syst. Appl. 202, 117278. https://doi.org/10.1016/j.eswa.2022.117278

Reymer, A., Schubert, G., 1984. Phanerozoic addition rates to the continental crust and crustal growth. Tectonics 3, 63–77. https://doi.org/10.1029/tc003i001p00063

Rezaei, A., Hassani, H., Moarefvand, P., Golmohammadi, A., 2020. Lithological mapping in Sangan region in Northeast Iran using ASTER satellite data and image processing methods. Geol. Ecol. Landsc. 4, 59–70. https://doi.org/10.1080/24749508.2019.1585657

Rose, M., Hassen, H.R., 2019. A Survey of Random Forest Pruning Techniques, in: 9th International Conference on Computer Science, Engineering and Applications (ICCSEA 2019). Presented at the 9th International Conference on Computer Science, Engineering and Applications, Aircc publishing Corporation, pp. 99–109. https://doi.org/10.5121/csit.2019.91808

Rotigliano, E., Martinello, C., Agnesi, V., Conoscenti, C., 2018. Evaluation of debris flow susceptibility in El Salvador (CA): a comparison between Multivariate Adaptive Regression Splines (MARS) and Binary Logistic Regression (BLR). Hung. Geogr. Bull. 67,

361–373.
https://doi.org/10.15201/hungeobull.67.4.5

Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. Remote Sens. Environ. 84, 350–366. https://doi.org/10.1016/s0034-4257(02)00127-x

Ruiz, P., Mateos, J., Camps-Valls, G., Molina, R., Katsaggelos, A.K., 2013. Bayesian active remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 52, 2186–2196.

Saadi, N., Watanabe, K., 2009. Assessing image processing techniques for geological mapping: a case study in Eljufra, Libya. Geocarto Int. 24, 241–253.

Sabins, F.F., 1999. Remote sensing for mineral exploration. Ore Geol. Rev. 14, 157–183.

Saljoughi, B.S., Hezarkhani, A., 2018. A comparative analysis of artificial neural network (ANN), wavelet neural network (WNN), and support vector machine (SVM) data-driven models to mineral potential mapping for copper mineralizations in the Shahr-e-Babak region, Kerman, Iran. Appl. Geomat. 10, 229–256. https://doi.org/10.1007/s12518-018-0229-z

Samat, A., Li, E., Wang, W., Liu, S., Lin, C., Abuduwaili, J., 2020. Meta-XGBoost for Hyperspectral Image Classification Using Extended MSER-Guided Morphological Profiles. Remote Sens. 12, 1973. https://doi.org/10.3390/rs12121973

Sarro, R., María Mateos, R., Reichenbach, P., Aguilera, H., Riquelme, A., Hernández-Gutiérrez, L.E., Martín, A., Barra, A., Solari, L., Monserrat, O., Alvioli, M., Fernández-Merodo, J.A., López-Vinielles, J., Herrera, G., 2020. Geotechnics for rockfall assessment in the volcanic island of Gran Canaria (Canary Islands, Spain). J. Maps 16, 605–613. https://doi.org/10.1080/17445647.2020.1806125

Sayed, F., Hammed, M., Shided, A., Hussein, A., 2023. Implementation of Remote Sensing Techniques in Structural and Lithological Mapping of Northwestern Margin of Red Sea Egypt. J. Min. Environ. 14. https://doi.org/10.22044/jme.2023.12628.2294

Schandelmeier, H., Richter, A., Harms, U., 1987. Proterozoic deformation of the East Saharan Craton in Southeast Libya, South Egypt and North Sudan. Tectonophysics 140, 233–246. https://doi.org/10.1016/0040-1951(87)90231-9

Schonlau, M., Zou, R.Y., 2020. The random forest algorithm for statistical learning. Stata J. Promot. Commun. Stat. Stata 20, 3–29. https://doi.org/10.1177/1536867X20909688

Schwartz, F.W., Lee, S., Darrah, T.H., 2021. A Review of Health Issues Related to Child Labor and Violence Within Artisanal and Small-Scale Mining. GeoHealth 5, e2020GH000326. https://doi.org/10.1029/2020GH000326

Sebtosheikh, M.A., Motafakkerfard, R., Riahi, M.A., Moradi, S., Sabety, N., 2015. Support vector machine

method, a new technique for lithology prediction in an Iranian heterogeneous carbonate reservoir using petrophysical well logs. Carbonates Evaporites 30, 59–68. https://doi.org/10.1007/s13146-014-0199-0

Seccatore, J., Veiga, M., Origliasso, C., Marin, T., De Tomi, G., 2014. An estimation of the artisanal small-scale production of gold in the world. Sci. Total Environ. 496, 662–667. https://doi.org/10.1016/j.scitotenv.2014.05.003

Shaker, A., 2023. Chapter 6 Effect sizes | STM1001 Topic 7: One-way ANOVA.

Shanthi, R., D, B., Venkatesh, R., Sankari, C., Manimaran, V., Basha, S.H., 2024. Enhancing Remote Sensing Image Classification Using Arithmetic Optimization Algorithm with Deep Learning Approach, in: 2024 Second International Conference on Advances in Information Technology (ICAIT). Presented at the 2024 Second International Conference on Advances in Information Technology (ICAIT), IEEE, Chikkamagaluru, Karnataka, India, pp. 1–6. https://doi.org/10.1109/ICAIT61638.2024.10690775

Shanti, M., Roobol, M.J., 1979. A late Proterozoic ophiolite complex at Jabal Ess in northern Saudi Arabia. Nature 279, 488–491. https://doi.org/10.1038/279488a0

Shebl, A., Abdellatif, M., Badawi, M., Dawoud, M., Fahil, A.S., Csámer, Á., 2023a. Towards better delineation of hydrothermal alterations via multi-sensor remote

sensing and airborne geophysical data. Sci. Rep. 13, 7406. https://doi.org/10.1038/s41598-023-34531-y

Shebl, A., Abdellatif, M., Elkhateeb, S.O., Csámer, Á., 2021. Multisource Data Analysis for Gold Potentiality Mapping of Atalla Area and Its Environs, Central Eastern Desert, Egypt. Minerals 11, 641. https://doi.org/10.3390/min11060641

Shebl, A., Abriha, D., Fahil, A.S., El-Dokouny, H.A., Elrasheed, A.A., Csámer, Á., 2023b. PRISMA hyperspectral data for lithological mapping in the Egyptian Eastern Desert: Evaluating the support vector machine, random forest, and XG boost machine learning algorithms. Ore Geol. Rev. 161, 105652. https://doi.org/10.1016/j.oregeorev.2023.105652

Shebl, A., Kusky, T., Csámer, Á., 2022. Advanced land imager superiority in lithological classification utilizing machine learning algorithms. Arab. J. Geosci. 2022 159 15, 1–13. https://doi.org/10.1007/S12517-022-09948-W

Shim, J.W., 2024. Enhancing cross entropy with a linearly adaptive loss function for optimized classification performance. Sci. Rep. 14, 27405. https://doi.org/10.1038/s41598-024-78858-6

Shim, K., Yu, J., Wang, L., Lee, S., Koh, S.-M., Lee, B.H., 2021. Content Controlled Spectral Indices for Detection of Hydrothermal Alteration Minerals Based on Machine Learning and Lasso-Logistic Regression Analysis. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7435–7447. https://doi.org/10.1109/jstars.2021.3095926

Shirmard, H., Farahbakhsh, E., Heidari, E., Pour, A.B., Pradhan, B., Müller, D., Chandra, R., 2022a. A Comparative Study of Convolutional Neural Networks and Conventional Machine Learning Models for Lithological Mapping Using Remote Sensing Data. Remote Sens. 2022 Vol 14 Page 819 14, 819. https://doi.org/10.3390/RS14040819

Shirmard, H., Farahbakhsh, E., Müller, R.D., Chandra, R., 2022b. A review of machine learning in processing remote sensing data for mineral exploration. Remote Sens. Environ. 268, 112750. https://doi.org/10.1016/j.rse.2021.112750

Simionato, J., Bertani, G., Osako, L.S., 2021. Identification of artisanal mining sites in the Amazon Rainforest using Geographic Object-Based Image Analysis (GEOBIA) and Data Mining techniques. Remote Sens. Appl. Soc. Environ. 24, 100633. https://doi.org/10.1016/j.rsase.2021.100633

Singh, R.K., Shanmugam, P., 2016. A Multidisciplinary Remote Sensing Ocean Color Sensor: Analysis of User Needs and Recommendations for Future Developments. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 5223–5238. https://doi.org/10.1109/JSTARS.2016.2520501

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62, 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Sudan Satti, K I Khalil, M Ahmed, El-Makky, Albarra Satti, 2020. Geological and Geochemical Studies of

Granitiod Intrusion, Tagotieb area Red Sea Hills. https://doi.org/10.13140/RG.2.2.30701.82401

Sun, B., Chen, X., Zhou, Q., 2017. Analyzing the Uncertainties of Ground Validation for Remote Sensing Land Cover Mapping in the Era of Big Geographic Data, in: Zhou, C., Su, F., Harvey, F., Xu, J. (Eds.), Spatial Data Handling in Big Data Era, Advances in Geographic Information Science. Springer Singapore, Singapore, pp. 31–38. https://doi.org/10.1007/978-981-10-4424-3_3

Sunan Sakti Syah Alam, B.Y.C., Itoi, R., Taguchi, S., Saibi, H., Yamashiro, R., 2019. Hydrogeochemical and isotope characterization of geothermal waters from the Cidanau geothermal field, West Java, Indonesia. Geothermics 78, 62–69. https://doi.org/10.1016/j.geothermics.2018.11.003

Szabó, S., Holb, I.J., Abriha-Molnár, V.É., Szatmári, G., Singh, S.K., Abriha, D., 2024. Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics. Appl. Soft Comput. 155, 111468. https://doi.org/10.1016/j.asoc.2024.111468

Szaniawska, L., 2018. Lithological maps visualizing the achievements of geological sciences in the first half of the 19th century. Pol. Cartogr. Rev. 50, 87–109. https://doi.org/10.2478/pcr-2018-0006

Tariq, A., Jiango, Y., Li, Q., Gao, J., Lu, L., Soufan, W., Almutairi, K.F., Habib-ur-Rahman, M., 2023. Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel

logistic regression and naive bayes tree models with optical remote sensing data. Heliyon 9.

Telmer, K.H., Veiga, M.M., 2009. World emissions of mercury from artisanal and small scale gold mining, in: Mason, R., Pirrone, N. (Eds.), Mercury Fate and Transport in the Global Atmosphere. Springer US, Boston, MA, pp. 131–172. https://doi.org/10.1007/978-0-387-93958-2_6

Tianshi, F., Zhiguo, P., Wei, J., Guanbo, P., Hao, L., 2021. Remote sensing abnormal extraction of hydroxyl alteration based on PCA method. IOP Conf. Ser. Earth Environ. Sci. 671, 012007. https://doi.org/10.1088/1755-1315/671/1/012007

Tj, A.G., Hermansyah, H., Kristadi, H.J., Riyanto, H., 2022. IMAGE PROCESSING TECHNIQUES FOR THE INFORMATION EXTRACTION OF THE LANDSAT TM IMAGERY APPLICATION TO GEOLOGICAL MAPPING OF THE NE JAVA BASIN*). Sci. Contrib. Oil Gas 16, 10–20. https://doi.org/10.29017/scog.16.1.1109

ud-geoai, 2025. ud-geoai/Statistical-region-growing-by-vectorseeds.

UNEP, 2012. Analysis of formalization approaches in the ASGM sector based on experiences in Ecuador, Mongolia, Peru, Tanzania and Uganda | Global Mercury Partnership [WWW Document]. URL https://www.unep.org/globalmercurypartnership/resources/report/analysis-formalization-approaches-asgm-sector-based-experiences-ecuador-mongolia (accessed 7.1.25).

Vail, J.R., 1978. Outline of the geology and mineralization of the nubian shield east of the Nile Valley, Sudan. Precambrian Res. 6, A39–A40. https://doi.org/10.1016/0301-9268(78)90116-X

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. Int. J. Appl. Earth Obs. Geoinformation 14, 112–128. https://doi.org/10.1016/J.JAG.2011.08.002

Van Der Werff, H., Van Der Meer, F., 2016. Sentinel-2A MSI and Landsat 8 OLI Provide Data Continuity for Geological Remote Sensing. Remote Sens. 8, 883. https://doi.org/10.3390/rs8110883

Vanmaercke, M., Panagos, P., Vanwalleghem, T., Hayas, A., Foerster, S., Borrelli, P., Rossi, M., Torri, D., Casali, J., Borselli, L., Vigiak, O., Maerker, M., Haregeweyn, N., De Geeter, S., Zgłobicki, W., Bielders, C., Cerdà, A., Conoscenti, C., De Figueiredo, T., Evans, B., Golosov, V., Ionita, I., Karydas, C., Kertész, A., Krása, J., Le Bouteiller, C., Radoane, M., Ristić, R., Rousseva, S., Stankoviansky, M., Stolte, J., Stolz, C., Bartley, R., Wilkinson, S., Jarihani, B., Poesen, J., 2021. Measuring, modelling and managing gully erosion at large scales: A state of the art. Earth-Sci. Rev. 218, 103637. https://doi.org/10.1016/j.earscirev.2021.103637

Vapnik, V.N., 1995. The nature of statistical learning theory. Springer-Verlag.

Vapnik, V.N., 1979. Reconstruction of Dependences from Empirical Data.

Venkata Mahesh Babu Batta, 2024. Machine Learning. Int. J. Adv. Res. Sci. Commun. Technol. 583–591. https://doi.org/10.48175/IJARSCT-17677

Vojtek, M., Vojteková, J., 2019. Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. Water 11, 364. https://doi.org/10.3390/w11020364

Volpi, M., Tuia, D., Camps-Valls, G., Kanevski, M., 2011. Unsupervised change detection in the feature space using kernels, in: 2011 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2011 - 2011 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Vancouver, BC, pp. 106–109. https://doi.org/10.1109/IGARSS.2011.6048909

Wang, N., Naz, I., Aslam, R.W., Quddoos, A., Soufan, W., Raza, D., Ishaq, T., Ahmed, B., 2024. Spatio-Temporal Dynamics of Rangeland Transformation using machine learning algorithms and Remote Sensing data. Rangel. Ecol. Manag. 94, 106–118. https://doi.org/10.1016/j.rama.2024.02.008

Wang, S., Huang, X., Han, W., Li, J., Zhang, X., Wang, L., 2023. Lithological mapping of geological remote sensing via adversarial semi-supervised segmentation network. Int. J. Appl. Earth Obs. Geoinformation 125, 103536. https://doi.org/10.1016/j.jag.2023.103536

Wang, Z.H., Zheng, C.Y., 2010. Rocks/minerals information extraction from EO-1 hyperion data base on SVM. 2010 Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2010 3, 229–232. https://doi.org/10.1109/ICICTA.2010.341

Waske, B., Menz, G., Benediktsson, J.A., 2007. Fusion of support vector machines for classifying SAR and multispectral imagery from agricultural areas, in: 2007 IEEE International Geoscience and Remote Sensing Symposium. Presented at the 2007 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Barcelona, Spain, pp. 4842–4845. https://doi.org/10.1109/IGARSS.2007.4423945

Werner, T.T., Bebbington, A., Gregory, G., 2019. Assessing impacts of mining: Recent contributions from GIS and remote sensing. https://doi.org/10.1016/j.exis.2019.06.011

Whiteman, A.J., 1972. The Geology of the Sudan Republic. Geol. Mag. 109, 76–77. https://doi.org/10.1017/S0016756800042448

WHO, 2016. ENVIRONMENTAL AND OCCUPATIONAL HEALTH HAZARDS ASSOCIATED WITH ARTISANAL AND SMALL-SCALE GOLD MINING.

willem (https://stats.stackexchange.com/users/159052/willem), 2017. F1/Dice-Score vs IoU.

Williamson, M.-C., McNeil, R.J., Day, S.J., McCurdy, M.M., Rainbird, R.H., Grunsky, E.C., 2015. Environmental impact of gossans revealed by orientation surveys for base metals in the Canadian Arctic Islands.

Environ. Econ. Significance Gossansed M-C Williamson 74–84.

Xi, Y., Mohamed Taha, A.M., Hu, A., Liu, X., 2022. Accuracy comparison of various remote sensing data in lithological classification based on random forest algorithm. Geocarto Int. 37, 14451–14479. https://doi.org/10.1080/10106049.2022.2088859

Yalcin, M., Kilic Gul, F., Yildiz, A., Polat, N., Basaran, C., 2020. The mapping of hydrothermal alteration related to the geothermal activities with remote sensing at Akarcay Basin (Afyonkarahisar), using Aster data. Arab. J. Geosci. 13. https://doi.org/10.1007/s12517-020-06083-2

Ye, Z., Zhu, H., 2020. Decentralized Principal Component Analysis by Integrating Lagrange Programming Neural Networks With Alternating Direction Method of Multipliers. IEEE Access 8, 182842–182852. https://doi.org/10.1109/ACCESS.2020.2981794

Yin, X., Fallah-Shorshani, M., McConnell, R., Fruin, S., Chiang, Y.-Y., Franklin, M., 2023. Quantile Extreme Gradient Boosting for Uncertainty Quantification. https://doi.org/10.48550/ARXIV.2304.11732

Yousefi, M., Tabatabaei, S.H., Rikhtehgaran, R., Pour, A.B., Pradhan, B., 2021. Application of Dirichlet Process and Support Vector Machine Techniques for Mapping Alteration Zones Associated with Porphyry Copper Deposit Using ASTER Remote Sensing Imagery. Minerals 11, 1235. https://doi.org/10.3390/min11111235

Yu, L., Porwal, A., Holden, E.J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Comput. Geosci. 45, 229–239. https://doi.org/10.1016/j.cageo.2011.11.019

Zabihi, M., Pourghasemi, H.R., Pourtaghi, Z.S., Behzadfar, M., 2016. GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. Environ. Earth Sci. 75, 665. https://doi.org/10.1007/s12665-016-5424-9

Zareef, M., Chen, Q., Hassan, M.M., Arslan, M., Hashim, M.M., Ahmad, W., Kutsanedzie, F.Y.H., Agyekum, A.A., 2020. An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. Food Eng. Rev. 12, 173–190. https://doi.org/10.1007/s12393-020-09210-7

Zebker, H., 2021. Accuracy of a Model-Free Algorithm for Temporal InSAR Tropospheric Correction. Remote Sens. 13, 409. https://doi.org/10.3390/rs13030409

Zeinelabdein, K.A.E., Nadi, A.H.E., 2014. The use of Landsat 8 OLI image for the delineation of gossanic ridges in the Red Sea Hills of NE Sudan. Am. J. Earth Sci. 1, 62–67.

Zeng, L., Ren, W., Shan, L., Niu, Y., Liu, X., 2023. Prediction and reliability analysis of reservoir lithology spatial distribution. Front. Earth Sci. 11, 1251218. https://doi.org/10.3389/feart.2023.1251218

Zhang, L., Yang, L., Ma, T., Shen, F., Cai, Y., Zhou, C., 2021. A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data. Geoderma 384, 114809. https://doi.org/10.1016/j.geoderma.2020.114809

Zhang, W., Goh, A.T.C., 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci. Front. 7, 45–52. https://doi.org/10.1016/j.gsf.2014.10.003

Zhao, J., Lee, C.-D., Chen, G., Zhang, J., 2024. Research on the Prediction Application of Multiple Classification Datasets Based on Random Forest Model, in: 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS). Presented at the 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS), IEEE, Shenyang, China, pp. 156–161. https://doi.org/10.1109/ICPICS62053.2024.10795875

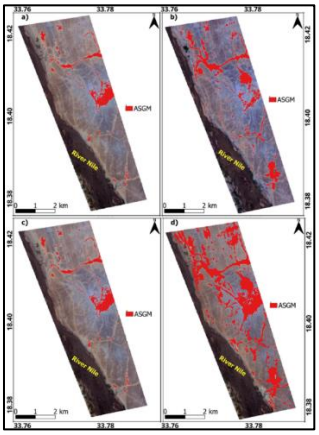
Zhen, Z., Fang, H., Liu, J., 2013. The Hierarchical Brain Network for Face Recognition. PLoS ONE 8, e59886. https://doi.org/10.1371/journal.pone.0059886

Zhu, W., Wang, Z., Yu, M., Zhang, X., Zhang, Z., 2023. Using support vector machine to explore the difference of function connection between deficit and non-deficit schizophrenia based on gray matter volume. Front. Neurosci. 17, 1132607.

Appendices

(Appendix 4-1) Accuracy measures and confusion matrices for Landsat data binary classification in 2003, 2013, and 2023 (UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy)

Binary RF			Reference					
			Non-ASM	ASM	Total	UA	PA	F1-Score
2003	Predicted	Non-ASM	361	19	380	0.93	0.912	0.935
		ASM	19	171	190	0.930	0.936	0.930
		Total	380	190	570	OA: 0.90		
2013	Predicted		Non-ASM	ASM	Total	UA	PA	F1-Score
		Non-ASM	351	29	380	0.894	0.931	0.916
		ASM	29	161	190	0.920	0.936	0.934
		Total	380	190	570	OA: 0.91		
2023	Predicted		Non-ASM	ASM	total	UA	PA	F1-Score
		Non-ASM	347	33	380	0.902	0.912	0.893
		ASM	33	157	190	0.855	0.827	0.884
		Total	380	190	570	OA: 0.88		



(Appendix 4-2) .Distributions of ASM between 2017 and 2024. a) and b) binary RF 2017 and 2024 respectively. c) and d) binary XGB classifier. ASM area (red) progressively increased based on both classifiers results.

(Appendix 4-3). Summary of Generalized Linear Model (GLM) performed with Landsat image (SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance, ω^2: effect size; $p < 0.05$, highlighted in bold).

	SS	df	F	p	ω^2
Model	0.0545	14	2.48	**0.046**	0.408
ML-models	0.0114	2	3.63	0.052	0.104
Training data	0.0163	4	2.59	0.079	0.125
ML-models × Training data	0.0268	8	2.14	0.098	0.179
Residuals	0.0236	15			
Total	0.0781	29			

(Appendix 4-4). Summary of the Generalized Linear Model (GLM) performed with the PRISMA image (SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance, ω^2: effect size; $p < 0.05$, highlighted in bold).

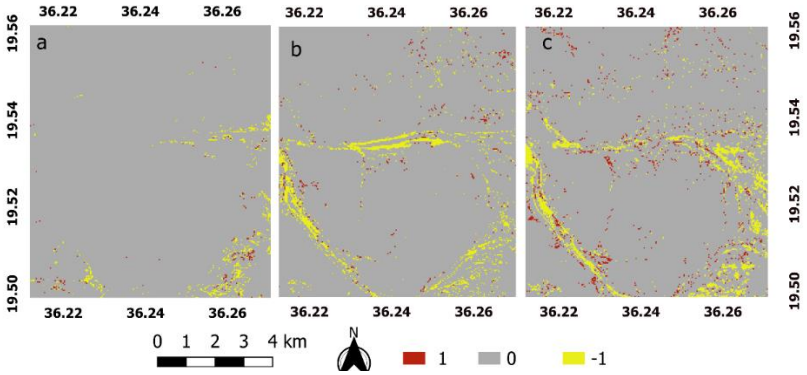
	SS	df	F	p	ω^2

Model	0.00877	14	0.3174	0.981	0.000
ML-models	7.97e-4	2	0.2020	0.819	0.000
Training data	0.00681	4	0.8632	0.508	0.000
ML-models × Training data	0.00116	8	0.0733	1.000	0.000
Residuals	0.02960	15			
Total	0.03837	29			

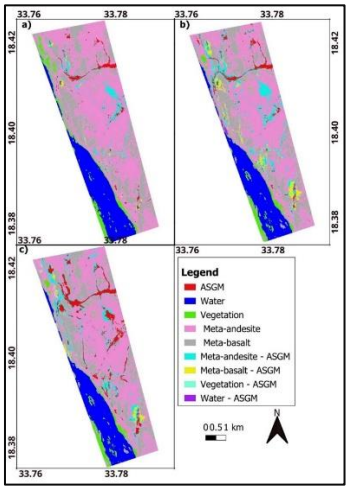
(Appendix 4-5). Accuracy measures and confusion matrices for Landsat data multiclass classification 2003, 2013 and 2023 (1: Marble, 2: meta-volcanic, 3: superficial deposits, 4: ASM, UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy)

Multiclass RF			Reference							
			1	2	3	4	total	UA	PA	F1-Score
2003	Predicted	1	226	19	9	8	262	0.842	0.809	0.823
		2	9	202	14	10	235	0.763	0.806	0.784
		3	14	23	273	38	348	0.768	0.737	0.752
		4	3	5	28	58	94	0.516	0.611	0.610
		total	252	249	324	114	939	OA: 0.804		
2013	Predicted		1	2	3	4	total	UA	PA	F1-Score
		1	275	0	0	0	275	0.905	0.889	0.905
		2	0	243	0	0	243	0.940	0.887	0.861
		3	10	0	121	50	181	0.940	0.811	0.827
		4	0	0	6	89	95	0.745	0.705	0.706
		total	285	243	127	139	939	OA: 0.874		
2023	Predicted		1	2	3	4	total	UA	PA	F1-Score
		1	191	0	0	0	191	0.926	0.889	0.905

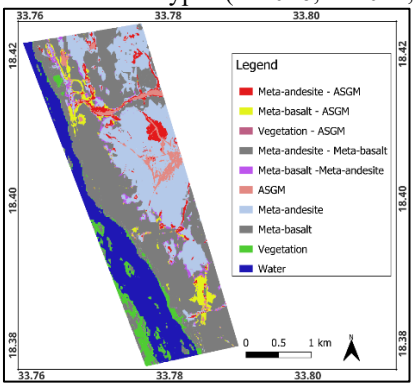
2	0	244	0	0	244	0.838	0.877	0.861
3	0	0	96	21	117	0.844	0.811	0.827
4	0	0	32	266	298	0.628	0.719	0.706
total	191	244	128	287	939	OA: 0.883		



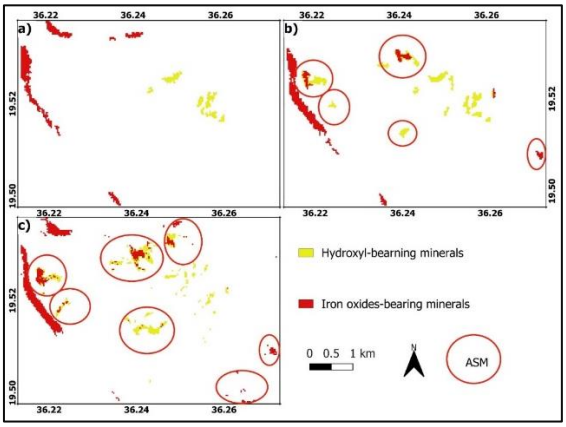
(Appendix 4-6). ASM regions by subtracting binary classifications from the multiclass (a: 2003, b: 2013, and c: 2023; -1: areas classified as ASM in both methods; 0: regions consistently classified as non-ASM; 1: areas detected as ASM in the multi-class approach but not in the binary classification).



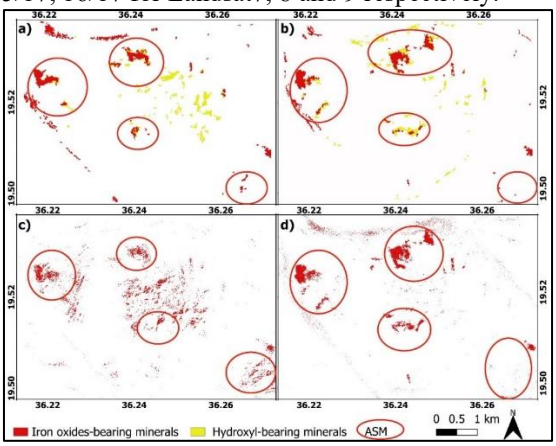
(Appendix 4-7) . Spatial distribution of the transition to ASM for geological units and land cover types (a: 2016, b: 2021, c: 2024)



(Appendix 4-8). Showing progressive increase in ASM areas through change detection analysis of multi-class RF classification between 2017 and 2024.



(Appendix 4-9). BR images showing the distribution of iron oxides and hydroxyl bearing minerals anomalous derived from Landsat7, 8, 9 and PRISMA. a), b) and c) b3/b1, b4/2 and b4/b2 of Landsat 7, 8, and 9. d), e) and f) b5/b7, b6/b7 for Landsat7, 8 and 9 respectively.

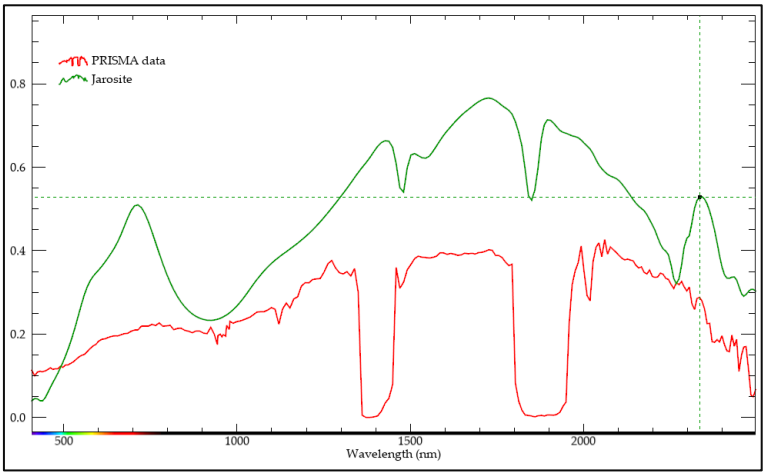


(Appendix 4-10). Band ratio b4/b2 and b11/b12 derived from Sentinel-2 a) 2015 and b) 2024 and the Planet data c) 2016 b3/b1, d) 2024 b6/b2 showing the distribution of iron oxides and hydroxyl bearing minerals anomalous.

(Appendix 4-11). The Eigenvector matrix values of bands, band ratio indices used d for DPCA using the Landsat 7, 8, 9 and Sentinel-2 to highlight various alteration types.

Eigenvectors(LC7 2003)	band 1	band 2	band 3	band 4
PC 1	0.211331	0.419694	0.614307	0.633895
PC 2	0.562149	0.531615	0.090456	-0.627049
PC 3	0.608110	-0.009256	-0.659204	0.442228
PC 4	-0.519162	0.735634	-0.424139	0.097060
Eigenvectors (LC7 2003)	band 1	band 4	band 5	band 7
PC 1	0.178919	0.541582	0.589881	0.571591
PC 2	0.256302	0.762961	-0.262461	-0.532273
PC 3	0.270050	0.077957	-0.742570	0.607936
PC 4	0.910698	-0.344242	0.178170	-0.142769
Eigenvectors (LC8 2015)	band 2	band 3	band 4	band 5
PC1	0.265856	0.408839	0.585879	0.647239
PC 2	0.439351	0.504923	0.229092	-0.706781
PC 3	0.763761	-0.044147	-0.593058	0.251002
PC 4	-0.391101	0.758917	-0.502539	0.136161
Eigenvectors (LC8 2015)	band 2	band 5	band 6	band 7
PC 1	0.219376	0.526073	0.584047	0.577937
PC 2	0.186913	0.770753	-0.192677	-0.577822
PC 3	0.301887	0.138671	-0.773046	0.540402
PC 4	0.908736	-0.331597	0.155448	-0.200193
Eigenvectors (LC9 2024)	band 2	band 3	band 4	band 5
PC 1	0.217980	0.414738	0.581675	0.664929
PC 2	0.567771	0.486114	0.173639	-0.641232
PC 3	0.682095	-0.081881	-0.623625	0.373007
PC4	-0.406039	0.764841	-0.492540	0.086924
Eigenvectors (LC9 2024)	Band 2	Band 5	Band 6	Band7
PC1	0.185717	0.553884	0.595858	0.551067
PC2	-0.032511	-0.754792	0.115303	0.644933
PC3	0.496536	0.216192	-0.734279	0.409325
PC 4	0.847292	-0.277061	0.304126	-0.335917
Eigenvectors(S2 2015)	Band 2	Band 4	Band 8a	Band 11
PC 1	-0.228349	-0.539160	-0.584102	-0.562128
PC 2	-0.258759	-0.528965	-0.170828	0.789973
PC 3	0.444797	0.391730	-0.768247	0.241867
PC4	0.826472	-0.525404	0.198593	-0.038151

Eigenvectors(S2 2015)	Band 2	Band 8a	Band 11	Band 12
PC 1	0.221738	0.588270	0.560746	0.538827
PC 2	0.160379	0.740625	-0.269299	-0.594332
PC 3	0.658782	-0.034935	-0.624946	0.417407
PC 4	0.700797	-0.322787	0.471683	-0.426857
Eigenvectors (S2 2025)	Band 1	Band 2	Band 3	Band 4
PC 1	-0.216821	-0.528411	-0.653422	-0.496800
PC 2	0.135094	-0.029649	-0.612549	0.778238
PC 3	0.500966	0.659805	-0.408400	-0.383276
PC 4	-0.826905	0.533441	-0.176163	0.025207
Eigenvectors (S2 2025)	Band 2	Band8a	Band 11	Band 12
PC 1	0.225758	0.673900	0.540980	0.449703
PC 2	0.030693	-0.674540	0.250876	0.693624
PC 3	0.876599	-0.004203	-0.464637	0.125177
PC 4	-0.423870	0.301390	-0.654610	0.548620



(Appendix 4-12). Spectral signature of the gossan pixel (red) in the study area matching Jarosite spectral reflectance (green).