School of Mining Engineering

A DATA SCIENCE FRAMEWORK FOR MINERAL RESOURCE EXPLORATION AND ESTIMATION USING REMOTE SENSING AND MACHINE LEARNING

Muhammad Ahsan Mahboob

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering.

Johannesburg, 2023

DECLARATION

I declare that this report is my own, unaided work. I have read the University Policy on Plagiarism and hereby confirm that no plagiarism exists in this report. I also confirm that there is no copying nor is there any copyright infringement. I willingly submit to any investigation in this regard by the School of Mining Engineering and I undertake to abide by the decision of any such investigation.

M Ahsan mahboob

Signature of Candidate

14 – August – 2023

Date

“Education is an invaluable tool to make a difference in the world and improve the lives of those around us. Let’s find the solutions with the poor and marginalised, not for them.”

Muhammad Ahsan Mahboob

ABSTRACT

Exploring mineral resources and transforming them into ore reserves is imperative for sustainable economic growth, particularly in low income developing economy countries. Limited exploration budgets, inaccessible areas, and long data processing times necessitate the use of advanced multidisciplinary technologies for minerals exploration and resource estimation. The conventional methods used for mineral resources exploration require expertise, understanding and knowledge of the spatial statistics, resource modelling, geology, mining engineering and clean validated data to build accurate estimations.

In the past few years, data science has become increasingly important in the field of minerals exploration and estimation. This study is a step forward in this field of data science and its integration with minerals exploration and estimation. The research has been conducted to develop a state-of-the-art data science framework that can effectively use limited field data with remotely sensed satellite data for efficient mineral exploration and estimation, which was validated through case studies. Satellite remote sensing has emerged as a powerful modern technology for mineral resources exploration and estimation. This technology has been used to map and identify minerals, geological features, and lithology. Using digital image processing techniques (band ratios, spectral band combinations, spectral angle mapper and principal component analysis), the hydrothermal alteration of potential mineralization was mapped and analysed. Advanced machine learning and geostatistical models have been used to evaluate and predict the mineralization using field based geochemical samples, drillholes samples, and multispectral satellite remote sensing based hydrothermal alteration information. Several machine learning models were applied including the Convolutional Neural Networks (CNN), Random Forest (RF), Support Vector Machine (SVM), Support Vector Regression (SVR), Generalized Linear Model (GLM), and Decision Tree (DT). The geostatistical models used include the Inverse Distance Weighting (IDW) and Kriging with different semivariogram models. IDW was used to interpolate data points to make a prediction on mineralization, while Kriging used the spatial autocorrelation to make predictions.

In order to assess the performance of machine learning and geostatistical models, a variety of predictive accuracy metrics such as confusion matrix, a receiver operating characteristic (ROC) curve, and a success-rate curve were used. In addition, Mean Absolute Error, Mean Square Error, and root mean square prediction error were also used. The results obtained based on the 10 m spatial resolution show that Zn is best predicted with RF with significant R^2 values of 0.74 ($p < 0.01$) and 0.7 ($p < 0.01$) during training and testing. However, for Pb, the best prediction is made by SVR with significant R^2 values of 0.72 ($p < 0.01$) and 0.64 ($p < 0.01$) for training and testing, respectively. Overall, the performance of SVR and RF outperforms the other machine learning models with the highest testing R^2 values. The experimental results also showed that there is no single method that can be used independently to predict the spatial distribution of geochemical elements in streams. Instead, a combinatory approach of IDW and kriging is advised to generate more accurate predictions. For the case study of copper prediction, the results showed that the RF model exhibited the highest predictive accuracy, consistency and interpretability among the three ML models evaluated in this study. RF model also achieved the highest predictive efficiency in capturing known copper (Cu) deposits within a small prospective area. In comparison to the SVM and CNN models, the RF model outperformed them in terms of predictive accuracy and interpretability. The evaluation results have showed that the data science framework is able to deliver highly accurate results in minerals exploration and estimation. The results of the research were published through several peer reviewed journal and conference articles. The innovative aspect of the research is the use of machine learning models to both satellite remote sensing and field data, which allows for the identification of highly prospective mineral deposits. The framework developed in this study is cost-effective and time-saving and can be applied to inaccessible and/or new areas with limited ground-based knowledge to obtain reliable and up- to-date mineral information.

PUBLICATIONS AS PART OF THIS RESEARCH

The publications listed below have emanated from this research work so far:

Journals

1. Mahboob, M. A., Celik, T., & Genc, B. (2022). Review of machine learning-based Mineral Resource estimation. Journal of the Southern African Institute of Mining and Metallurgy, 122(11), 655-664.

2. Mahboob, M. A., Celik, T., & Genc, B. (2021). A Comparative Study on Machine Learning Algorithms for Geochemical Prediction Using Sentinel-2 Reflectance Spectroscopy. Journal of Mining and Environment (JME) Vol, 12(4), 987-1001.

3. Mahboob, M. A., Celik, T., & Genc, B. (2020). Predictive modeling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments. Arabian Journal of Geosciences, 13, 1-21.

4. Mahboob, M. A., Genc, B., Celik, T., Ali, S., & Atif, I. (2019). Mapping hydrothermal minerals using remotely sensed reflectance spectroscopy data from Landsat. Journal of the Southern African Institute of Mining and Metallurgy, 119(3), 279-289.

5. Mahboob, M. A., Celik, T., & Genc, B. (2023). Predictive Modelling of Mineral Prospectivity Using Satellite Remote Sensing and Machine Learning Methods, A Data Science Approach. Mining, (under review).

Conference

1. Mahboob, M. A., Genc, B., & Atif, I. (2022, August). Harnessing the Power of Machine Learning for Mapping of Iron Oxide Mineralization using Sentinel Satellite Spectral Imagery. In 2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 285-289). IEEE.

2. Mahboob, M. A., & Genc, B. (2019, July). Evaluation of ISODATA Clustering Algorithm for Surface Gold Mining Using Satellite Data. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.

3. Mahboob, M. A., Genc, B., & Atif, I. (2019, April). Automatic extraction of geological lineaments in association with mineralization using remote sensing and geographical information system. In 2019 Proceedings of the 26th International Mining Congress and Exhibition of Turkey, Antalya, (pp. 285-289).

DEDICATION

To my loving parents for all their prayers and support, to my wonderful wife, Dr Iqra Atif, for her faith in me, to my brilliant son, Muhammad Yousuf Ahsan, for graciously sacrificing the time they deserved from me so much, and to my siblings for their encouragements and support.

ACKNOWLEDGMENTS

Alhamdulillah, glory be to Allah, the Almighty, the Most Gracious, and the Most Compassionate, I would like to begin by thanking Allah for His Blessing on the completion of this dissertation. Then, our Prophet Muhammad (ﷺ), who has guided us to a better life today, is also praised.

I would like to thank my parents for their prayers, unending love, and support throughout my entire life. Thank you very much for the motivation I required to nurture a dream, pursue, and believe in myself, and achieve my life's goals.

I would also like to express my sincere gratitude to the following extraordinary individuals for their guidance, support, encouragement, and contributions to the success of my research work:

- My supervisors, Professor Bekir Genc (School of Mining Engineering) and Professor Turgay Celik (School of Electrical and Information Engineering), for their invaluable and vital academic assistance and guidance during the research and prompt review of the research papers and thesis, and especially for their confidence in me.
- Prof Sarfraz Ali (Former Dean and Director; National University of Sciences and Technology Baluchistan Campus, Quetta Pakistan), for providing all the necessary resources and motivation to complete my PhD.
- Prof Frederick Cawood (Emeritus & Visiting Professor, School of Mining Engineering, University of the Witwatersrand), for providing moral support and encouragement to complete my PhD.
- Prof Cuthbert Musingwini (Head of School of Mining Engineering, University of the Witwatersrand), for providing School’s support in every possible way to complete my PhD.
- The current and the former Staff of School of Mining Engineering, Wits Mining Institute and Sibanye-Stillwater Digital Mining (DigiMine) Laboratory research group for the encouragement to ensure I complete this thesis.
- To all my teachers, from my earliest days in grade school to the mentors during my PhD journey. Each of you has been instrumental in shaping my academic path and assisting me in becoming the person I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviation	Description
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
TOA	Top-Of-Atmosphere
As	Arsenic
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
ATCOR	Atmospheric/Topographic CORrection
Au	Gold
Ba	Barium
BGS	British Geological Survey
BP	Back Propagation
Ce	Cerium
CGS	Council for Geosciences
CMAES	Covariance Matrix Adaptation Evolution Strategy
CNN	Convolutional Neural Network
CPU	Central Processing Units
CT	Convergence Threshold
Cu	Copper
CV	Coefficient of Variation
D8	Eight Flow Direction
DEM	Digital Elevation Model
DINF	D-Infinity
DN	Digital Numbers
DNN	Deep Neural Network
DT	Decision Tree
EM	Electromagnetic
ESRI	Environmental Systems Research Institute
FATA	Federally Administrated Tribal Areas
FCC	False Colour Composites
Fe	Iron
FLAASH	Fast Line-of-sight Atmospheric Analysis of Hypercubes
FN	False Negative

FP	False Positive
FPR	False Positive Rate
GAI	Geochemical Accumulation Index
GDEM	Global Digital Elevation Model
GIS	Geographic Information Systems
GLM	Generalized Linear Model
GPU	Graphical Processing Units
HT	Hough Transform
IDW	Inverse Distance Weighting
I_G	Gini index
IG	Information Gain
IK	Indicator Kriging
ISODATA	Iterative Self-Organizing Data Analysis Technique
JNB	Jenks Natural Breaks
LCDM	Landsat Data Continuity Mission
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LINE	Automatic Lineament Extraction Model
LLRBF	Local Linear Radial Basis Function
MAE	Mean Absolute Error
MFD	Multi-Flow Direction
MI	Maximum Iterations
MLA	Machine Learning Algorithms
ML	Machine Learning
Mn	Manganese
MODTRAN	MODerate resolution atmospheric TRANsmission
MPM	Mineral Prospectivity Mapping or Modelling
MR	Mineral Resources
MRE	Mineral Resources Estimation
MREE	Mineral Resources Exploration and Estimation
MSE	Mean Square Error
MSI	Multispectral Instrument
NASA	National Aeronautics and Space Administration
NBC	Naïve Bayes Classifier
Ni	Nickle

NIR	Near-Infrared
NN	Nearest Neighbour
OK	Ordinary Kriging
OLI	Operational Land Imager
Pb	Lead
PCA	Principal Component Analysis
ppb	Parts-Per-Billion
ppm	Parts-Per-Million
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red-Green-Blue
RK	Regression Kriging
RMSE	Root Mean Square Error
RMSPE	Root Means Square Prediction Error
ROC	Receiver Operating Characteristic
S6	Satellite Signal in the Solar Spectrum
SAM	Spectral Angle Mapper
SAS	Sub-Arctic Summer
Sb	Antimony
SBC	Spectral Band Combination
SBR	Spectral Band Ratio
SDE	Standard Deviational Ellipse
SE	Standard Error
SK	Simple Kriging
Sn	Tin
SPABC	Simultaneous Perturbation Artificial Bee Colony
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SVR	Support Vector Regression
SWIR	Shortwave-Infrared
Ti	Titanium
TIN	Triangular Irregular Network

TIR	Thermal-Infrared
TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
TN	True Negative
TOA	Top of Atmosphere
TP	True Positive
TPR	True Positive Rate
UK	Universal Kriging
V	Vanadium
WNN	Wavelet Neural Network
Zn	Zinc
Zr	Zirconium

1. INTRODUCTION

1.1. Research Background and Context

Mineral exploration is an essential activity that has the potential to bring economic growth to low income developing economy countries. It provides the ability to identify and extract valuable minerals or energy sources, which can then be used to improve the quality of life in these countries. Unfortunately, mineral exploration in underdeveloped countries presents a range of challenges and limitations that can hamper progress (Gocht et al., 2012). One of the main challenges faced by underdeveloped countries is the lack of resources to fund mineral exploration. As shown in Figure 1.1, the contribution of low income developing economy countries to the world budget for mineral exploration is even less than 0.01%, which is significantly lower than that of other countries.

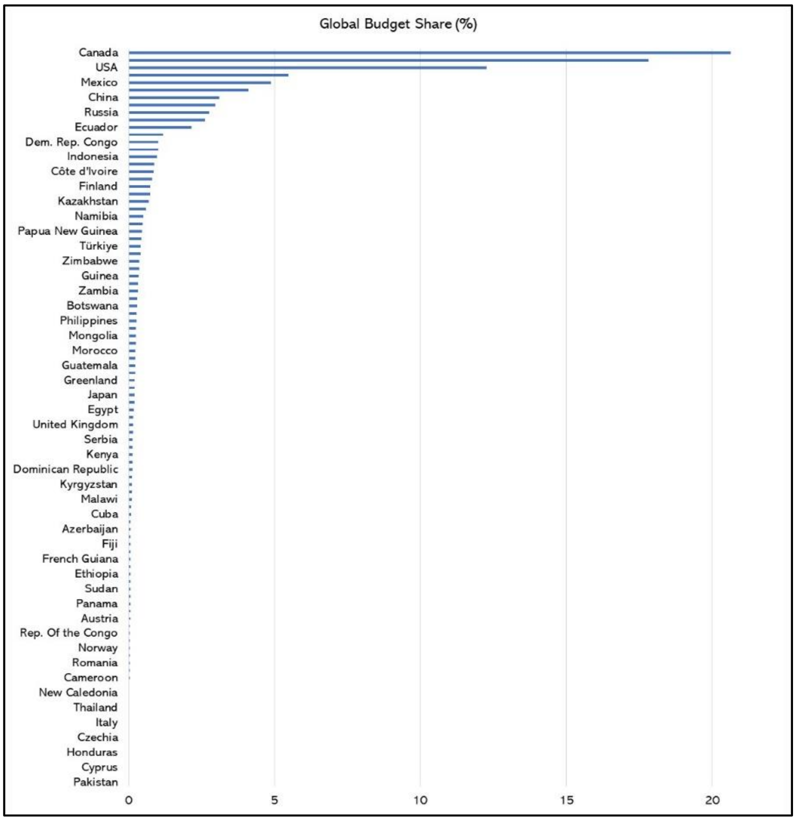


Figure 1. 1 The share of countries in global exploration budget as extracted from S&P Global Market Intelligence report (Global, 2022)

This means that there is often a significant gap between the resources required for exploration and the resources available to invest. This can lead to a lack of investment in exploration, which can hinder the potential for economic growth. Additionally, in some countries, the lack of infrastructure can make it difficult to conduct mineral exploration activities. Many of these countries lack the roads, railways, pipelines, and other transport systems necessary to facilitate the transportation of goods and people. Without these systems in place, it is difficult for miners to access the sites they wish to explore. Another challenge is the lack of scientific and technological expertise to conduct mineral exploration (Litvinenko, 2020). This can be an issue, as mineral exploration requires specialised skills that are not always available in these countries. Finally, the political and social dynamics of many underdeveloped countries often present significant challenges to mineral exploration. These countries typically have weak or non-existent legal systems, making it difficult for miners to secure the rights to explore and exploit mineral resources. Finally, the lack of a regulatory framework to govern exploration and exploitation activities in these countries is also a major concern. Without a clear set of regulations in place, it becomes difficult to control how resources are explored and exploited, and how the benefits of such activities are distributed. Due to all of these challenges and issues, there is an immediate need for a robust, reliable, and efficient method of mining exploration and exploitation not only for low-income developing economies, but also for developed economies.

Satellite remote sensing has emerged as a powerful modern technology for mineral resources exploration in underdeveloped countries. With the help of satellite imagery, mineral resources exploration can be conducted in a much more efficient and cost-effective manner (Shirmard et al., 2022). Remote sensing from satellites has numerous advantages over traditional exploration methods. Satellites can cover a larger area in a shorter period of time, and are able to capture images in a variety of wavelengths, which can be used to identify different mineral deposits (Bhattacharya et al., 2012; Mahboob et al., 2019; Manuel et al., 2017). Furthermore, satellite images also provide detailed information about resources in remote and hard-to-reach areas. Additionally, satellite remote sensing of mineral resources has the potential to provide detailed information on the location and characteristics of mineral deposits, enabling decisions to be made about where to focus exploration efforts. It can also help to identify areas where mineral deposits have been previously overlooked, helping to unlock valuable resources that could otherwise remain untapped. The majority of commercial and open-source remote sensing satellites are equipped with sensors for measuring the spectral reflectance and/or emission spectra between 400 and 12,000 nanometres (0.4 and 12 micrometres) of the electromagnetic (EM) spectrum. This spectrum range typically consists of the visible, near-infrared, shortwave infrared, and thermal infrared regions of the electromagnetic spectrum and is particularly suitable for confident mineral exploration (Fu et al., 2023). Several researchers have applied the satellite remote sensing data for minerals exploration and associated fields for example Salawu et al. (2023) applied remote sensing with geophysical datasets for gold exploration; Frutuoso et al. (2021) also applied remote sensing for gold exploration using open source Landsat-8 satellite datasets. Atwizukye (2022); Fu et al. (2023); Pour and Hashim (2012); Zhao et al. (2021) applied satellite remote sensing data with several digital image processing techniques for efficient copper mineral exploration and mapping. Ghorbani et al. (2019) utilized remote sensing for regional scale prospecting of non-sulphide zinc deposits. Cardoso-Fernandes et al. (2019); Santos et al. (2019) applied remotely sensed satellite data for lithium exploration.

There are several techniques that can be applied on remote sensing data to convert it into useful mineral information. Generally, these techniques can be categorized into two groups 1) knowledge-driven models and 2) data-driven models (Todorovski & Džeroski, 2006). The knowledge driven models require prior site information, i.e., experts' knowledge, and only incorporate the multispectral data of the remotely sensed satellite images. These techniques include Spectral Angle Mapper (Kayet et al., 2018; Lotfi et al., 2017; Qiu et al., 2015), Principal Component Analysis (PCA) (Koga et al., 2018; Sadowski et al., 2015), Minimum Noise Fraction (MNF) (Abubakar et al., 2018; Ayoobi & Tangestani, 2017), Band Ratio (BR) (Ge et al., 2018; Mahboob et al., 2019; Ousmanou et al., 2023), Constrained Energy Minimization (CEM) (Ousmanou et al., 2023; Pour et al., 2018a) and ISODATA technique (Li et al., 2023; Mahboob & Genc, 2019). Due to their lack of self-modification during the classification of image pixels, the results of these techniques sometimes lead to misclassification of the minerals present in the area both in terms of over and under estimation. On the other hand, the data-driven models estimate the mineral deposits based on spatial associations between evidential features and locations of known deposits. Usually they are Machine Learning Algorithms (MLA) such us Artificial Neural Networks (ANNs) (Franco-Sepulveda et al., 2019; Rigol-Sanchez et al., 2003), Decision Trees (DTs) (Cao et al., 2023; Zou et al., 2021), Regression Trees (RTs) (Kaplan et al., 2021; Mahboob et al., 2022), Random Forest (RF) (Mahboob et al., 2022; Mohamed Taha et al., 2022), Support Vector Machines (SVMs) (Tao et al., 2022) and Convolutional Neural Network (CNN, or ConvNet) (Li et al., 2021; Yang et al., 2021) are powerful data driven methods that incorporate not only the satellite imagery spectral information but also other field-based measurements (Figure 1, as representing general framework of MLAs). The results of data-driven models are quite accurate and reliable because they are trained at each data input stage.

1.2. Problem Statement

Mineral exploration can be considered as a multistage search process in which only the last stage, drilling, is usually conclusive. At each stage, an effort is made to reduce the area to which the next, typically more costly, stage of search is applied. Most of the low income developing economy countries are rich in mineral resources, offering a great potential for economic development and prosperity. But due to poor economic situation, lack of political will, limited data availability for most of the region and the areas with data has less and poor quality, the mineral reserve estimations are not up to the mark. The other big reason is the lack of expertise in interpretation of the available datasets. This research aims to explore mineral resources by integrating limited field data with remotely collected satellite data and employ advanced machine learning algorithms in order to obtain a more accurate estimation of available mineral resources in specific mineralized areas.

1.3. Research Objectives

The research aims to improve mineral prospectivity mapping and modelling by combining an integrated data science approach with machine learning models based on limited field data and remotely sensed satellite data. The specific objectives of the research are given as following:

- Geospatial mineral resource exploration using satellite remote sensing data and advance machine learning algorithms (MLAs);
- Comparative analysis of MLAs along with geostatistical modelling and comparison of different estimation techniques; and
- Comparative analysis of MLAs for potential mineral prospectivity mapping and modelling.

1.4. Contribution to Knowledge

This research would contribute to enhancing the strategy of potential mineral exploration particularly for low income developing economy countries. The development of a data science framework for mineral resource exploration and estimation using remote sensing and machine learning can have a significant impact on the efficiency and cost-effectiveness of mineral resource exploration and estimation. Government and private organizations can more accurately map areas of interest, identify mineral resources, make more informed decisions regarding their exploitation, and improve data-driven strategies for mineral resources exploration by utilizing the work completed in this research.

1.5. Structure of the Research Report

This thesis consists of ten chapters. Chapter 1 is the introduction and it covers the background of the research topic, the problem statement, the justification for the research, and the research objectives. The research methodologies followed by this study and the sources of the information are introduced. The research validation criteria, as well as the contribution to knowledge, are discussed.

Chapter 2 reviews the applications of machine learning-based mineral resource estimation techniques. The systematic literature review was performed on published articles related to applications of Machine Learning (ML) methods in the field of Mineral Resource Estimation (MRE) within a specific period of time. . The main purpose was to summarize the published work regarding the types of ML methods used in MRE, the comparisons between ML and non-ML methods, the performance evaluation of ML and non-ML methods, and the factors mainly considered in the application of ML methods. The outcome of this chapter is published as a peer reviewed journal article *Mahboob, M. A., Celik, T., & Genc, B.*

(2022). Review of machine learning-based Mineral Resource estimation. Journal of the Southern African Institute of Mining and Metallurgy, 122(11), 655-664.

Chapter 3 is a comparative study on machine learning algorithms for geochemical prediction using sentinel-2 reflectance spectroscopy. In this chapter several Machine Learning models were constructed and compared for the prediction of Lead (Pb) and Zinc (Zn) concentrations associated with the potential mineralization and mining activities at Central Wales (in the United Kingdom) using the satellite-based remotely- sensed spectral reflectance data. The outcome of this chapter is published as a peer reviewed journal article *Mahboob, M. A., Celik, T., & Genc, B. (2021). A Comparative Study on Machine Learning Algorithms for Geochemical Prediction Using Sentinel-2 Reflectance Spectroscopy. Journal of Mining and Environment (JME) Vol, 12(4), 987-1001*.

Chapter 4 is on the predictive modelling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments. In this chapter the geostatistical (IDW and Kriging) methods, which are widely used interpolation methods, have been selected and applied to test their efficiency in geochemical characterization and prediction along unsampled streams. To further aid the identification of mineral enriched streams and potential mineral prospects, a Geochemical Accumulation Index (GAI) has been developed, thereby providing a novel methodology. The outcome of this chapter is published as a peer reviewed journal article *Mahboob, M. A., Celik, T., & Genc, B. (2020). Predictive modeling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments. Arabian Journal of Geosciences, 13, 1-21*.

Chapter 5 is on the automatic extraction of geological lineaments in association with mineralization using remote sensing and geographical information system. In this chapter the advanced digital image processing algorithms were applied on the open-source remote sensing Landsat – 8 satellite datasets to map the lineaments at Waziristan region situated in North-West of Pakistan. The outcome of this chapter is published as a peer reviewed conference article *Mahboob, M. A., Genc, B., & Atif, I. (2019, April). Automatic extraction of geological lineaments in association with mineralization using remote sensing and geographical information* *system. In 2019 Proceedings of the 26th International Mining Congress and Exhibition of Turkey, Antalya, (pp. 285-289)*.

Chapter 6 is the mapping hydrothermal minerals using remotely sensed reflectance spectroscopy data from Landsat. In this chapter the identification of hydrothermally altered rocks and associated features with hydrothermal mineralization in Gauteng and Mpumalanga regions of South Africa was examined using open-source remote sensing Landsat – 8 satellite reflectance spectroscopy. The outcome of this chapter is published as a peer reviewed conference article *Mahboob, M. A., Genc, B., Celik, T., Ali, S., & Atif, I. (2019). Mapping hydrothermal minerals using remotely sensed reflectance spectroscopy data from Landsat. Journal of the Southern African Institute of Mining and Metallurgy, 119(3), 279-289*.

Chapter 7 is the evaluation of ISODATA clustering algorithm for surface gold mining using satellite data. In this chapter an unsupervised image classification algorithm ISODATA clustering was developed and evaluated on satellite remote sensing data for identification of gold mining areas. The outcome of this chapter is published as a peer reviewed conference article *Mahboob, M. A., & Genc, B. (2019, July). Evaluation of ISODATA Clustering Algorithm for Surface Gold Mining Using Satellite Data. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE*.

Chapter 8 is the harnessing the power of machine learning for mapping of iron oxide mineralization using sentinel satellite spectral imagery. In this chapter, three computationally intensive digital image processing techniques were evaluated for efficient mapping of iron oxide deposits in the Northern Cape region of South Africa using multispectral and multi-spatial Sentinel-2 satellite data. The outcome of this chapter is published as a peer reviewed conference article *Mahboob, M. A., Genc, B., & Atif, I. (2022, August). Harnessing the Power of Machine Learning for Mapping of Iron Oxide Mineralization using Sentinel Satellite Spectral Imagery. In 2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 285-289). IEEE*.

Chapter 9 is about the predictive modelling of mineral prospectivity using satellite remote sensing and machine learning algorithms. In this chapter a data science approach was applied to create nine predictor maps, incorporating limited field data and satellite remote sensing information. A confusion matrix, statistical measures, and a Receiver Operating Characteristic (ROC) curve were used to evaluate the prediction models efficacy on both the training and test datasets. The outcome of this chapter is submitted as a peer reviewed journal publication and is currently under review..

Chapter 10 is the highlights of the research findings and drawn conclusions. In addition, the chapter contains considerations for future research work.

2. REVIEW OF MACHINE LEARNING-BASED MINERAL RESOURCE ESTIMATION

This chapter is based on the following publication

Mahboob, M. A., Celik, T., & Genc, B. (2022). Review of machine learning-based Mineral Resource estimation. *Journal of the Southern African Institute of Mining and Metallurgy*, *122*(11), 655-664.

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

2.1. Introduction

Mineral Resources estimation (MRE) is one of the most important and critical stages in the mining value chain. The whole mining project depends on the reliable estimation of the grade of the mineralization. The spatial distribution of Mineral Resources (MR) depends on several known and unknown factors that cannot be incorporated in the traditional/conventional geostatistical models (Hosseini et al., 2017; Rossi & Deutsch, 2013). The basic assumption made by most of the mineral grade estimation models is that a spatial relationship exists between the grades at any two locations and that this relationship is a function of the distance between the two locations. Since the 1950s, many Mineral Resource Estimation (MRE) models have been proposed based on statistical methods, and later by incorporating the spatial dimensions in the estimations, which improved the results significantly. However, the spatial estimation techniques are based on several assumptions and predict the MR values with some uncertainty levels. Many MRE models based on spatial statistics only incorporate the three-dimensional location of the measured value (X, Y, Z) along with the grade and thickness information. However, several other parameters like topographical variations, directions of geological structures, type of geology etc. are also very important for reliable MRE. These parameters are often neglected or not incorporated in the conventional statistical and spatial statistical methods. In the last 10 years, machine learning (ML)-based methods have become more popular in Resource estimation research. Several researchers (Chatterjee, Bandopadhyay, et al., 2010b; Biswajit Samanta et al., 2006; Tahmasebi & Hezarkhani, 2010b; Zhang et al., 2017a) have reported that ML-based methods have emerged as prediction models and as major alternatives to geostatistics for MRE. Despite the huge number of research studies that used ML-based methods, inconsistent results have been reported regarding the accuracy of the methods, and comparison between ML and non-ML methods, and comparisons among several different ML-based methods. For example, in a comparison of ML and non-ML methods for MRE, Dutta et al. (2010) concluded that the ML methods produced more accurate results; however, Biswajit Samanta et al. (2005) showed that non-ML methods outperform the ML-based methods in producing reliable MREs. In a comparison among the ML-based methods, Tahmasebi and Hezarkhani (2010b), showed that artificial neural networks perform better than regression models. Chatterjee and Bandopadhyay (2011a) however, reported opposing results. Due to the discrepancies in the research studies which applied ML methods for MRE, practitioners in the fields of mining and geosciences may be hesitant to use ML models more practically and confidently. As opposed to other fields of study where the ML methods have been well tested and applied successfully, the applications of these methods in MRE face several challenges, like limited training data-sets, the uncertainty in existing data-sets and geological conditions, as well as the human factor. Although there is an increasing trend in academic research towards the applications of ML methods, most of the experts choose non-ML methods for MRE. To facilitate the applications of ML methods in the mining industry, it is very important to systematically summarize the empirical evidence on ML methods in current research and practice. The applications of ML methods are novel in the field of MRE, hence very little literature is available on the subject. There is no existing systematic literature review on the applications of ML methods in the field of mining related to MRE. In this paper, literature review was performed on articles published from January 1 1990 to June 30 2019, related to ML methods applications in the field of MRE. The main purpose is to summarize the published works regarding the types of ML methods used in MRE, the comparisons between ML and non-ML methods, the performance evaluation of ML and non-ML methods, and the factors mainly considered in the application of ML methods.

2.2. Methodology

The systematic literature review methodology proposed by Kitchenham and Charters (2007) was used to conduct and report the review. The main steps include the definition of research questions, design of search strategy, selection criteria for studies, quality assessment, extraction of relevant data, and analysis, as given in Figure 2.1.

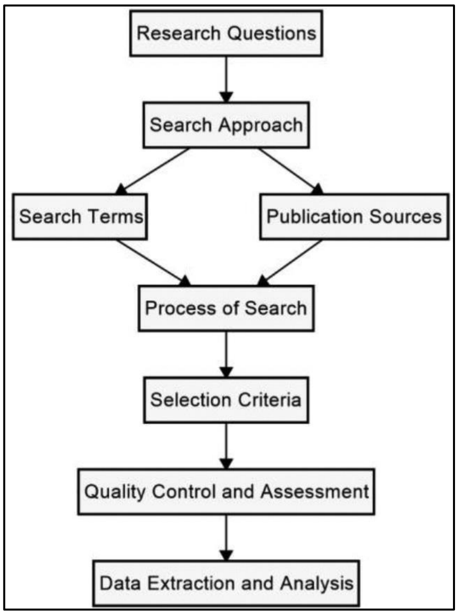


Figure 2. 1 The methodology used for the systematic literature review (conceptualised from Wen et al. (2012))

2.2.1. Research questions

Four research questions were defined based on the objective of summarizing the published work regarding the types of ML models used in MRE.

1. Which ML methods/models have been used for MRE?

The aim was to identify the ML methods/models that have been used in MRE to provide MRE researchers and practitioners with a range of possible methodologies to consider.

2. Does any publication on the comparison of ML against non-ML methods exist?

This question is concerned with the comparison of ML with non-ML methods in terms of accuracy, if performed in the studies.

3. Do ML methods outperform non-ML methods?

The aim is to compare the accuracy of the ML methods against non-ML methods.

4. Are there any ML methods that distinctly outperform other ML methods?

The comparison of different ML methods in order to identify these ML methods which perform better than others.

2.2.2. Search approach

The search strategy was based on the search terms, sources of publications, and process of search as explained below.

Search terms

The following steps were applied in order to search the terms for MRE (Wen et al., 2012):

(a) Selection of major keywords based on the research questions
(b) Identify possible different spellings and synonyms for major keywords
(c) Check the major keywords given in the relevant books and papers
(d) Usage of Boolean operator (OR) to combine the different spellings and synonyms
(e) Usage of AND operator to link the major keywords.

The following are the main keywords identified from the published work on machine learning models and techniques used for MRE.

Mineral AND (grade OR ore OR reserve OR resources) AND (estimation OR prediction) AND (machine learning OR artificial intelligence OR mining OR data mining) AND (geochemical OR exploration OR boreholes) AND (neural networks OR support vector machine OR support vector machine OR regression tree OR random forest OR Kriging OR nearest neighbour) AND (modelling OR spatial OR analysis).

Although significant research has been done on the application of ML techniques in the oil and gas or petroleum industry, that was not considered in this literature review, being outside the scope of the paper.

Publication sources

Five reliable and widely searched electronic databases (IEEE Xplore, ScienceDirect, Springer Nature, Web of Science, and Google Scholar) were used to search the most relevant literature. All the other databases are largely covered by these five primary databases, and hence have also been used by many literature reviews studies in several fields. The main keywords developed previously were used to search for journal and conference papers in the five databases. Except for Google Scholar, the search was conducted on the title, abstract, and keywords. Google Scholar returned several millions of irrelevant records when terms were searched on full text, hence the search was limited to the titles of the publications. The searches were restricted to the time period from January 1 1990 to June 30 2019, as publications on the application of ML methods to MRE begin to appear in the early 1990s; for example, Wu and Zhou (1993a) used neural networks for reserve estimation in 1993.

Search process

A comprehensive search of the relevant keywords among all the databases is very important. Therefore, the search process was designed and divided into two phases to identify relevant published papers.

(a) Phase A: Search the five databases individually and list all the papers that resulted from the searches.

(b) Phase B: The references in each resulting paper were scanned to identify additional relevant papers, which were added to the original. The inclusion and exclusion criteria set for the papers were as follows:

Inclusion criteria

1. ML methods used to pre-process the data
2. Applied more than one ML method and/or combined with non- ML methods
3. Comparative studies that compare different ML methods and/ or with non-ML methods
4. Studies with both conference and journal publications; only journal papers were selected for inclusion
5. Studies that have multiple versions or duplications; only the most recent and complete study was included in the list.

Exclusion criteria

1. Qualitative studies without proper ML methods details and used with other than borehole or geochemical data.
2. Review papers.

2.3. Quality Control and Assessment

The quality control process assists with the selection of the most relevant research papers. A series of research questions were formulated in order to assess the rigorousness, reliability, and relevance of the papers. The questions are given in Table 2.1, based on the methodology developed by Wen et al. (2012). The three options were assigned to each question, i.e., yes, partially, and no. These options were scored as 1, 0.5, and 0 respectively. The quality was assessed by summing the scores for answers against each question. The studies with a minimum score of 3.0 (50% of a perfect score) were selected for data extraction and analysis to ensure the quality of this literature research.

Table 2. 1 Research Questions defined for the quality assessment of the published research studies

Code	Question

QA1	Are the aims of the research clearly defined?
QA2	Are the estimation methods well defined and deliberate?
QA3	Is the estimation accuracy measured and reported?
QA4	Is the proposed estimation method compared with other methods?
QA5	Are the findings of the study clearly stated and supported by the results?
QA6	Are the limitations of the study analyzed thoroughly?

By applying the selection criteria, 50 papers were identified. After scanning the references in these papers, eight additional relevant papers were found. Hence, a total of 58 relevant papers were initially considered. However, after applying the quality control criteria, only 31 papers were selected for data extraction. The quality assessment is discussed in detail in the following sections. The complete list of the 31 selected papers can be found in Appendix A.

2.4. Data Extraction and Analysis

The selected research studies were exploited in order to collect the data that can answer the specific research questions of this exercise. All the reliable and relevant research papers were divided into specific sections, as given in Table 2.2. These sections were analysed combinedly to provide more meaningful information and to enhance the understandings. The extracted data was both quantitative (number. of boreholes or data-sets) and qualitative (data type, publisher, publication type). Different visualization techniques such as bar charts, pie charts, etc. were also used to enhance the data extracted from the research studies.

The vote counting method was used in order to compare the accuracy and application of different ML models (Malhotra, 2015). The vote counting method counts the number of times a model i.e., Model A outperformed Model B, or vice versa. With this method, a general idea of whether an estimation ML model outperforms another model in the estimation of mineral grades emerged.

Table 2. 2 Information extracted from the research studies for the analysis

Sr. no	Extracted sections
1	Title of paper
2	Title of journal/conference
3	Publisher
4	Link
5	Year of publication
6	Type (journal/conference)
7	Data-set type (borehole/image)
10	Error assessment technique
11	Research question 1
12	Research question 2
13	Research question 3
14	Research question 4
15	Relevant to research
16	Publication domain

2.5. Results and Discussion

A total of 31 research studies that dealt with the application of ML methods in the field of MRE were identified. These studies were published between 1990 and 2019 in journals and conference proceedings. Of these, 30 (approx. 97%) papers were published in journals, only one (approx. 3%) was published in conference proceedings, and none were found as book chapters. The places of publication are given in Appendix B. The publishers were mainly Elsevier, Springer, the IEEE, and Taylor & Francis. All of the research studies were experimental and none of them were survey research. In terms of quality control, only studies with a minimum quality score of 50% were selected, hence all the studies were of a high-quality level.

The publication history is summarized in Figure 2.2 and it shows that the oldest paper found was published in the year 1993, followed by eight years wherein the selected publishers did not publish any paper until 2002. A significant increase in ML-based mineral grade estimation papers was found in 2010, with five papers published in that year. In both 2013 and 2017, four papers were published.

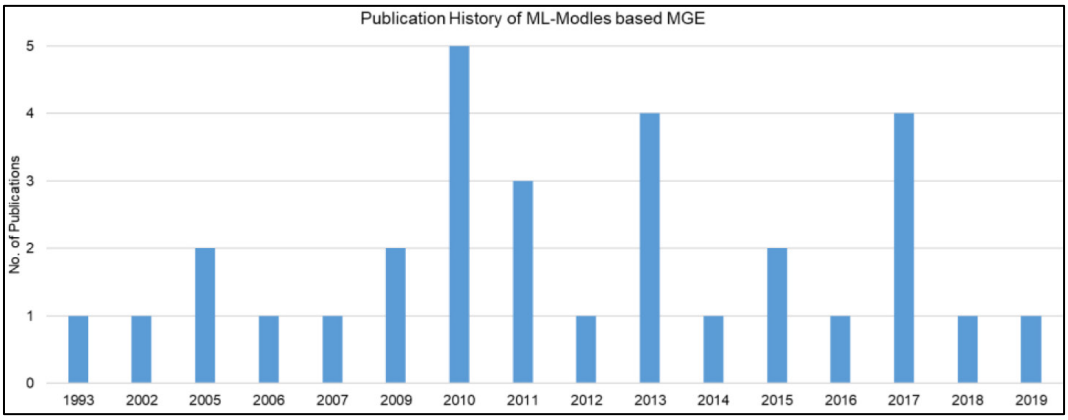


Figure 2. 2 Annual ML research publications from 1990 to 2019

However, only one paper was found to be published in 2018 and 2019. These statistics show that limited research has been conducted on the application, testing and validation of ML techniques in the field of mineral grade estimation.

Through this exercise, fourteen types of ML-based methods and techniques were identified that have been used for mineral resource exploration and estimation as listed below:

- Support vector machine (SVM)
- Support vector regression (SVR)
- Artificial neural networks (ANN)
- Adaptive neuro-fuzzy inference system (ANFIS)
    -
- Local linear radial basis function (LLRBF) neural network
- Simultaneous perturbation artificial bee colony algorithm (SPABC)
- Back propagation (BP)
- Covariance matrix adaptation evolution strategy (CMAES)
- Particle swarm optimization (PSO)
- Naïve Bayes classifier (NBC)
- Radial basis function (RBF)
- Wavelet neural network (WNN)
- Random forest (RF)

- Regression kriging (RK)

Among all these models and techniques, SVM, SVR, and ANN are the most commonly used and applied for MRE. Together they were found in 60% of the selected studies, as shown in Figure 2.3. Detailed information about which techniques were used in which study are given in Appendix C.

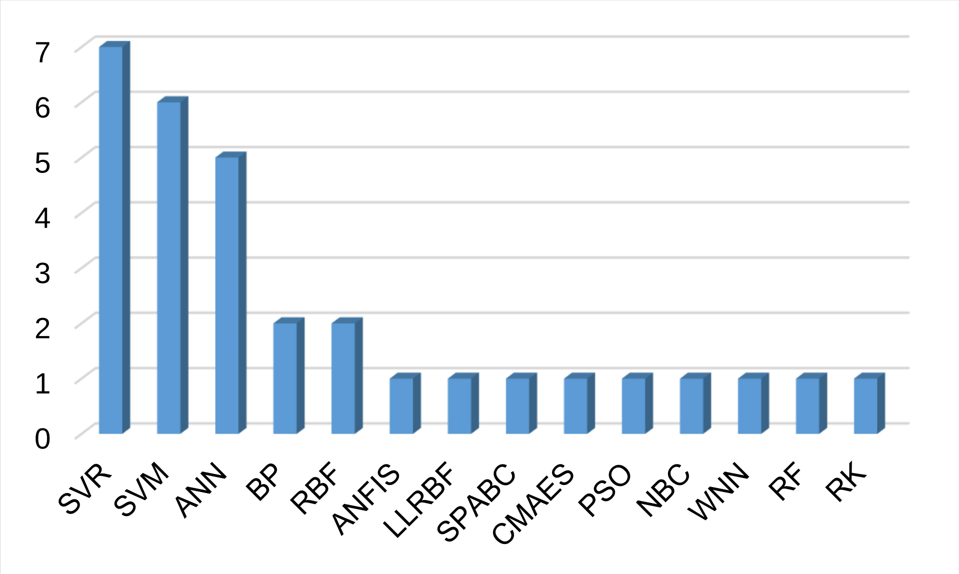


Figure 2. 3 Frequency of use of ML models by the selected research studies

On the other hand, the non-ML methods mostly used in MRE are kriging, ordinary kriging, and inverse distance weight (IDW). Only 41% of studies (13 papers) compared the results of ML with non-ML methods in terms of mineral grade estimation. All of these studies concluded that ML models outperformed the non-ML methods except S-04, which concluded that ordinary kriging performed better than artificial neural networks. The ML methods, which most frequently outperformed the other ML methods were SVM and SVR. A brief description of most of the common ML techniques, including SVM, SVR, and ANN, are given in following sections.

The support vector machine is a supervised empirical machine-learning algorithm, based on statistical learning theory (Vapnik, 1999). (SVM has recently been introduced in the field of mining and mineralization. The SVM is usually used for data classification and prediction; however, multi-class SVM can also be generated by combining multiple binary classifiers. It showed several unique advantages in a small data sample with nonlinear and multi-dimensional patterns recognition. The main objective of SVM is to locate a hyperplane that can separates data-points of one type from another. The best hyperplane is the one with the largest margin between two classes, and hence can separate the classes distinctively as shown in Figure 2.4.

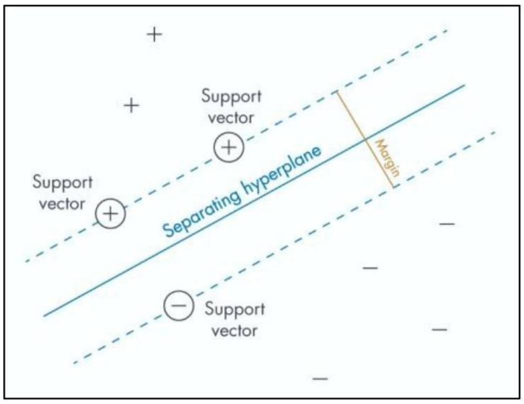


Figure 2. 4 Hyperplane separating the support vectors during the classification by SVM (MatLab, 2020)

The maximum distance between two parallel hyperplanes results in the minimum classification error. SVM has been extensively used in several field of engineering, science, and natural languages. In the mining industry, it has recently been used for mineral classification (Patel et al., 2017a), mineral prospectivity (Abedi et al., 2012), and automatic lithological classifications (Othman & Gloaguen, 2017; Yu et al., 2012). Support vector regression differs greatly from other regression models. Whereas the other linear regression models try to minimize the difference between the estimated and the true value, SVR tries to fit the best line within a threshold value. In MRE, SVR tries to categorize all the estimation lines in two forms, those that pass through the threshold boundary and those that don’t. The lines that do not pass the threshold boundary are not considered as the difference between the estimated grade value and the true grade value has exceeded the error threshold defined by ε (epsilon) as shown in Figure 2.5. On the other hand, the lines that pass are considered for a possible support vector to estimate the grade value at an unknown location.

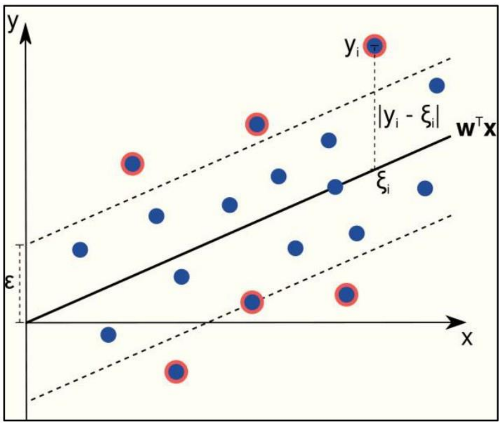


Figure 2. 5 Illustration of an SVR regression function separated by the ξ band for data-sets (Rosenbaum et al., 2013)

Artificial neural networks is another strong machine-learning approach. This biologically inspired computational technique has wide applications in several science and engineering fields. A unique quality of ANNs is that they are able to create empirical relations between independent and dependent variables and extract the hidden variability and complex knowledge from training data-sets. The relations between independent and dependent variables can be built without assumptions about any mathematical depiction of the phenomena. ANN models have several benefits over regression-based ML methods, including their ability to deal with noisy data. An ANN model has thousands of interconnected artificial neurons made up of inputs and outputs. The input nodes receive the actual mineral grade values at known locations based on the internal weighting system and the neural network tries to learn the hidden patterns and produces the output, as shown in Figure 2.6.

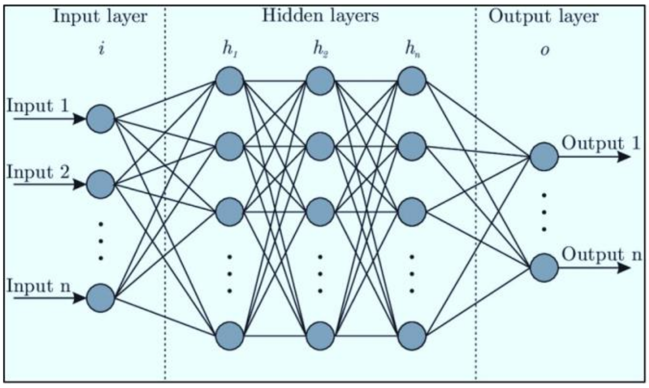


Figure 2. 6 General architecture of an ANN-Artificial Neural Network (Bre et al., 2018)

An ANN model compares the actual output with what it was meant to produce, i.e., the desired output. The difference between both is corrected using back-propagation so the ANN works regressively, going from the final output node to the input nodes in order to fine-tune the weight of its connections until the variation between the actual grade value and estimated grade value produces the lowest possible error.

The types of data used for MRE estimation, regardless of technique, can be divided into three general categories:

- Exploratory boreholes
- Images
- Stream sediments

Twenty-three studies used exploratory borehole data, six used images photographs, and only one study utilized stream sediment data for the mineral grade estimation. The study S-13 used 3,500 exploratory boreholes and concluded that ML-based SVR is the most accurate technique for mineral grade estimation compared to the non-ML based ordinary kriging. Similarly, research study S-08, in which the authors used stream sediments, also concluded that ANN, which is another ML method, outperformed kriging, a non-ML method, when used for mineral grade estimation. The studies in which images were used did not compare the results of ML models with non-ML models.

In terms of publication domains, the studies were divided into the following three main categories:

- Minerals/mining
- Computer science
- Geoscience

A total of seven studies were published in the field of minerals/mining, 12 in computer sciences, and 12 in the field of geosciences. This reveals that most of the publications fall into multidisciplinary applied domains. Another important aspect of this (Systematic Literature Review) SLR is to identify the most applicable statistical/geostatistical technique used for the error assessment of the predicted mineral grades. More than 73% of the research studies used mean square error (MSE) and root mean square error (RMSE) as error assessment techniques, followed by the standard error (SE), and mean absolute error (MAE) with a total of 16%. The remaining 11% of the error assessment techniques include mainly generalization error, percentage of accuracy, co-efficient of determination, and out-of-bag error.

This review has found that most of the recent and advance ML methods and techniques such as deep convolution neural network, hierarchical convolutional deep maxout network, and hidden trajectory (generative) models, are seldom used within the field of mineral grade estimation. Therefore, researchers/ practitioners are encouraged to apply these techniques and test their applicability for mineral grade estimation.

In addition, researchers/practitioners are also encouraged to explore the other non-tested ML methods for mineral grade estimation. In order to become acquainted with the unexplored ML methods and to apply them in a more efficient way, researchers/practitioners should keep a close watch on the related disciplines like machine learning, deep learning, data science, and artificial intelligence, as these disciplines may provide ideas for new ML methods and techniques (Wen et al., 2012).

Even though this investigation has found that the ML models are more accurate and perform better than the non-ML methods, the results of error assessment between ML and non-ML methods, and between different ML methods, are still inconclusive. Hence, it is strongly recommended that the scientific community develop a common framework for evaluating the performance of different ML techniques, as well as against non-ML methods. The results of the studies may vary because of different data-sets and/ or different experimental designs.

2.6. Conclusion

This systematic literature review examined ML-based MRE models in terms of the type of ML methods or techniques, the error estimation of applied ML methods, comparison between different ML and non-ML methods, as well as amongst ML methods. The extensive systematic literature review was based on research studies published in the period 1990-2019, with a total of 31 studies meeting the requirements of five research questions.

The key findings of the literature review are that the significant ML methods applied for the MRE are, in order of application, SVR, SVM and ANN. Few studies actually compared the results of ML methods with non-ML methods for MRE. Those studies concluded that ML methods outperformed the non-ML methods in general. SVM and SVR are the most applied and tested ML methods, which yield much better results than other ML methods. Very few papers have been published in the fields of mineral and mining, whereas most of them were published in the computer and geosciences fields.

This review provides recommendations for researchers for future work as well as guidelines for practitioners. More research should be conducted detailing studies on the application of ML methods and drawing of conclusions in terms of their applicability, validation, and accuracy. Researchers should also develop a framework in terms of data usage and accuracy assessment against non-ML methods used for mineral grade estimation.

From this review, it is very clear that the application of ML methods in the industry for MRE are limited, and hence more studies should be conducted and analysis done in order to find the possible barriers to ML method applications. It is strongly recommended that ML models be used in parallel with non-ML (conventional statistical) models in the early stages of mineral grade estimations. After the error assessment and quality check, the validated ML methods can then be used for estimation of Mineral Resources. Moreover, it is advisable to consult researchers/experts from the fields of machine learning and/or data science in order to check the strengths and weaknesses of the potential ML methods and interpretation of results before and after application accordingly.

3. A Comparative Study on Machine Learning Algorithms for Geochemical Prediction Using Sentinel-2 Reflectance Spectroscopy

This chapter is based on the following publication

Mahboob, M. A., Celik, T., & Genc, B. (2021). A Comparative Study on Machine Learning Algorithms for Geochemical Prediction Using Sentinel-2 Reflectance Spectroscopy. *Journal of Mining and Environment (JME) Vol*, *12*(4), 987-1001

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

3.1. Introduction

The identification and mapping of potential mineralization through the geochemical exploration of stream sediments has been successfully done over the past several decades (Lin et al., 2020; Yousefi et al., 2012). These geochemical observations at the surface of earth usually represent the diverse effects of the primary and secondary geological processes inside the earth (Kirkwood et al., 2016). The field-based sampling of stream sedimentation is not only time-consuming but also expensive; usually the samples are only collected from the downstream areas with less slope and easy accessibility. However, reflectance spectroscopy of satellite-based remote sensing data provides a unique edge over mapping of geochemically enriched areas using stream sedimentation not only at a larger scale but it is also time- and cost-effective (Choe et al., 2008; Cyples et al., 2020; Wang et al., 2020). In the research work conducted by Martinez et al. (2009), the MODIS satellite data was used in order to quantify the Amazon River sedimentation. It was concluded that sedimentation was assessed through satellite, and the field datasets showed a very good agreement with a mean difference lower than 1%. Another research work conducted by Abedi and Norouzi (2016) tested the ASTER and LANDSAT satellite datasets along with the geochemical and geological data for mineral exploration using the TOPSIS method. The research work concluded that the proposed methodology based on satellite and field data was satisfactory for mapping of the porphyry copper deposit. The research work conducted by Afzal et al. (2015) also used the ASTER satellite data for exploration and mapping of Cu mineralization in Iran using the stream sedimentation data. A good correlation was found between the satellite reflectance data and field-based sedimentation values, and concluded that the satellite multispectral data could be used for mineral exploration using stream estimations. However, one of the key components in using the satellite data is the spatial auto-correlation of satellite data and the field measurements (Mondini, 2017), a phenomenon usually used to map the features in a region based on the systematic spatial similarities between the satellite reflectance and laboratory measurements. Usually, the spatial interpolation models like inverse distance weighted and kriging are used in order to predict the presence or absence of a geochemical value of a variable at the non-surveyed regions using the surveyed datasets. However, these interpolation models usually incorporate the estimation errors such as the smoothing effects in the prediction of geochemical variables (Yousefi, 2017). Besides the smoothing effects, the complexity, non-linearity, and spatial variability in the geochemical datasets makes it challenging to predict the values at unknown locations based on the known location datasets.

The advancements in machine learning (ML) have made it a strong alternative choice to overcome the classical interpolation challenges and to develop the predictive models by analysing and learning from the data by incorporating the spatial autocorrelation. Several researchers (Ahmed et al., 2020; Dornan et al., 2020; Tehrany et al., 2019) have applied ML as a tool for modelling the geospatial phenomenon by building the models capable of identifying the patterns in geospatial data and predicting from these models. The applications of ML in geochemical modelling are very limited, and there are very few publications on this topic. However, the ML-based predictive models are potentially powerful for geochemical mapping and mineral explorations (Chen & Wu, 2017; Coimbra et al., 2014; Wang et al., 2019; Zuo & Xiong, 2018). This research work aims to construct and compare the ML models for prediction of the Pb and Zn concentrations associated with the potential mineralization and mining activities at Central Wales (in the United Kingdom) using the satellite-based remotely-sensed spectral reflectance data.

3.2. Materials and Methods

3.2.1. Geology setting

The studied area for this research work is in Central Wales of Great Britain (Figure 3.1). It has a prolonged history of gold mining, with the main centre of activity being the "Dolgellau Gold Belt", where the Clogau and Gwynfynydd mines are very active gold mines, particularly towards the South, in the Welsh Basin, where gold has been mined since the Roman times. Besides gold, several other secondary products such as lead, copper, zinc, iron, and nickel sulphides are present throughout the Wales area. The general geology of Central Wales is consisted of Ordovician and Silurian marine sedimentary rocks. The rocks settled at the bottom of the Welsh Basin are dominated by series of sandstone, siltstone, and mudstone. Many of these classifications are referred to as turbidites because they were settled from violent, sediment-laden undersea flows, which flooded off the shallower shelf areas onto the deep floor of the basin (Toghill, 2011). There has been extensive mining of Pb and Zn as Ball and TK (1976) has reported that the area is enriched with these two minerals, and they are most commonly found within the interbedded sequences of sandstone, siltstone, and mudstone.

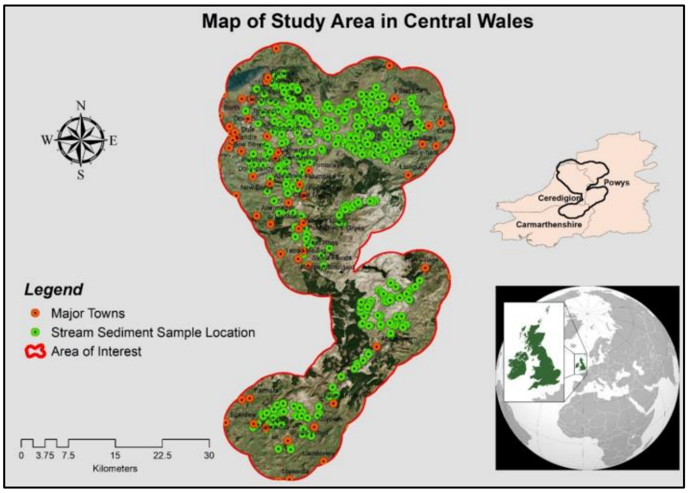


Figure 3. 1 Geographical map of the studied area highlighting the major towns and locations of the collected samples (Mahboob et al., 2020)

3.3. Machine Learning Models

Machine learning models have been applied, tested, and proved in several real-world applications. The main objective of this research work is to predict Pb and Zn in Central Wales using satellite spectral reflectance data through following machine learning models and their comparative evaluation in terms of predictions. Regardless of the machine learning model type applied, the data was split into the training and testing purposes at 70% and 30%, respectively (Chen et al., 2017).

3.3.1. Support vector regression (SVR)

SVR was initially introduced by Drucker et al. (1997), and has been successfully applied for hidden pattern identifications in the data due to its strong predictive capability, flexibility, and robustness (Lim et al., 2002). SVR is more suitable and applicable to the classification problems; however, the SVR concepts can be generalized to become applicable to the regression problems. Several researchers have applied SVR for urban land cover mapping (Okujeni et al., 2013), hotspot detection (Pozdnoukhov & Kanevski, 2007), predication of organic carbon content in soils (Tan et al., 2015), landslide modelling and mapping (Miao et al., 2018), and in mining applications (Ding et al., 2020; Nourali & Osanloo, 2019). In the SVR algorithm, all the data is plotted as points in an n-dimensional space (where n is the total of features in the data), and then the classification is performed on the n-dimensional space with the main objective to find the hyper-plane that can differentiate the two feature classes distinctively.

For the regression problems, rather than finding a hyperplane that can mainly distinguish the training points, SVR uses an ε-insensitive loss function for the computation of the hyperplane so that the predicted response values of the training points have at least an ε deviation from their actual response values. The hyperplane along with ε will define an ε-insensitive band (decision boundary) for the regression. For a set of training data $\{(x_1, y_1), \ldots (x_i, y_i)\}$, where x_i is the input data and y_i is the target output, following Equation 3.1 can be considered for general SVR.

$$y = f(x) = w, x + b = w^T x + b$$ *Equation 3. 1*

where w, x denotes the dot product of the input data x and the weight vector w. The function f can be calculated through an ε-insensitive band as flat as possible, which is usually mentioned as flatness in order to seek a small w. The approximation of f can be described as Equations 3.2 and 3.3.

$$min_w \frac{1}{2} ||w||^2$$ *Equation 3. 2*

subject to $\begin{cases} y_i - w^T x_i - b \leq \varepsilon \\ w^T x_i + b - y_i \leq \varepsilon \end{cases}$ *Equation 3. 3*

For the above equation, it can be seen that SVR is to perform a linear regression with an ε- insensitive loss function that can be linear or quadratic, as given in Equations 3.4 and 3.5, respectively:

for linear

$$L\left(y, f\ (x)\right) = \begin{cases} 0 & if\ |y - f(x)|\ \leq\ \varepsilon \\ |y - f(x)| -\ \varepsilon & otherwise \end{cases}$$ Equation 3. 4

for quadratic

$$L\left(y, f\ (x)\right) = \begin{cases} 0 & if\ |y - f(x)|\ \leq\ \varepsilon \\ (|y - f(x)| -\ \varepsilon)^2 & otherwise \end{cases}$$ Equation 3. 5

3.3.2. Generalized linear model (GLM)

GLM is used to focus on the arbitrary distributions (non-normal) response of the linear regression models. The main aim of a GLM is not to model the dependent factor as a linear combination of independent factors but to model a function of dependent factor as a linear combination of independent factors. Many research studies have applied GLM for predictive modelling, e.g., Youssef et al. (2016) have applied different machine learning models including GLM for landslides prediction, concluding that GLM could be used efficiently for landslide susceptibility mapping with a reasonable accuracy of 76.9%. Another research work conducted by Miller and Franklin (2002) has applied GLM along with other classification trees for an effective mapping of vegetation alliances. For a GLM model, the three main components are a family function f, link function k, and parameters required to train the model. The family function can be Gaussian, binomial, fractional-binomial, ordinal, quasi-binomial, multinomial, Poisson, gamma, Tweedie, and negative-binomial. In this research work, the Tweedie family function has been applied, which further consist of gamma, normal, Poisson, and their combinations, also because it is highly recommended, if the dataset has the positive continuous and exact zero responses. The variation in the Tweedie function is directly proportional to the *pth* power of the mean-variance var (*yi*) = δεp, where δ is the distribution factor and p is the power of variation. The Tweedie distribution is characterized by the power of variation p, while δ is an unknown constant. The values of p will be defined as per Equation 3.6:

$$
\begin{gathered}
p = 0: for\ Normal\ case \\
p = 1: for\ Poisson\ case \\
p\ \in\ (1,2): for\ Compound\ Poisson, non \\
- negetive\ with\ zeros\ case \\
p = 2: for\ \text{Gamma}\ case \\
p = 3: for\ \text{Inverse} - \text{Gaussian}\ case \\
p > 2: for\ positive\ reals\ case
\end{gathered}
\qquad \text{Equation 3. 6}
$$

The following maximum likelihood Equation 3.7 is used to fit the model:

$$
\sum_{i=1}^{N} \log\big(\alpha(y_i,\delta)\big) + \begin{cases} \dfrac{1}{\delta}\left(y_i\ \log(\varepsilon_i)\ - \dfrac{\varepsilon_i^{2-p}}{2-p}\right), & p = 1 \\ \dfrac{1}{\delta}\left(y_i\ \dfrac{\varepsilon_i^{1-p}}{1-p} - \log(\varepsilon_i)\right), & p = 2 \\ \dfrac{1}{\delta}\left(y_i\ \dfrac{\varepsilon_i^{1-p}}{1-p} - \dfrac{\varepsilon_i^{2-p}}{2-p}\right), & p \neq 1,\ p\ \neq 2 \end{cases} \qquad \text{Equation 3. 7}
$$

where the function $\alpha(y_i,\delta)$ is assessed using an unbounded sequence increase, and will not have a logical explanation. However, because δ is an unknown constant, $\sum_{i=1}^{N} \log\big(\alpha(y_i,\delta)\big)$ can be considered as a constant as well and can be ignored. Therefore, the final function to minimize with the penalty will be as per the following Equation 3.8:

$$\min_{\alpha,\alpha_0} \mu\left(\beta||\alpha||_1 + \frac{1}{2}(1-\beta)||\alpha||_2\right)\begin{cases} \left(y_i\ \log(\varepsilon_i)\ - \frac{\varepsilon_i^{2-p}}{2-p}\right), & p=1 \\ \left(y_i\ \frac{\varepsilon_i^{1-p}}{1-p} - \log(\varepsilon_i)\right), & p=2 \\ \left(y_i \frac{\varepsilon_i^{1-p}}{1-p} - \frac{\varepsilon_i^{2-p}}{2-p}\right), & p \neq 1,\ p \neq 2 \end{cases} \qquad \text{Equation 3. 8}$$

Usually, the following function (Equation 3.9) is always used in the GLM model for the Tweedie function:

$$g\ (\varepsilon) = \begin{cases} \varepsilon^q =\ \omega = X\alpha, q \neq 0 \\ \log(\varepsilon) =\ \omega = X\alpha, q = 0 \\ q = 1-p \end{cases} \qquad \text{Equation 3. 9}$$

The link power q can also be set to some other values as per the value of p. The resultant deviation will be as per the following Equation 3.10:

$$D = 2\ \times \begin{cases} \sum_{i=1}^{N} y_i\ \log\left(\frac{y_i}{\varepsilon_i}\right) - \frac{\left(y_i^{2-p} -\ \varepsilon_i^{2-p}\right)}{2-p}\ , & p=1 \\ \sum_{i=1}^{N} \frac{y_i}{1-p}\ \left(y_i^{1-p} -\ \varepsilon_i^{1-p}\right) - \log\left(\frac{y_i}{\varepsilon_i}\right)\ , & p=2 \\ \sum_{i=1}^{N} \frac{y_i\left(y_i^{1-p} -\ \varepsilon_i^{1-p}\right)}{1-p} - \frac{\left(y_i^{2-p} -\ \varepsilon_i^{2-p}\right)}{2-p}\ , & p \neq 1,\ p \neq 2 \end{cases} \qquad \text{Equation 3. 10}$$

3.3.3. Deep neural network (DNN)

Multi-layer Feedforward Artificial Neural Network, also known as DNN, is used in deep learning, and is trained with stochastic gradient descent using back-propagation. DNN comprises multiple levels of non-linear processes like neural nets with many hidden layers, as shown in Figure 3.2, and is also suitable for tabular datasets like the geochemical samples.

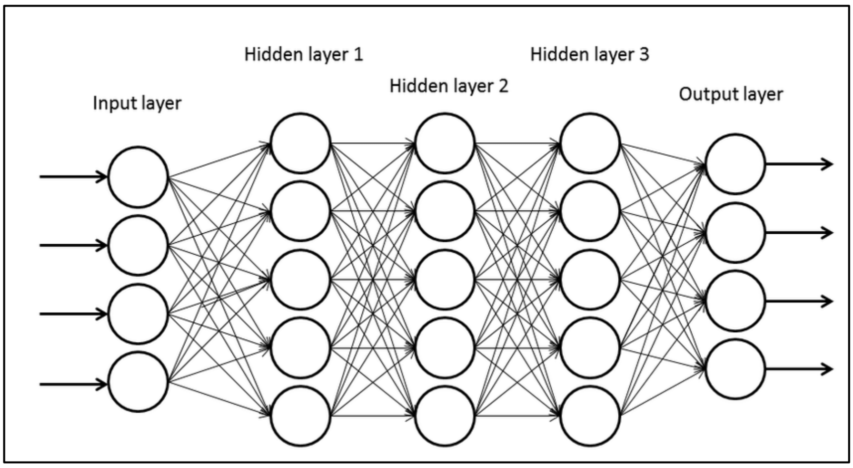


Figure 3. 2 General architecture of deep neural network

A classical DNN with a single hidden layer can be mathematically stated with these Equations 3.11:

$$u_j = \sum_{i=1}^{N_{imp}} X_i W_{ij} + W_{0j}$$

$$H_j = f(u_j)$$

$$V_k = \sum_{i=1}^{N_{hid}} L_j m_{jk} + m_{0k}$$

$$O_k = g(V_k)$$

Equation 3. 11

The results of the hidden layer (H_j) will be acquired by summing the products of the inputs (X_i) and the weight vectors W_{ij} in addition to a hidden layer's bias term W_{0j}, and then translating this sum using a transfer function f. The commonly used transfer functions are logistic and hyperbolic tangent (Hussain & Jeong, 2015). Similarly, the outcome of the output layer O_k is obtained by summing the products of hidden layer's outputs L_j and weight vectors m_{jk} and output layer's bias term (m_{0k}), and converting the sum using the transfer function f.

3.3.4. Decision tree (DT)

DTs or regression trees are non-parametric techniques that can explain the response of a dependent variable. The algorithm maps the whole dataset by representing different variables as internal nodes as the inputs and the leaf nodes as the outputs. Usually, the decision trees do not require the data normalization and other data preparation requirements before its application, and can be applied to the data with outliers. This algorithm is also known as a white-box algorithm, hence the behaviour of the model along with the structure of predictions can be analysed through the visual and instinctive interpretation of the results.

3.3.5. Random forest (RF) regression

The RF regression algorithm can be the improved form of the decision tree algorithm, and was first introduced by Breiman in 2001 (Gislason et al., 2006). In RF, several small decision trees are generated from the random subsets of the dataset. Usually it categorizes each input vector into the random trees in order to build a forest and decide the output of each class based on the vote majority (Fawagreh et al., 2014). The final model is a voting model of all the generated random trees. Each random tree predicts each subset of the dataset by following the branches of the tree in line with the splitting rules and assessing the leaf. The importance of each effective factor can be estimated based on the mean decrease accuracy received at the end of the model. As all single predictions are taken with the same importance, and are based on the subsets of the dataset, the final prediction inclines to show a less variability than the single predictions. The concept of pruning can reduce the complexity of the model by replacing sub-trees that provide little predictive power only with leaves.

3.4. Preparation of Input Datasets

The dataset used in this work was obtained from the British Geological Survey (BGS). The stream sediment baseline geochemistry data was estimated from the samples collected across the central Wales region during the late 90s by BGS in Figure 3.1. The concentration values of the stream sediments are usually the representation of the parent geological material in the region and is developed over millions of years. Sampling was based on the collection of heavy minerals accumulated from the 1st and 2nd order streams. An active stream sediment was moved through a 2 mm sieve, collected in a wooden pan (about 3-4 kg), and condensed by panning to about 60 g. This process was repeated using an additional sediment from the same site, and the two concentrates combined were inspected on-site for heavy minerals and collected in a Kraft bag. A total of 407 samples were collected and analysed for Zn and Pb. The geochemical sample points were split into the training and the testing group with 80% (325 samples) and 20% (82 samples), respectively. The criterion to split the data was that the sample points located in the homogenous geological group and on the same streams should be selected and separated from the training data as the test data (Figure 3.3).

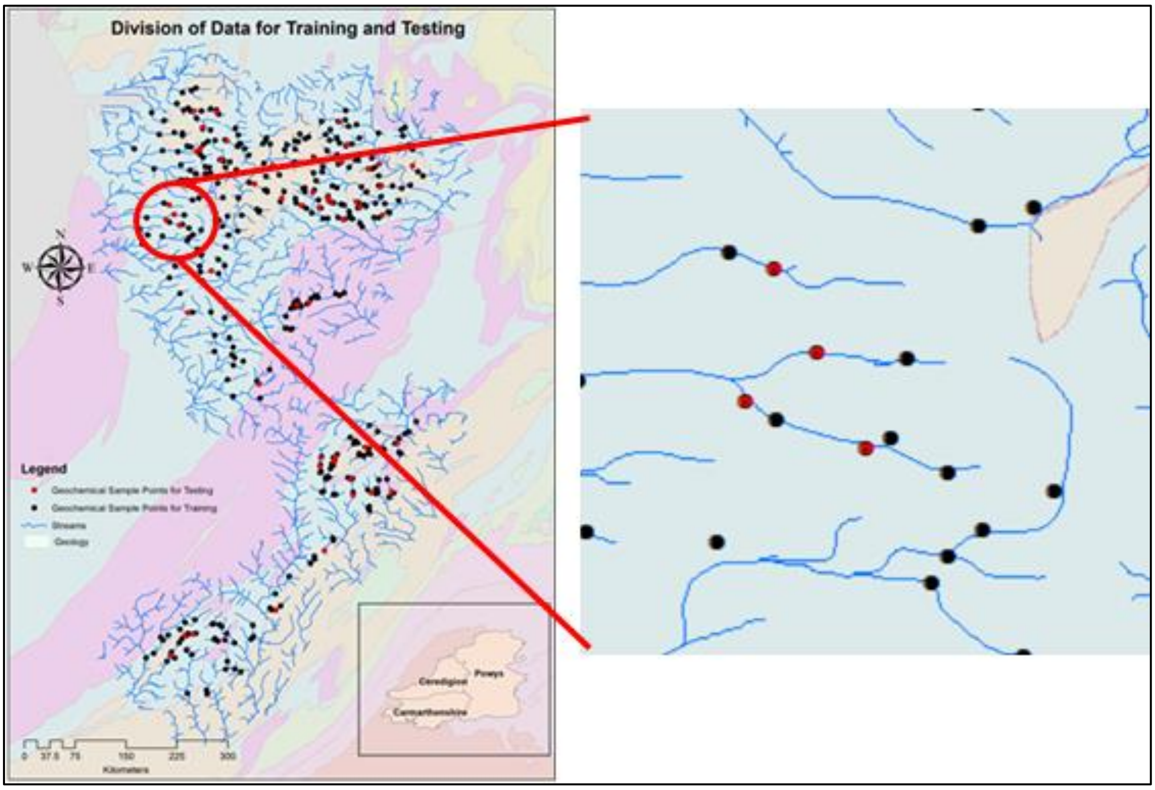
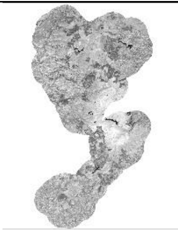
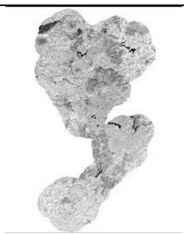


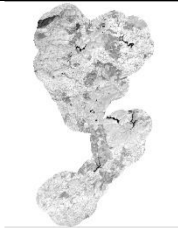
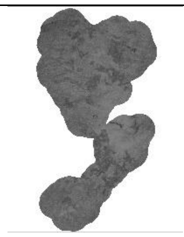
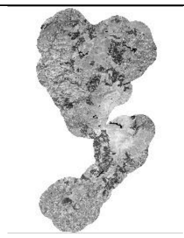
Figure 3. 3 Splitting of geochemical sample points into train (black dots) and test (red dots) group (N = 407)

Prediction of Pb and Zn in the stream sediments was made using the five machine learning algorithms at stream levels using the Multispectral Instrument (MSI) Sentinel-2 satellite data of 2015. Although there is a time difference between the filed sampling and the satellite data acquisition, it will take a very long for the parent material to be changed. The time required for developing a geological profile is usually altered by the significant tectonic activities (earthquakes, tsunami, landslides), earth's inner temperature, gravity, climate, and weathering. Hence, the time difference between the collection of samples and the satellite data acquisition cannot be significant compared to the geological timelines, and therefore, the spectral profiles of the region will merely be changed. The Sentinel-2 satellite has thirteen spectral bands, which range from the Visible and Near Infrared (NIR) to the Short-Wave Infrared (SWIR) of the electromagnetic spectrum. The spectral and spatial profile of Sentinal-2 satellite is given in Table 3.1.

Table 3. 1 Data characteristics of Sentinel-2 satellite

Band No	**Spectral band**	**Central wavelength (nm)**	**Bandwidth (nm)**	**Spatial resolution**	**Satellite image of spectral band of the studied area**
B1	Coastal aerosol	442.7	21	60	
B2	Blue	492.4	66	10	
B3	Green	559.8	36	10	

B4	Red	664.6	31	10	
B5	Red edge–1	704.1	15	20	
B6	Red edge–2	740.5	15	20	
B7	Red edge–2	782.8	20	20	
B8	NIR	832.8	106	10	
B8A	Narrow NIR	864.7	21	20	

B9	Water vapour	945.1	20	60	
B10	SWIR–Cirrus	1373.5	31	60	
B11	SWIR–1	1613.7	91	20	
B12	SWIR–2	2202.4	175	20	

Sentinel-2 has a 12-bit radiometric resolution, which means that the sensor can differentiate 4096 grey-levels in each spectral band. Due to its high spatial, spectral, and radiometric resolutions, the Sentinel-2 data is usually called the super-spectral satellite data. Many research works have applied, tested, and validated the Sentinel-2 dataset for several applications, i.e. geology (Van der Meer et al., 2014), hydrology and hydrogeology (Karaman et al., 2018), mining (Lobo et al., 2018), and mineral exploration (Mielke et al., 2014). Similarly, the dataset has also been successfully used for stream sedimentation mapping (Karim et al., 2019) and geochemical mapping (Cardoso-Fernandes et al., 2018). Usually, the stream sediments have a huge capacity to hold the traces of different elements along with their chemical compositions that can be a good indicator of related mineralization. The Sentinel-2 satellite data was divided into four input subsets (Table 3.2) based on the spatial resolution of each band; however, the topographic elevation extracted from the ASTER Global Digital Elevation Model (GDEM) and the slope was also considered a part of each subset. The satellite image of the studied area was overlaid with stream sampling locations, and the value of each spectral band was extracted against each sample location and prepared in a separate database to be further used in the machine learning models.

Table 3. 2 Input datasets composed of Sentinel-2 satellite and topographic variables

Input subset	**Parameters**	**Properties**
DS-1	B2, B3, B4, B8, elevation, and slope	All the spectral bands of Sentinel-2 satellite with 10 m spatial resolution along with topographic elevation and slope
DS-2	B5, B6, B7, B8A, B11, B12, elevation, and slope	All the spectral bands of Sentinel-2 satellite with 20 m spatial resolution along with topographic elevation and slope
DS-3	B1, B9, B10, elevation, and slope	All the spectral bands of Sentinel-2 satellite with 60 m spatial resolution along with topographic elevation and slope
DS-4	B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B10, B11, B12, elevation, and slope	All the spectral bands of Sentinel-2 satellite along with topographic elevation and slope

All the five machine learning models were trained and validated on the input subsets of DS-1 to DS-4 for prediction of the Zn and Pb stream sediments, as shown in summary Figure 3.4.

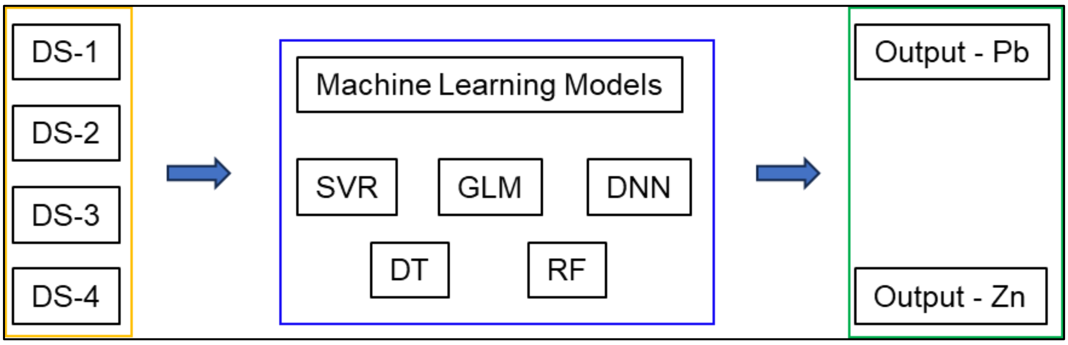


Figure 3. 4 Summary of the four input sets and machine learning models used for predictions of Pb and Zn

Performing the models were evaluated based on the coefficient of determination (R^2) as per Equation 3.12, which represents the total variation in the predictions made by the model as given in the following equation, and its value ranges between 0 (poor) and 1 (perfect) (Piepho, 2019):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \qquad \text{Equation 3. 12}$$

where y in the equation is an estimation of the average predictions, and N is the number of stream sample points.

3.5. Results and Discussion

The results obtained on the prediction of Pb and Zn by applying various machine learning applications are given in this section.

The results based on DS-1 showed that Zn was best predicted with the RF model with the R^2 values of 0.74 and 0.7 during training and testing. However, for Pb, the best prediction was made by SVR with the R^2 values of 0.72 and 0.64 for training and testing, respectively. In the Random Forest (RF) model, the weight allocated to the near-infrared (NIR) spectral band was determined to be the highest, while the weight assigned to the topographic slope was found to be the lowest. The research work conducted by Cozzolino and Moron (2004) also concluded that the Zn mineral ore could be separated from waste and other materials using the NIR reflectance spectroscopy. Pb also showed a good correlation with the NIR region along with SWIR due to its high reflectance in these two spectral bands, as also discussed by Hauff (2008). Using the DS-2 subset of the input dataset, Zn was best predicted through the SVR model with the R^2 values of 0.73 and 0.63 during training and testing. However, for Pb, the best prediction was made by DNN with the R^2 values of 0.72 and 0.61 for training and testing, respectively. DS-2 has eight input parameters with six Sentinel-2 satellite spectral bands including the NIR edge and SWIR. These spectral bands are highly suitable for the mapping of Zn and Pb on the surface of the earth and in the outcrops regions (Chattoraj et al., 2020; Hauff, 2008; Hunt, 1977). The third input subset DS-3 showed that Zn and Pb both were best predicted through the SVR model with the R^2 values of 0.61 and 0.55 for Zn and 0.49 and 0.39 for Pb during training and testing, respectively. DS-3 has five input parameters with three Sentinel-2 satellite spectral bands and the R^2 value for DS-3 was less than the first two input subsets. This might be related to the spectral bands of DS-3, which are usually suitable for monitoring the atmospheric conditions and are less suitable for the mapping of features on the surface of the earth (Vanhellemont, 2019). The other main reason could be the spatial resolution of the spectral bands, which is 60 m, and is coarser than DS-1 and DS-2. The fourth input subset DS-4 showed that Zn and Pb both were best predicted through the RF model with the R^2 values of 0.74 and 0.64 for Zn and 0.73 and 0.63 for Pb during training and testing, respectively. DS-4 was composed on 15 parameters including all the spectral bands of Sentinel-2 satellite ranged from visible, NIR, and SWIR along with topographic elevation and slope.

Overall, SVR and RF outperformed the other machine learning models (i.e., GLM, DNN, and DT) with the highest testing R^2 value. Similarly, the input subset DN-1 consisting of visible and near-infrared spectral bands along with topographic elevation and slope of the studied area can be considered as the most suitable for prediction of Zn and Pb. The second-best input subset was DS-4, which consisted of all the spectral bands, the topographic elevation, and the slope of the studied area. The R^2 values of all the ML models were significant at $p < 0.01$, both for testing and training, as given in Table 3.3.

Table 3. 3 Coefficients of determination as obtained for Zn and Pb prediction using machine learning models

			SVR	**GLM**	**DNN**	**DT**	**RF**

DS-1	Zn	Training	0.71	0.52	0.73	0.67	0.74
		Testing	0.69	0.5	0.69	0.63	0.7
	Pb	Training	0.72	0.6	0.64	0.44	0.64
		Testing	0.64	0.51	0.55	0.35	0.54
DS-2	Zn	Training	0.73	0.7	0.72	0.56	0.7
		Testing	0.63	0.59	0.6	0.46	0.59
	Pb	Training	0.7	0.7	0.72	0.52	0.71
		Testing	0.58	0.59	0.61	0.42	0.62
DS-3	Zn	Training	0.61	0.59	0.6	0.41	0.57
		Testing	0.55	0.52	0.52	0.34	0.47
	Pb	Training	0.49	0.47	0.48	0.25	0.41
		Testing	0.39	0.35	0.36	0.16	0.28
DS-4	Zn	Training	0.56	0.54	0.72	0.72	0.74
		Testing	0.44	0.44	0.59	0.59	0.64
	Pb	Training	0.6	0.6	0.62	0.52	0.73
		Testing	0.48	0.49	0.52	0.4	0.63

The research work conducted by Perez et al. (2011) and Abbaszadeh et al. (2013) also concluded that SVR was better for mineral resource exploration and estimations, while Sheng et al. (2015) concluded that RF was better for mineral resource exploration and estimations. The spatial distribution of the original and predicted Zn and Pb stream sediment contents (Figure 3.5) shows significant degrees of association.

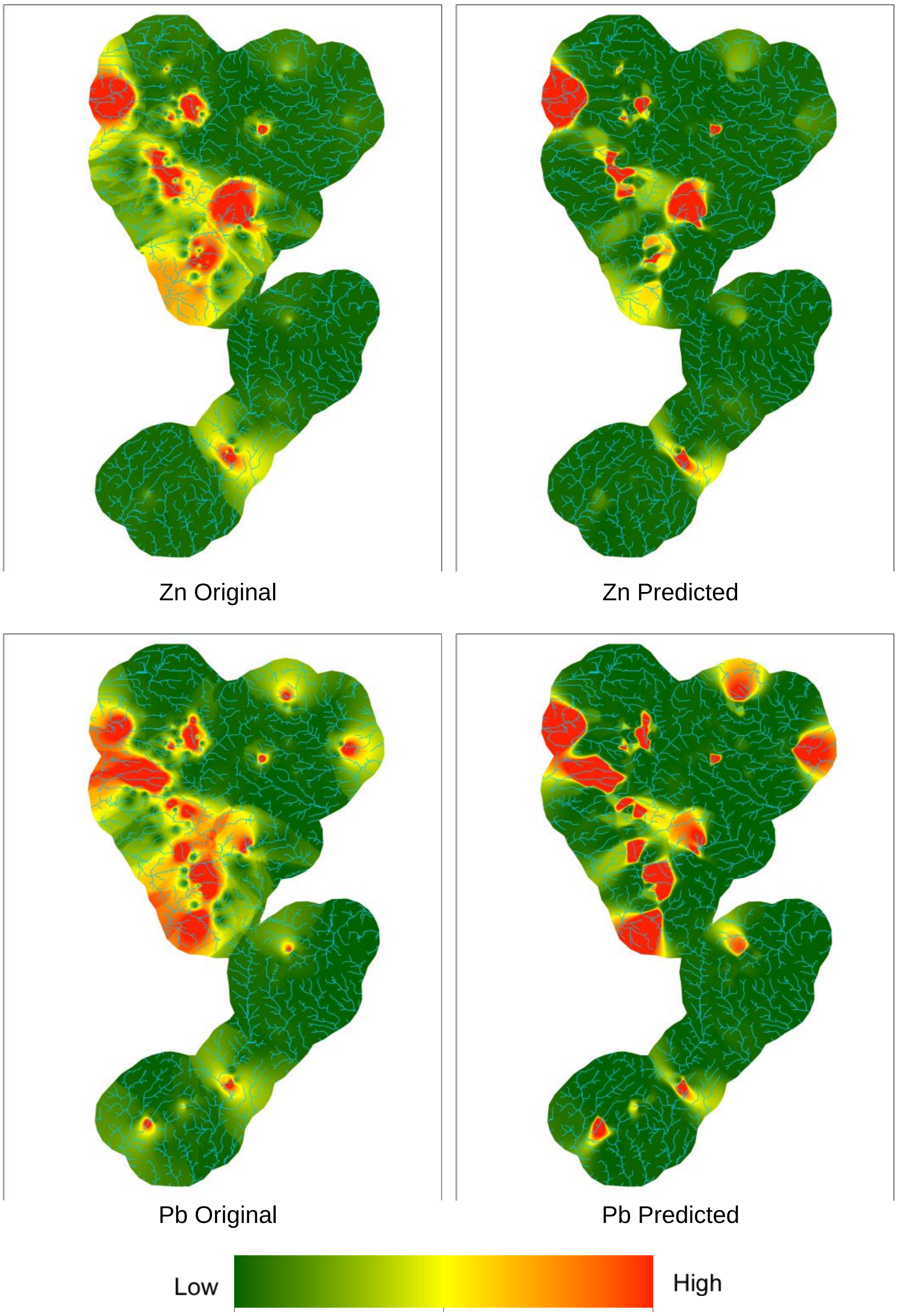


Figure 3. 5 Spatial distribution of original and predicted Zn (through RF) and Pb (through SVR) concentrations in the study area

The predicted Zn and Pb geochemical surfaces are also under the bedrock permeability and geological fault line map of the region as well (Figure 3.6). Permeability is the porosity of a rock; a higher permeability means a higher movement of liquid through the bed rock geology, and *vice versa*.

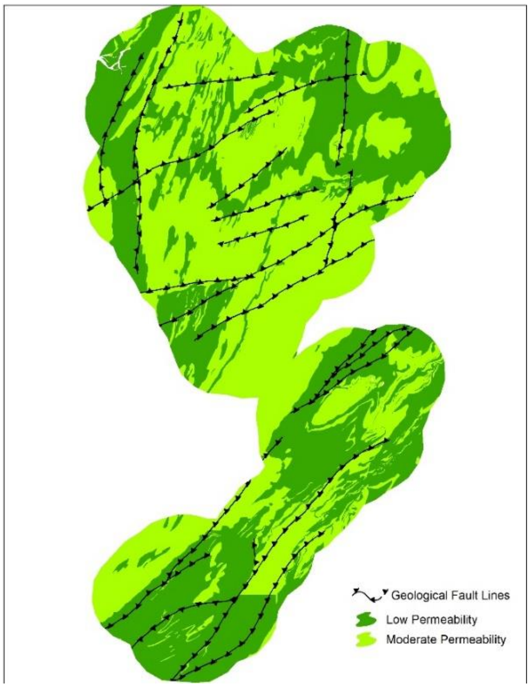


Figure 3. 6 Bedrock permeability and geological fault line map of the Central Wales

There are only two, i.e., low and moderate, zones of bedrock permeability present in the studied area, and high concentration pockets of Zn and Pb are present in the low zones. The research work conducted by Freedman (1972), Bouabdellah and Sangster (2016), and Gao et al. (2019) also concluded that the higher concentrations of Zn and Pb were associated with the lower permeability of the bedrock. The other important observation is that the predicted high concentration values of Zn and Pb are in a close vicinity of the geological fault lines, which also indicate this mineralization in the studied area.

Although the values of the predicted stream sediment concentrations are higher than the original, the predicted values coincide spatially with the original dataset.. The higher Zn and Pb content distribution can also be related to the ongoing mining activities of the same elements in the studied area. However, other areas in the central Wales are enriched in these elements as well, and could be the future exploitation zones.

3.6. Conclusions

Five machine learning algorithms including support vector regression, generalized linear model, deep neural network, decision tree, and random forest regression were trained and tested for prediction of the Zn and Pb stream sediments using the reflectance spectroscopy of Sentinel-2 satellite data along with topographic elevation and slope. For a better prediction and in order to check the impacts of different spectral channels and spatial resolution of the data, the input datasets were divided into four subsets, and the accuracy was measured using the coefficient of determination (R^2). The prediction of stream sediments based on the DS-1 input dataset was the best with the RF for Zn and SVR for Pb as compared to the other models and input datasets with the measured R^2 values of 0.7 (for Zn) and 0.64 (for Pb). This could be related to the NIR band in the dataset in association with the high spatial resolution of 10 m. The input dataset DS-4 could be concluded as the second-best for the predictions of the Zn and Pb stream sediments using the RF machine learning model with the R^2 values of 0.64 (for Zn) and 0.63 (for Pb). DS-4 consisted of 15 input parameters including 13 Sentinel-2 spectral bands, the topographic elevation, and the slope of the area. The spatial locations of the predicted concentration values are not only under the original dataset but also with the existing mines present in the studied area. The proposed methodology for the prediction of the concentration of the stream sediments is useful for the mapping of the mineral enriched zones and future mineral resource exploration in the studied area. However, applying this methodology to the new studied area requires the new training and testing of the models so that the model can adapt to the new data for predictions of particular stream sediments. In the future research works, the field-based reflectance spectroscopy, topographic factor (aspect, surface curvature), and hydro-meteorological factors (streams flow, rainfall) can be added to the machine learning models along with the hyper-spectral satellite-based reflectance datasets for more accurate concentration predictions of the stream sediments.

4. PREDICTIVE MODELLING AND COMPARATIVE EVALUATION OF GEOSTATISTICAL MODELS FOR GEOCHEMICAL EXPLORATION THROUGH STREAM SEDIMENTS

This chapter is based on the following publication

Mahboob, M. A., Celik, T., & Genc, B. (2020). Predictive modeling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments. *Arabian Journal of Geosciences*, *13*, 1-21

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

4.1. Introduction

Stream sediments usually result from the erosion and transportation of soil and rock remnant, and other materials present upstream of the sampling locations within a basin (Merritt et al., 2003). Thus, stream sediments are considered to be representative of the geochemistry of the rocks present in the upstream drainage basin. Usually, stream sediments are the primary sampling source for geochemical exploration, primarily in areas where distinct drainage systems exist due to the local topography (Fletcher, 1997; Salminen & Gregorauskien, 2000; Zuluaga et al., 2017). Geochemical exploration sampling from stream sediments is a suitable technique for identification of areas of interest with significant mineralization content, particularly in the early stages of mining for undiscovered minerals (Arndt et al., 2017; Carranza, 2017; Cheng, 2007, 2012; de Mulder et al., 2016; Khalajmasoumi et al., 2017; Nieto et al., 2014). For the best results, it is always preferred to get as many samples as possible to work with a dense and large sampling dataset. However, high-density sampling of stream sediments is challenging due to the cost involved in getting geographical accessibility. Therefore, a limited number of samples are usually collected from the field and after analysing in the lab are used for estimation for larger areas using geostatistical interpolation models (Ottesen et al., 1989).

Geostatistical models have been extensively used as a powerful tool for providing accurate estimates of any spatial/temporal phenomena at unsampled locations using sampled data along with a quantification of the related uncertainty (Cai et al., 2018; Dagdelen & Vega, 1997; Goovaerts, 2000; Liu et al., 2006; Y. Liu et al., 2018; Lu & Wong, 2008; Moral, 2010; Piel et al., 2013; Ssempiira et al., 2017). Several spatial interpolation methods, including geostatistics, have been developed in the past to better predict spatial distributions based on limited samples. However, the accuracy of the interpolated surface varied widely between models (Robinson & Metternicht, 2006). In general spatial interpolation methods can be divided into three main categories, i.e. with respect to the number of points used for prediction (Global/Local), with respect to the surface smoothness (Exact/Approximate), and with respect to the error estimation (Deterministic/Stochastic) (Li & Heap, 2014).

The most common interpolation methods used for geospatial characterization of geochemical data from streams are Nearest Neighbour (NN) (Li & Heap, 2014; Yang et al., 2004), Triangular Irregular Network (TIN) (Li & Heap, 2014; Wu et al., 2011; Yang et al., 2004), Inverse Distance Weighting (IDW) (Li & Heap, 2014; Lu & Wong, 2008), Radial Based Functions (RBF) (Ding et al., 2017) and Kriging (Kleinschmidt et al., 2000; Panahi et al., 2004; Wu et al., 2011). In general, Kriging is the most applied, tested and validated geostatistical model in the field of spatial interpolation. The most significant advantage of this model over several other spatial and statistical models is that Kriging considers the spatial correlation of the data for the predictions. Several types of Kriging have evolved over the time, but the most commonly used are Ordinary Kriging (OK), Simple Kriging (SK), Universal Kriging (UK), Indicator Kriging (IK) and Co-Kriging (Panahi et al., 2004). Among all these techniques, the predistortions are highly dependent not only on the location of the data but also on the accurate development of semivariogram model. But the development of the required semivariogram model is time taking and subjective. In addition to that, the transformation of the data may also be required in non-stationary conditions and directional trends needed to be considered (Nieto & Toffait, 2007). Hence different geospatial characterization methods have been used in the past to characterize geochemical mineralization which tends to overextend predictions to areas not really influenced by the original stream mineralization condition. As it is not required that a stream sediment sample (S_1)

be collected at a particular location (L_1) on a stream, which represents the geochemical value at L_1, it is quite possible that it is a representative of the upstream location L_2 or L_3 i.e., the potential source of the mineralized zone (Figure 4.1).

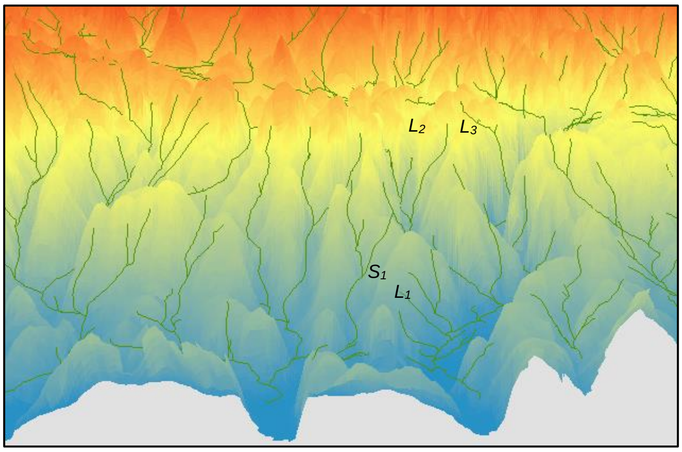


Figure 4. 1 3-D topographical representation of streams along with geochemical sampling and value locations

Therefore, this research focuses on the development of a new approach for geospatial characterization based on a continuous geochemical prediction surface up to the extent of sampling streams. Hence geochemically spatial predictions of continuous streams will provide improved estimations in streams considering the source of the mineralization.

Geochemical data are usually geospatial data as it can be expressed as x, y, and z, where x and y are location coordinates (latitude, longitude or easting, northing), and z represents the recorded value (i.e., elements concentration) at those coordinates. Geochemical data are usually stored in a spatial format as point data with location and value, hence can be processed in Geoinformatics (Zuo et al., 2016). With the advancement in computer technology, several geographic information systems (GIS) based numerical codes are available with a strong component of Geostatistics embedded in them. Surface interpolations through geostatistical modelling in GIS are very powerful for estimating geochemical values (Johnston et al., 2012; Li & Heap, 2014). This research study has also incorporated a very strong GIS numerical package ESRI's ArcGIS Geostatistical Analyst, which is equipped with advanced geostatistical predictive models for predictive modelling. The two widely used interpolation methods, including the geostatistical (IDW and Kriging), were selected and applied to test their efficiency for geochemical characterization and prediction along the unsampled streams. Also, these, models have been extensively used in the past, but none of them has been applied and tested for the prediction of geochemical values at individual stream levels, which is a new proposed methodology of this research study. Furthermore, the Geochemical Accumulation Index (GAI) has been developed in this research to identify the mineral enriched streams and for identification of potential mineral prospects in the study area.

4.2. Materials and Methods

4.2.1. Study area

The study area for this research is located in Central Wales of Great Britain (Figure 4.2) and has a prolonged history of gold mining, with the main centre of activity being the "Dolgellau Gold Belt" where the Clogau and Gwynfynydd mines were very active gold mines. Particularly towards South, in the Welsh Basin, where gold has been mined since the Roman times. Besides gold, several other secondary products such as lead, copper, zinc, iron, nickel sulphides and others are present throughout the Wales area.

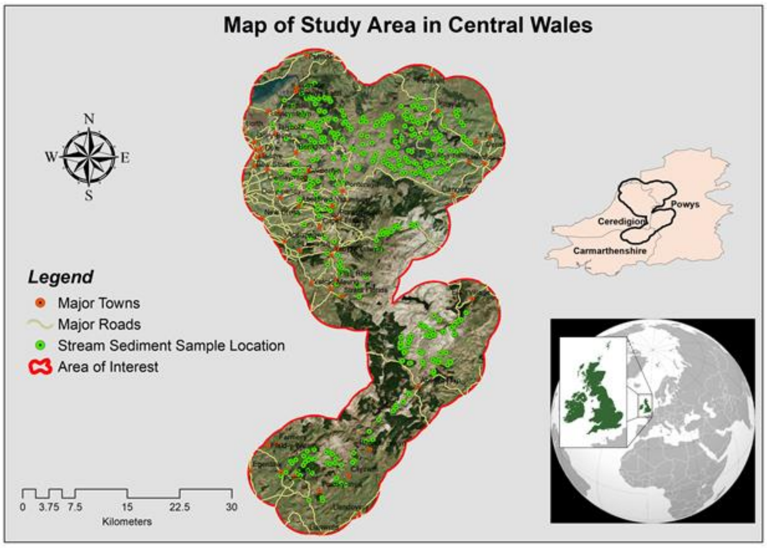


Figure 4. 2 Study area map highlighting the location of major towns, roads and stream sediments sample points

The dataset used in this study (WF/MR/93/013) was obtained from the British Geological Survey (BGS) (Brown, 1993). The stream sediment baseline geochemistry data was estimated from samples collected across the central Wales region by BGS as shown in Figure 4.2. Sampling was based on the collection of heavy minerals accumulated from the 1st and 2nd order streams. Active stream sediment was moved through a 2 mm sieve, collected in a wooden pan (about 3-4 kg) and condensed by panning to about 60 g. This process was repeated using additional sediment from the same site, and the two concentrates combined together were inspected on-site for heavy minerals and collected in a Kraft bag. A total of 407 samples were collected and analysed for 15 different elements, including titanium (Ti), manganese (Mn), iron (Fe), vanadium (V), nickel (Ni), copper (Cu), zinc (Zn), zirconium (Zr), tin (Sn), antimony (Sb), barium (Ba), cerium (Ce), lead (Pb), arsenic (As) and gold (Au). Gold was estimated on 60g of the sample by atomic absorption spectrometry after grinding and dissolution in aqua regia (a yellow-orange fuming liquid), with a lower confidence limit of detection as 10 parts-per-billion (ppb) Au. All other elements (Ti, Mn, Fe, V, Ni, Cu, Zn, Zr, Sn, Sb, Ba, Ce, Pb, and As) were determined by X-ray fluorescence analysis of milled sub-samples in parts-per-million (ppm). The descriptive statistic of the 15 elements is given in Table 4.1.

Table 4. 1 Summary statistics of geochemical elements in the study area (N = 407)

Elements	Min	Max	Mean	Variation coefficient	Skewness (Pearson)
Ti	730.0	7050.0	5247.0	19.6	-2.0
Mn	10.0	9210.0	825.3	100.3	5.9
Fe	9400.0	76100.0	46805.7	25	-0.8
V	0.0	360.0	102.7	36.3	0.0
Ni	7.0	656.0	47.9	107.6	7.9
Cu	1.0	19917.0	549.4	414.5	5.4
Zn	33.0	17454.0	777.2	309.4	4.7
Zr	0.0	304.0	175.6	22.4	-1.2
Sn	0.0	260.0	10.9	243.4	5.6
Sb	0.0	693.0	14.0	414	7.1
Ba	62.0	2907.0	457.8	38.7	7.1
Ce	10.0	100073.0	12516.1	133	2.4
Pb	5.0	13397.0	1157.4	248.5	2.6
As	0.0	80.0	12.8	67.5	3.3

Au	0.0	8765.0	36.3	1258.8	17.6

From Table 4.1, it can be observed that the distributions of most geochemical parameters are positively skewed, and their frequency diagrams do not follow a normal distribution (Bai & Ng, 2005; Beedles & Simkowitz, 1978). Coefficient of variation (CV) values <10% and >90% indicate low and high variability respectively in the data (Reed et al., 2002), and in this dataset, Ti and Au have the lowest and the highest variability, respectively. It can be observed through summary statistics that the data need to be transformed to a normal distribution and as all the variables have variability so spatial and geostatistical modelling can be applied for further predictions (Komnitsas et al., 2010). Further, the correlation coefficient (Pearson coefficient) matrix, as shown in Figure 4.3 was developed to highlight the statistical relationship between geochemical variables (Facchinelli et al., 2001). The correlation coefficient matrix reveals that the correlation coefficient value r ranges from negative 0.66 (between Ti and Pb) to positive 0.69 (between Fe and V). The r values are statistically significant at a 0.05 (5%) confidence level.

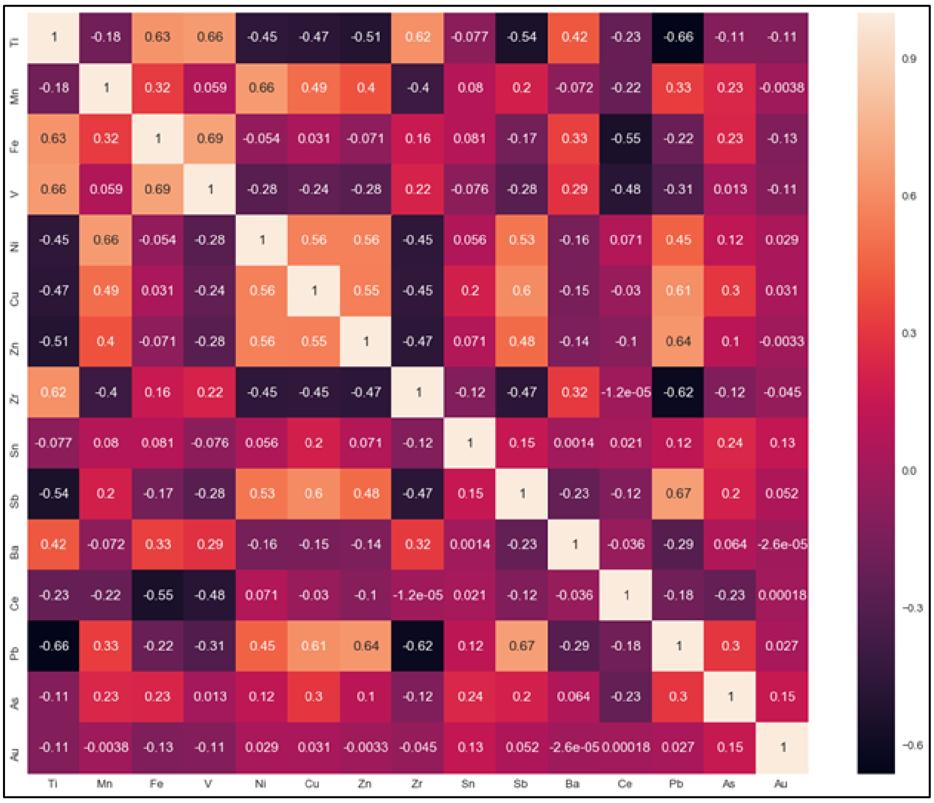


Figure 4. 3 The correlation matrix between different geochemical variables highlighting the positive and negative associations

4.2.2. Directional distributional trend of geochemical data

Usually, the standard deviational ellipse (SDE) is used to capture the spatial trend for a set of measured observations (Lefever, 1926). The SDE is based on standard deviations computed on spatial coordinates (x, y) of the observations and has been applied to describe bivariate features. Most commonly, it is applied to assess the geographical distribution trend of the features concerned by analyzing both of their dispersion and direction. For a given set of n samples $\mathcal{M} = \{(x_i, y_i, z_i)\}_{i=1}^{n}$ the corresponding SDE is defined by three parameters: spatial mean (or average spatial location) $(\bar{x}, \bar{y})$, spatial dispersion (or concentration) (σ_x, σ_y) and orientation of the geochemical data θ. In addition to the traditional spatial mean center (gravity on the distribution), weighted spatial mean or median of the data could also be used as other options (Wang et al., 2015). The average spatial location of the SDE is computed using the Equation 4.1:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{Equation 4. 1}$$

the average spatial location is subtracted from each sample to center the observed samples at the origin as indicated in Equation 4.2.

$$\tilde{x}_i = x_i - \bar{x}, \tilde{y}_i = y_i - \bar{y} \qquad \text{Equation 4. 2}$$

the centered samples are then used to compute the orientation of the SDE as given in Equation 4.3.

$$\theta = \arctan\frac{A + \sqrt{A^2 + 4B^2}}{2B} \qquad \text{Equation 4. 3}$$

where

$$A = \left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right), \quad B = \sum_{i=1}^{n} \tilde{x}_i\tilde{y}_i \qquad \text{Equation 4. 4}$$

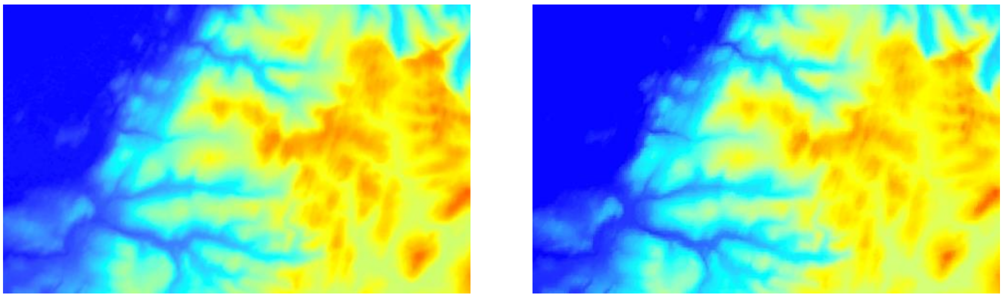
and finally, the spatial dispersion of the SDE is computed using the shifted and oriented samples based on Equation 4.5.

$$\begin{aligned} \sigma_x &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\tilde{y}_i\sin\theta + \tilde{x}_i\cos\theta)^2} \\ \sigma_y &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\tilde{y}_i\cos\theta - \tilde{x}_i\sin\theta)^2} \end{aligned} \qquad \text{Equation 4. 5}$$

A rule-of-thumb drawn from the Rayleigh distribution recommends that a first, second and third standard deviational ellipse will cover approximately 63%, 98% and 99.99% of the features (geochemical data) in two dimensions (x and y) (Wang et al., 2015). In this research, the directional distributional trend of all the 15 elements was assessed with one standard deviation, which covers 63% of the data in ArcGIS software.

4.2.3. Stream delineation using digital elevation model

Global Digital Elevation Map (GDEM) derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) was used to extract the streams present in the study area. The GDEM was re-projected from Geographical Coordinate Systems to Projected Coordinates Systems for spatial referencing. The pre-processing of GDEM is always required and recommended before the extraction of any spatial information. Usually, a raw DEM often has a large number of “sinks” (or false depressions), i.e. single or multiple pixels with extreme low elevation and are completely enclosed by immediate higher elevation pixels (Jones, 2002). The sinks, if present in the DEM, then must be removed using the fill operation available in any raster GIS packages. In this research study, FILL operation of ArcHydro model of ESRI’s ArcGIS was applied as shown in Figure 4.4. The FILL operation eliminates the sinks by either (a) raising the elevation on the sink cell to that of its lowest neighbor or (b) lowering the elevation of the lowest neighbor adjacent to the sink (Jones, 2002).



Before Fill After Fill

Figure 4. 4 The effect of FILL operation on the GDEM as on left side false depressions were filled after applying the operation

After pre-processing of the GDEM, the next step is to determine the flow direction, which is the ability to control the direction of water from every pixel in the raster to its downslope neighbor. Several algorithms can be used to determine the direction of flow, including the Multi-Flow Direction (MFD), D-Infinity (DINF), and D8 (Płaczkowska et al., 2015). Among all these, the D8 (eight flow direction) method is the simplest and the most effective. D8 assigns the flow (in terms of coding) from individual pixel to one of its eight connecting cells, either adjacent or diagonal, in the direction with the steepest downward slope.

Once the flow direction raster is created, the next step is to calculate areas where water may accumulate as a result of rainfall, known as the Flow Accumulation raster. Mainly, the value to each cell in the resulting raster the value of a pixel in the flow accumulation raster is the summation of all pixels that flow into that particular pixel. The objective is to simulate the flow, or possible flow, of water to form creeks, streams, and/or rivers. A threshold value is required to determine the minimum number of cells that can generate the flow, i.e. in this study a threshold of 150 was set, which means that a minimum of 150 cells should contribute to generating a flow into a cell (Wei et al., 2016). The threshold value is selected from a range [100, 359] according to the number of connected pixels which fall on the stream to generate an adequate amount of flow. After highlighting the streams in the area, the next step is to convert them from raster to vector by using the Streams to Feature tool of ArcHydro. The mineral prediction at stream level will provide more realistic information and help to mark the mineral enriched zones as compared to a classical square regional predicted surface.

4.3. Spatial Interpolation Methods

4.3.1. Inverse distance weighting (IDW)

IDW is a local, exact and deterministic interpolation model, which means it predicts a value at unsampled locations using a subset of sampled values, and then it will be the same, i.e. the line pass through the points and gives no error assessment. Due to its simplicity and computationally non-intensiveness, several researchers, without having much knowledge in spatial statistics and geostatistics will use IDW as a default method to produce a surface when data values exist only at sampled locations. This model assumes that data values that are nearby to one another are more similar than those that are at far distances (Wong, 2016). To calculate a value for any unsampled location, IDW uses the measured values close to the unsampled location.

The IDW model predicts the unknown geochemical value $\hat{g}^{(e)}(\mathbf{r}_u)$ of element e at location $\mathbf{r}_u$, using the observed (or known) element values $g^{(e)}(\mathbf{r}_i)$ at sampled locations $\mathbf{r}_i$ as Equation 4.6:

$$\hat{g}(\mathbf{r}_u) = \sum_{i,\mathbf{r}_i \in \mathcal{N}_u^k} w_i g(\mathbf{r}_i)$$ Equation 4. 6

where w_i is the weight for the observed value $g^{(e)}(\mathbf{r}_i)$ for geochemical element e and $\mathcal{N}_u^k$ is the set of k nearest sample locations around the spatial location $\mathbf{r}_u$. In other words, the predicted value $\hat{g}^{(e)}(\mathbf{r}_u)$ at $\mathbf{r}_u$ is a weighted sum of observed values $g^{(e)}(\mathbf{r}_i)$ with Equation 4.7.

$$\sum_{i=1}^{k} w_i = 1$$ Equation 4. 7

the weight w_i is inversely proportional to the α^{th} power of Euclidean distance between $\mathbf{r}_u$ and $\mathbf{r}_i$ and defined as Equation 4.8:

$$w_i = \frac{||\mathbf{r}_i - \mathbf{r}_u||^{-\alpha}}{\sum_{j=1}^{k} ||\mathbf{r}_j - \mathbf{r}_u||^{-\alpha}}$$ Equation 4. 8

where $||\mathbf{r}_i - \mathbf{r}_u||$ is defined as Euclidean distance between $\mathbf{r}_u$ and $\mathbf{r}_i$.

The α parameter is quantified as a geometric form for the weight while other conditions are possible. This condition indicates that if α is more significant than 1, then the distance-decay influence will be more as compared to an increase in distance, and vice versa. Therefore, a small α tend to give predicted values as averages of $g^{(e)}(\mathbf{r}_i)$ in the neighborhood, while large α assign more weights to the nearest points and gradually decrease for points far away. As a result, when $\alpha \to 0$

$$\begin{aligned}
\lim_{\alpha\to 0}\hat{g}(\mathbf{r}_u) &= \lim_{\alpha\to 0}\sum_{i=1}^{k} w_i g(\mathbf{r}_i);\\
&= \lim_{\alpha\to 0}\sum_{i=1}^{k}\frac{||\mathbf{r}_i-\mathbf{r}_u||^{-\alpha}}{\sum_{j=1}^{k}||\mathbf{r}_j-\mathbf{r}_u||^{-\alpha}}g(\mathbf{r}_i);\\
&= \lim_{\alpha\to 0}\sum_{i=1}^{k}\frac{1}{\sum_{j=1}^{k}1}g(\mathbf{r}_i);\\
&= \frac{1}{k}\sum_{i=1}^{k} g(\mathbf{r}_i),
\end{aligned}\qquad \text{Equation 4. 9}$$

and $w_i \to 1/k$. Then, the predicted value is the average of all sampled values. Similarly when $\alpha \to \infty$, $\hat{g}(\mathbf{r}_u) = g(\mathbf{r}_c)$ where $\mathbf{r}_c$ is the closest sample location.

Usually, the geoscientists and mineralogists use the power $\alpha = 2$, which is known as the inverse distance squared weighted model. In this research, three different IDW power values $\alpha = 1$, $\alpha = 2$ and $\alpha = 3$ had been tested to generate the geochemical prediction surfaces. Even though there is no theoretical reasoning in the selection of a particular value over others, and the influence of changing power should be examined by visualizing the output and observing the validation statistics. The measured values near to the unsampled location have more impact on the predicted value than those farther apart. IDW considers that each sampled point has a local impact that reduces with distance. So this model uses weights assigned to the sample values as it gives more weights to points nearby to the unsampled location, and the weights reduce as a function of distance; therefore, the name inverse distance weighting.

4.3.2. Kriging

Kriging predicts the structure of spatial variability through a variogram and incorporates the spatial autocorrelation (Hattis et al., 2012). The Kriging prediction is modeled as the summation of a global trend λ (which is the general trend in the entire data) and a local stochastic variation ε as given in Equation 4.10 (Matheron, 1963):

$$\hat{g}(\mathbf{r}_i) = \lambda(\mathbf{r}_i) + \varepsilon(\mathbf{r}_i)$$ Equation 4. 10

where $\mathbf{r}_i$ represents the spatial coordinates. Depending on the global trend λ, several types of Kriging models have been established. For example, Simple Kriging (SK) assumes λ = 0; Ordinary Kriging (OK) assumes λ as an unknown constant mean; and Universal Kriging (UK) assumes λ as a general polynomial trend given as $z = ax + by + c$, where $x,\ y$ are the variables for the latitude and longitude respectively, and z can be statistically analyzed from the past data. In this research study OK is used because it provides more realistic and reliable predictions based on mean squared errors; it is an unbiased predictor for sparsely sampled regions (Cressie, 1993); and it minimize the influence of spatial outliers (Triantafilis et al., 2001), if present in the data. The OK method predicts the $\hat{g}(\mathbf{r}_u)$ at $\mathbf{r}_u$ using the weighted sum of the data as given Equation 4.11:

$$\hat{g}(\mathbf{r}_u) = \sum_{i,\mathbf{r}_i \in \mathcal{N}_u^k} w_i \mathbf{g}_i(\mathbf{r}_i)$$ Equation 4. 11

The choice of weight in Equation 4.11 should be made in such a way, that w_i yields the lowest mean square estimation error (Cressie, 1993). Besides this the selection of w_i also depends on the type of semivariogram model as suggested by Deutsch and Journel (1992). The experimental semivariogram $\hat{\gamma}(r)$ can be defined as the average square difference of the geochemical data values between the samples, parted by the lag vector r. If there are no explicit directional dependencies present among the data values, then Matheron method-of-moments estimator can be used to develop unidirectional experimental semivariogram as mentioned in Equation 4.12.

$$\hat{\gamma}(r) = \frac{1}{2k(r)} \sum_{i=1}^{k(r)} \{[g(\mathbf{r}_i) - g(\mathbf{r}_i + r)]^2\}$$ Equation 4. 12

where $k(r)$ is the number of sample pairs at lag distance r, (Deutsch & Journel, 1992) after which $\hat{\gamma}(r)$ is then fitted to the model function $\gamma_g(r)$. The weights w_i for ordinary kriging can be given by the $(k+1)\times(k+1)$ linear systems of Equations 4.13 and 4.14.

$$\sum_{i,\mathbf{r}_i\in\mathcal{N}_u^k} w_i\gamma_g(\mathbf{r}_i,\mathbf{r}_j)+\mu=\gamma_g(\mathbf{r}_j,\mathbf{r}_u),\quad j=1,\ldots,k$$ Equation 4. 13

$$\sum_{i,\mathbf{r}_i\in\mathcal{N}_u^k} w_i=1$$ Equation 4. 14

where k is the number of geochemical samples within the proximity of $\hat{g}(\mathbf{r}_u)$; $\gamma_g(\mathbf{r}_i,\mathbf{r}_j)$ is the semivariogram between two geochemical samples $\mathbf{r}_i$ and $\mathbf{r}_j$; $\gamma_g(\mathbf{r}_j,\mathbf{r}_u)$ is the semivariogram between $\mathbf{r}_j$ and predicted $\mathbf{r}_u$; and μ is a linear external parameter called the Lagrange factor. The variance of OK prediction is given by Equation 4.15 whereas the Lagrange factor compensates the uncertainty related to the mean value.

$$\sigma_{\mathrm{E}}^2\hat{g}(\mathbf{r}_u)=\sum_{i,\mathbf{r}_i\in\mathcal{N}_u^k} w_i\gamma_g(\mathbf{r}_i,\mathbf{r}_u)+\mu$$ Equation 4. 15

4.3.3. Semivariogram estimation

In literature, there are several types of semivariogram models (Olea, 2006; Varouchakis & Hristopulos, 2013), but the most commonly used are Gaussian and Circular, and these are also tested in this study along with Exponential and Spherical. The general Equations for all these semivariogram models are given as follows from 4.16 to 4.20:

$$\gamma_g(\mathbf{r})=\sigma_g^2\left[1-\exp\left(-\frac{|\mathbf{r}|}{\xi}\right)\right]$$ Equation 4. 16

$$\gamma_g(\mathbf{r})=\sigma_g^2\left[1-\exp\left(-\frac{|\mathbf{r}|^2}{\xi^2}\right)\right]$$ Equation 4. 17

$$\gamma_g(\mathbf{r})=\sigma_g^2\,[1.5\ |\mathbf{r}|/\xi-0.5\,(|\mathbf{r}|/\xi)^3]\,\theta(\xi-|\mathbf{r}|)$$ Equation 4. 18

where

$$\theta(\xi - |\mathbf{r}|) = \begin{cases} 0 & \text{if} \xi - |\mathbf{r}| < 0; \\ 1 & \text{if} \xi - |\mathbf{r}| \geq 0, \end{cases}$$ Equation 4. 19

and

$$\gamma_g(\mathbf{r}) = \sigma_g^2 \left\{ 1 - \frac{2}{\pi} \cos^{-1}\left(\frac{\mathbf{r}}{\xi}\right) + \frac{2r}{\pi} \sqrt{1 - \frac{\mathbf{r}}{\xi^2}} \right\}$$ Equation 4. 20

in the above Equations σ_g^2 is the variance; $|r|$ is the Euclidean norm of the lag vector r, and ξ is the characteristics length. After the development of semivariogram models, the most appropriate one is selected by comparing the root mean square prediction error statistics between the observed and predicted geochemical values. Depending on the semivariogram model, the prediction error at location S_0 can be computed in terms of the standard deviation as Equation 4.21:

$$\overline{E}(\mathbf{r}_u) = \sum_{i=1}^{k} w_i \gamma \;\; (||\mathbf{r}_i - \mathbf{r}_u||)$$ Equation 4. 21

where $||\mathbf{r}_i - \mathbf{r}_u||$ is the distance among the locations $\mathbf{r}_i$ and $\mathbf{r}_u$.

4.3.4. Sample size requirements for variogram computation

As per recommended in the literature, a minimum of 100 to 150 data points are required to achieve a steady variogram (Voltz & Webster, 1990). This requirement is fulfilled in this research, with 407 geochemical points available for each element. Because the number of sample points was more than the minimum number recomeded in literature, anisotropy (directional trend) can also be determined if present in the data. The ArcGIS Geostatistical Analyst can compute the optimum parameters, such as Major Axis (Range), Minor Axis (Range), and Angle of Rotation (Direction) to constitute the anisotropic effect.

4.4. Potential Mineral Prospects

The potential mineral prospectivity was done by developing the Geochemical Accumulation Index (GAI) using the multivariate overlay analysis to map the high, medium and low mineral enriched streams. Multivariate overlay analysis has been used by several research studies (Correia & Waitzberg, 2003; Hou et al., 2017) to classify the regions in different classes of underlying physical phenomenon. All the predicted surfaces of both IDW and Kriging were classified into three main classes based on Jenks Natural Breaks (JNB) algorithm (Khamis et al., 2018). With JNB, classes were made for each geochemical predicted map based on features similarity and relatively significant data value differences in classes. The standardized classes were combined through linear combination by adding them together to obtain the final high, medium and low mineralized streams.

4.5. Results and Discussion

The directional distributions (standard deviational ellipse) were used to map the geographical trend of different geochemical elements, as shown in Figure 4.5. The ellipses were developed using the first standard deviation (almost 63% of the data) as it is recommended by several research studies (Kent & Leitner, 2007; Seidl et al., 2015) to avoid the effects of any potential spatial outlier if present in the data.

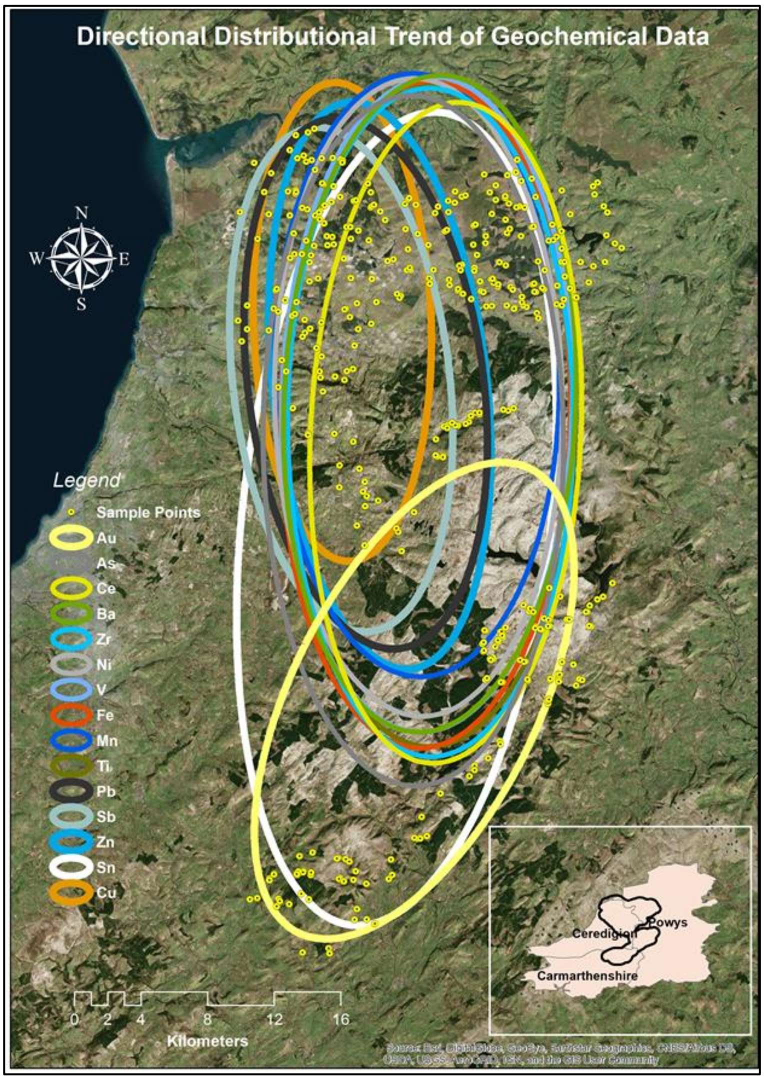


Figure 4. 5 Directional distribution through standard deviational ellipses of all the geochemical elements (N = 407)

In general, all the elements are distributed North to South except the gold, which showed the directional trend in North-East to South-West. For a detailed analysis, these 15 geochmeical elements were grouped in five different classes of direction and geographical concentration, as shown in Figure 4.5.

The results showed that nine geochemical elements (As, Ba, Ce, Fe, Mn, Ni, Ti, V, Zr) have North to South distribution whereas Mn and As are concentrated more towards the North and South respectively. The main regions of As mineralization may be due to metasedimentary rocks and form post-orogenic granites in the

South-East area of the study. Cohen et al. (1999) also found similar results in Wales and associated the variability of As with the geological rocks located within the area. The three elements Pb, Sb and Zn can be grouped because of the same directional trend, i.e. North-West to South-East with Zn as a bigger ellipse as it has more variability. Sn also has North-South concentration but the largest deviational ellipse, which means it has the highest variability and 63% of the values distributed all over the study area. Cu has the most concentrated values, i.e. smaller standard deviational ellipse and is mainly distributed North-South but at West of the study area. Gold has an altogether different distribution, which is North-East to South-West, but 63 percent of the data values are geographically concentrated in the lower part of the study area.

4.5.1. Stream delineation

Stream delineation was performed from the ASTER global digital elevation model with 30m spatial resolution using the D8- algorithm. The flow direction raster shows that the majority of the flow in the basin was from North and North-East to South-West and West, which may be due to the presence of the Irish Sea on the West of Wales. The flow accumulation raster shows the major flow lines (streams) in the area based on the directional raster. Finally, the streams were extracted using that flow accumulation raster with a threshold value of 150, as discussed in the methodology section of this paper.

Figure 4.6 showed that all the sample locations were correctly overlaid on the streams, which can be taken as the verification of the accuracy of extracted streams and the used threshold value. These streams were further converted into a vector data format, and a buffer of 100m was applied to them, to be further used in the prediction models, as shown in Figure 4.7.

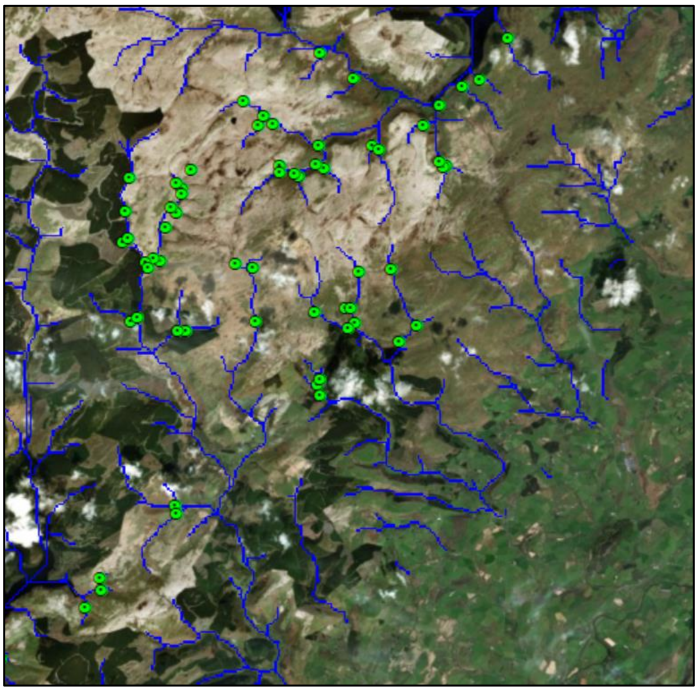


Figure 4. 6 Extracted streams (solid blue lines) from the digital elevation model and overlaid on the sample locations (solid green dots)

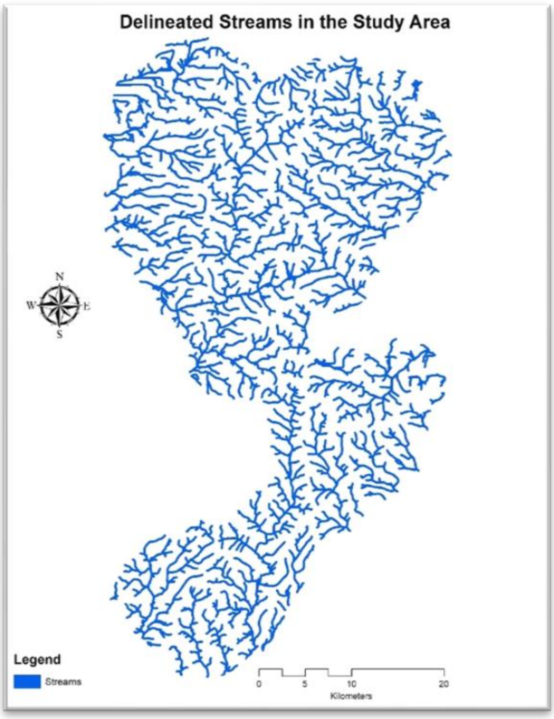


Figure 4. 7 The extracted streams in the study area

4.5.2. Criteria for comparison

The geochemical sample points were split into two groups named as the training and the test group with 70% and 30% of the points respectively. The criterion to split data was that the sample points located in the homogenous geological group and on the same streams were selected and separated from the training data as test data.

Out of the total 407 geochemical sample points, 285 were used to train the model, and the remaining 122 points were used to test the accuracy of the predicted surfaces. The root means square prediction error (RMSPE) statistics were used to assess the difference between predicted and the actual geochemical values (Robinson & Metternicht, 2006). The formula for RMSPE is given as the following Equation 4.22:

$$\mathrm{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{Z(x_i) - \hat{Z}(x_i)\right\}^2} \qquad \text{Equation 4. 22}$$

where $\hat{Z}(x_i)$ is the predicted geochemical value, $Z(x_i)$ is the geochemical observed (known) value, *N* is the total number of samples. Ideally, the value of RMSPE should be zero if there is no error in prediction; however, a low RMSPE suggest less error in the predicted surface and vice-versa.

4.5.3. Spatial interpolation and interpretation

The predicted/interpolated geochemical surfaces developed by IDW were controlled and changed by varying the two IDW parameters, i.e. the power function and the number of sample points used in the prediction. Starting with the lower power (α) as 1 to a higher power of 3 were tested, whereas the number of the nearest sample points were varied from 5 to 15 with increments of 5 at each step. The other parameters, like sector type, which is the directional search for the inclusion of several points were kept fixed, i.e. four sectors with a 45° offset, angle 0 and auto calculation of semi-major and minor axis.

For an initial prediction, the model was set with low power (α) as 1 and a maximum of 5 neighbor points were selected. The number of neighbor points was increased up to a maximum of 15 with the same power. The same procedure was repeated for power 2 and 3. The results showed that a single power factor and number of neighbors could not be used for all the geochemical variables, all the three power value and a different number of neighbors produced best geochemical prediction surfaces for different elements. The lowest root mean square prediction error (RMSPE) statistics (Table 4.2) were used to select the most appropriate combination model parameters for the geochemical elements. To discuss the effects of power and the number of neighbors Sn is taken as a case study, the prediction surfaces with all the variation of powers and neighbours are shown in Figure 4.8. The comparison of Sn predicted surface, as shown in Figure 4.8, revealed that areas with a low (near to zero) concentration remain almost the same in all the model types, but the area of medium and high concentration varies with different models. The change is quite significant at the edges and in the middle, particularly the southern part of the map. The maps of all the predicted streams geochemical properties through the IDW models based on the lowest RMSPE are given in Figure 4.9.

The results of geochemical properties of stream sediments showed that Au was present in the lower streams of the study area towards the South and South-West whereas no or very low gold concentration was found in the Northern part of the study area. There were only traces of Gold present in the study area, and distinct anomalies in the area could be associated with the alluvium dominated (containing several significant placer deposits) or metasedimentary rocks (Cohen et al., 1999). As, Fe, Ti, V and Zr had variable concentration and could be found throughout the streams present in the study area. The highest and above-average concentration of Ba was observed in the Eastern and Western streams to the North, respectively.

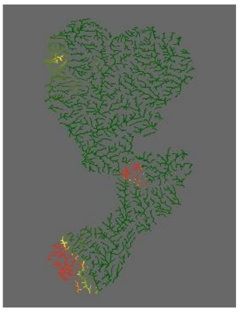
Table 4. 2 RMSPE of all the geochemical elements against different IDW models (Green highlights the combination of power and neighbouring points with the lowest RMSPE values for each element)

	IDW-P-1 (5/5)	**IDW-P-1 (10/5)**	**IDW-P-1 (15/5)**	**IDW-P-2 (5/5)**	**IDW-P-2 (10/5)**	**IDW-P-2 (15/5)**	**IDW-P-3 (5/5)**	**IDW-P-3 (10/5)**	**IDW-P-3 (15/5)**
Ti	935.306	930.288	928.567	982.25	970.99	966.118	1030.16	1022.93	1020.6
Mn	579.784	599.78	615.654	551.378	550.218	549.433	551.375	549.231	547.605
Fe	10176.4	10103.139	10131.154	10801.5	10658.5	10602.4	11411.1	11327.8	11301.3
V	34.497	34.064	34.213	36.225	35.652	35.54	38.311	37.98	37.918
Ni	41.582	42.233	42.779	38.871	38.621	38.51	38.632	38.43	38.328

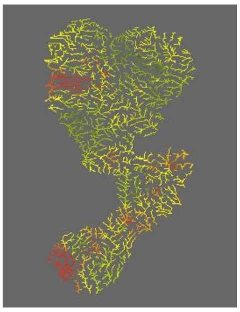
Cu	1624.36	1650.882	1669.518	1698.91	1683.29	1674.35	1799.99	1795.61	1793.09
Zn	2047.331	2063.756	2079.226	2130.86	2113.68	2105.43	2278.01	2269.98	2266.43
Zr	32.354	32.309	32.658	34.193	33.917	33.838	35.523	35.394	35.349
Sn	26.992	26.824	26.695	28.873	28.634	28.453	29.796	29.704	29.643
Sb	46.614	46.112	46.034	45.513	45.329	45.508	48.993	48.634	48.566
Ba	186.463	182.52	181.299	201.156	198.719	197.45	213.379	212.64	212.31
Ce	14657.94	14734.557	14827.495	15512.6	15383.3	15353.4	16499.4	16410.6	16394.1
Pb	2305.09	2322.459	2348.981	2324.85	2318.12	2320.98	2387.3	2381.34	2380.93

As	7.599	7.616	7.654	7.681	7.669	7.677	7.88	7.867	7.865
Au	525.482	513.607	509.136	585.032	580.134	578.308	619.582	618.648	618.407

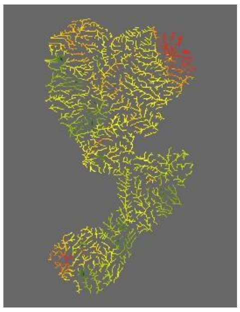
Figure 4. 8 Prediction maps of Tin (Sn) for different power and neighboring sample points using the IDW model



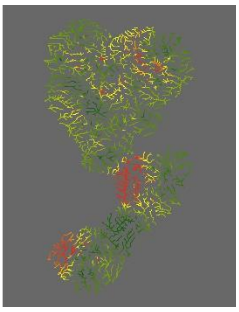
Au



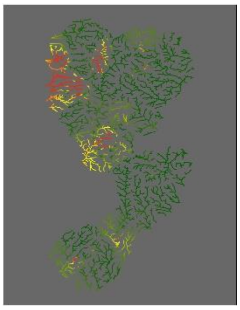
As



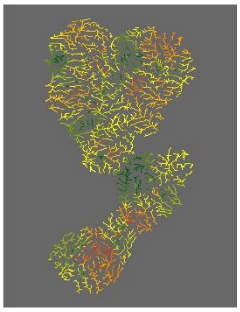
Ba



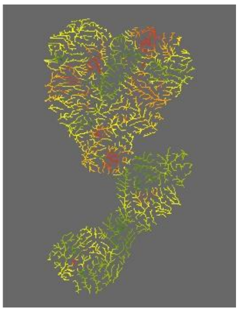
Ce



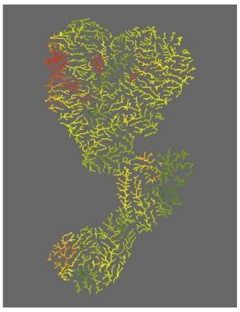
Cu



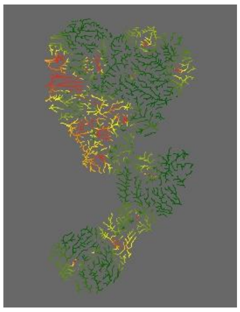
Fe



Mn



Ni



Pb

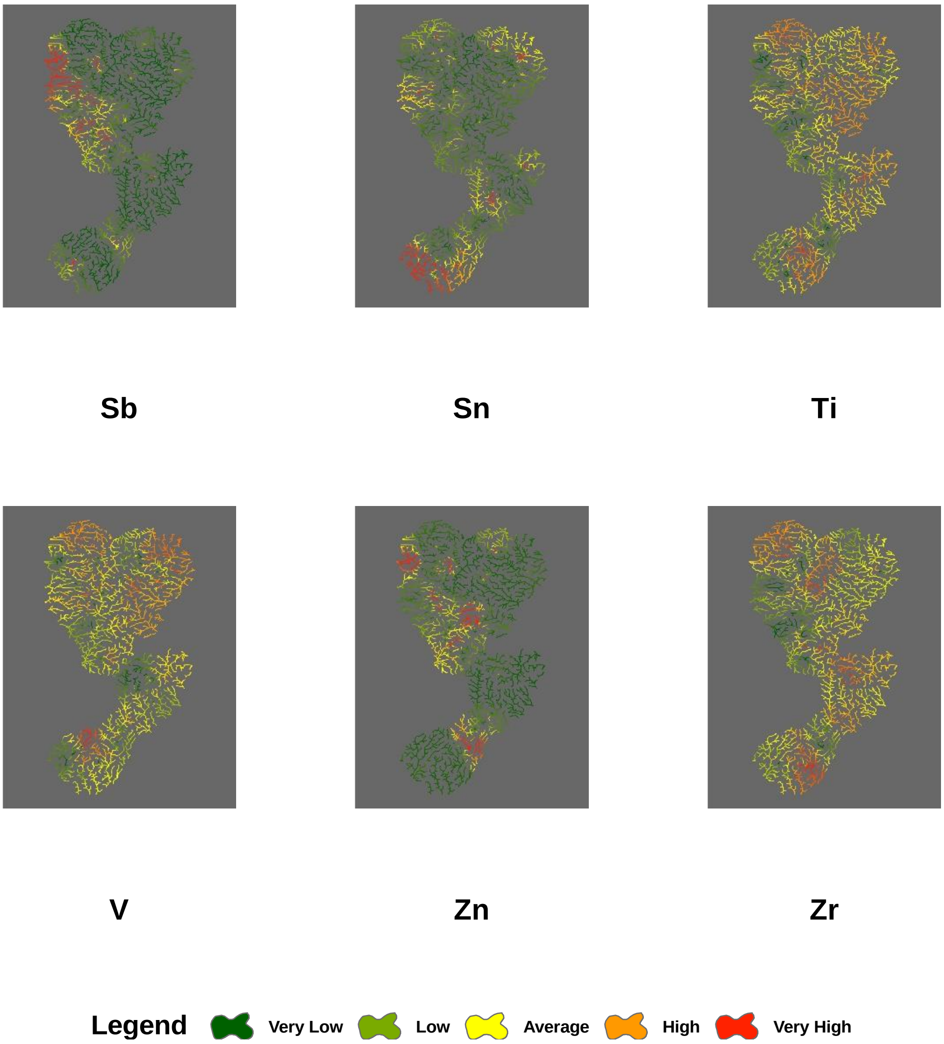


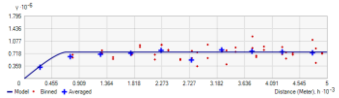
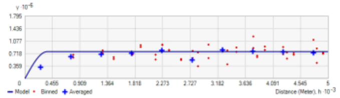
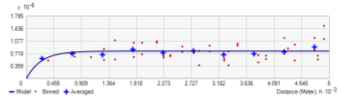
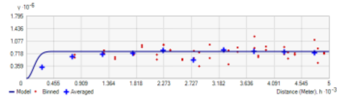
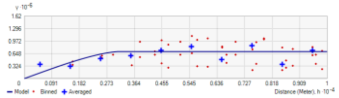
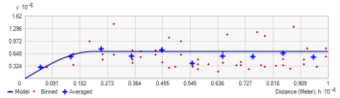
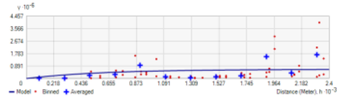
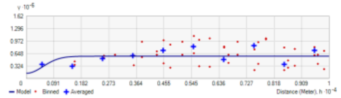
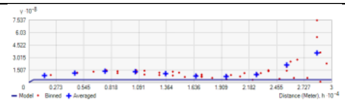
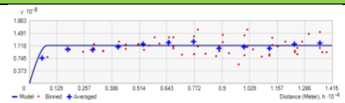
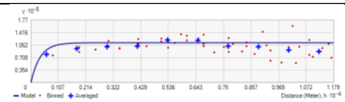
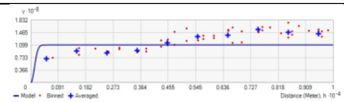
Figure 4. 9 Predicted stream surfaces of all the geochemical elements based on the lowest RMSPE of IDW model

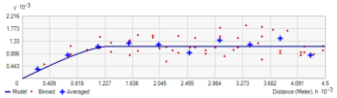
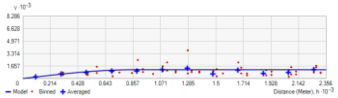
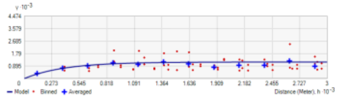
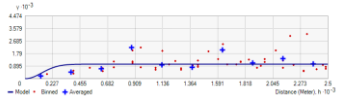
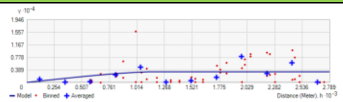
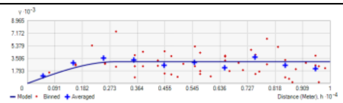
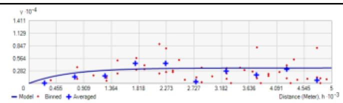
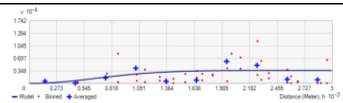
The high concentrations of Cu were mainly in the middle streams towards the North of the study area, whereas average values were found in the middle. Ce was significant in the lower streams of the study area towards the South with a little higher amount in the North-West region of the study area. Mn showed less variability and is mainly concentrated in the upper streams more towards the North. Ni and Pb were found in high concentration near the Eastern streams of the study area. Sb followed the same trend of concentration as Ni and Pb but in more streams. Sn was mainly dominated in the lower streams towards the South with medium and a few high quantities in the North-West and North-East of the study area. The highest concentration of Zn was observed in the North-Eastern and middle streams with average values in the lower streams of the study area.

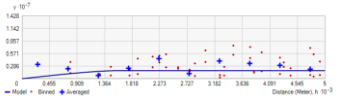
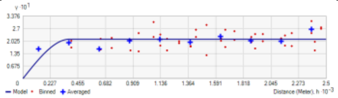
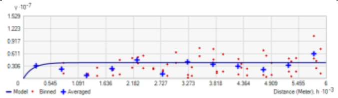
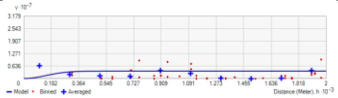
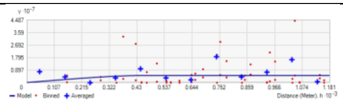
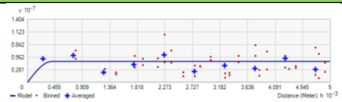
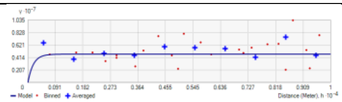
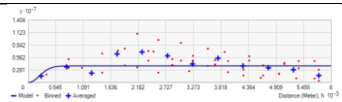
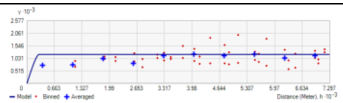
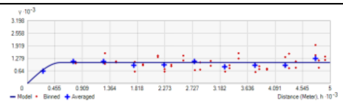
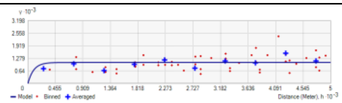
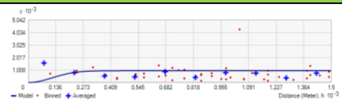
For Kriging, interpolation is the first and the most important thing to develop an empirical semivariogram model and then fitting that model as structural analysis in Kriging; this process is usually known as variography. The four most commonly applied theoretical semivariogram models circular, spherical, exponential and gaussian were tested against the streams sediments geochemical data. It was tried to make the semivariogram models fit the average of the binned data; also the directional variations were incorporated for micro-tuning of the models.

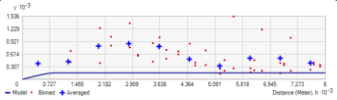
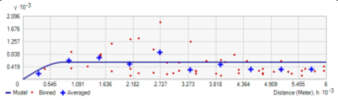
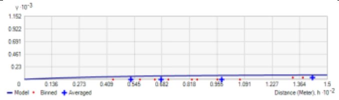
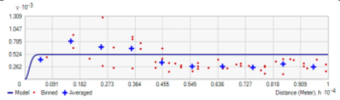
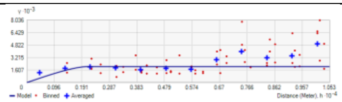
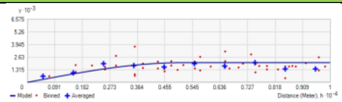
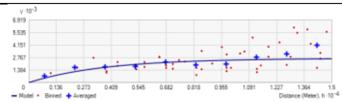
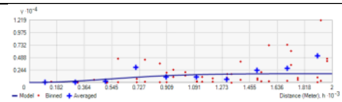
The selection of the most appropriate semivariogram model for a particular geochemical element was made based on the lowest yielded RMSPE by each model. The semivariogram of all the models, along with their RMSPE, is given in Figure 4.10.

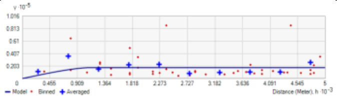
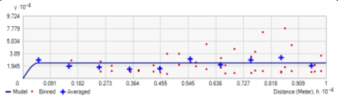
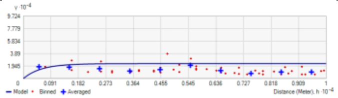
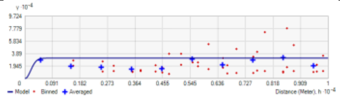
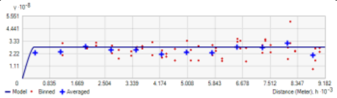
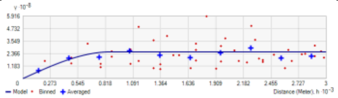
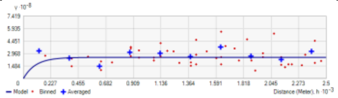
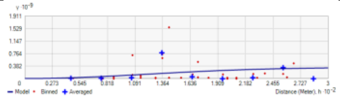
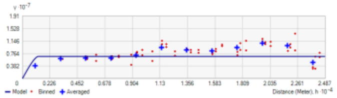
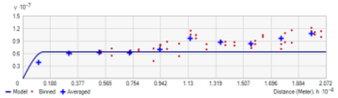
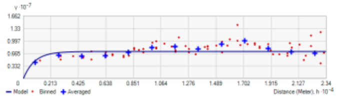
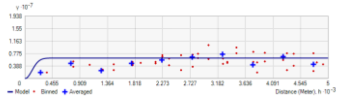
The analysis showed that out of the 15 streams sediments geochemical elements, only two can be better predicted with the circular semivariogram model, which is V and Ba. The spherical semivariogram model can be used for Mn, Ni, Cu, Zr, Sn, Ce, As and Au, whereas the exponential semivariogram model is suited best for Ti, Fe, Sb, and Pb. Only one geochemical element Zn can be best predicted by the gaussian semivariogram model. To discuss the effects of the semivariogram model on stream sediments, Sn was taken as a case study, the prediction surfaces with all the circular, spherical, exponential and gaussian semivariogram models are shown in Figure 4.11.

	Circular	Spherical	Exponential	Gaussian
Ti				
	910.090	908.099	906.227	908.121
Mn				
	530.237	525.452	535.214	529.565
Fe				

	10123.668	10100.240	10098.145	10107.177
V				
	33.372	33.960	34.549	33.749
Ni				
	38.319	35.302	36.709	35.930

Cu				
	1641.324	1605.028	1614.670	1620.419
Zn				
	2035.104	2030.435	2033.545	2028.555
Zr				

	32.259	31.704	31.985	32.444
Sn				
	26.346	25.286	28.331	27.197
Sb				
	49.526	45.621	43.955	46.515

Ba				
	178.687	182.180	179.187	183.357
Ce				
	14679.096	14632.843	14671.989	14778.979
Pb				
	2308.287	2311.019	2304.180	2316.875

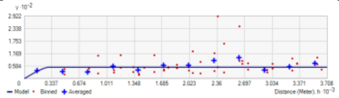
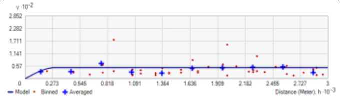
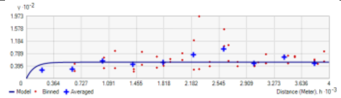
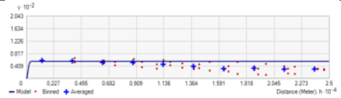
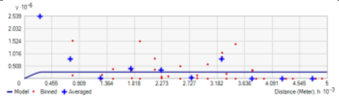
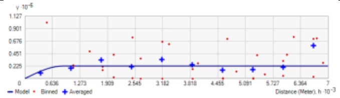
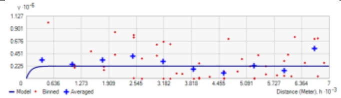
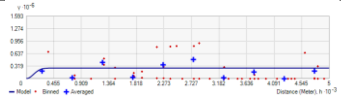
As				
	7.714	7.474	7.922	8.013
Au				
	521.306	507.338	514.021	510.830

Figure 4. 10 Semivariogram models of all the geochemical elements along with different RMSPE

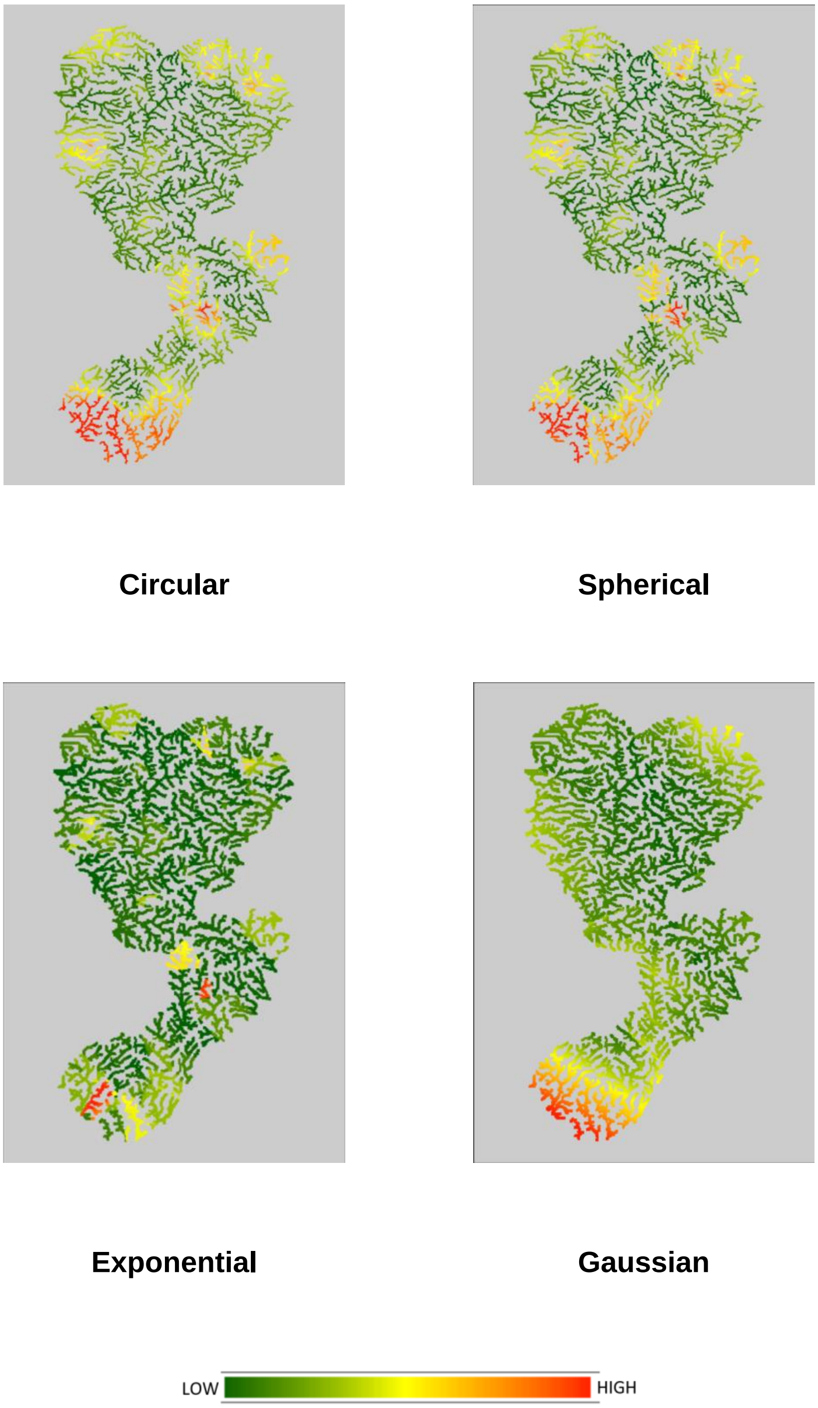
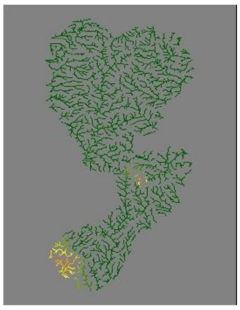
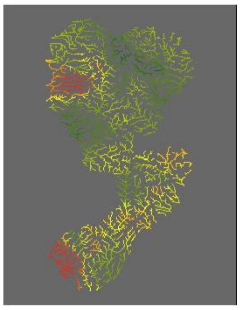
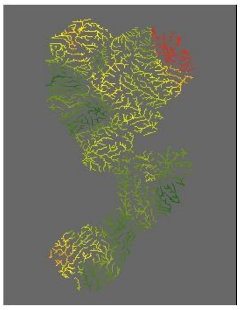


Figure 4. 11 Prediction maps of Tin (Sn) for different semivariogram models used in Kriging interpolation

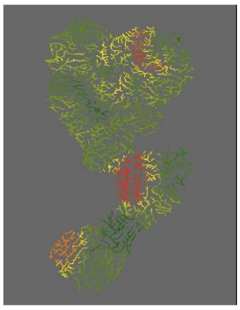
The comparison of Sn predicted surface through four semivariogram models, as given in Figure 4.11 showed that the spherical model is more accurate than the others as it produced more continues geochemical values in streams. The same is supported by the semivariogram model of Sn, as mentioned in the Figure 4.10. The circular model also showed some continuity but is not as accurate as of the spherical. This continuity might be due to the circular semivariogram model, which shows a little variation in the start and then levels off. The exponential semivariogram model of Sn suggested abrupt changes in the geochemical values and the same is highlighted in the predicted surface. The gaussian model showed some variability but is unable to depict appropriately because most of the geochemical sample points were not fitted through the model. The change is quite significant at the edges in the North and the middle, particularly the Southern part of the map. The maps of all the predicted streams geochemical properties through the Kriging model based on the lowest RMSPE of the best-fitted semivariogram model is given in Figure 4.12. The trend of prediction surfaces by Kriging is in line with the IDW results but significantly more realistic and continues.

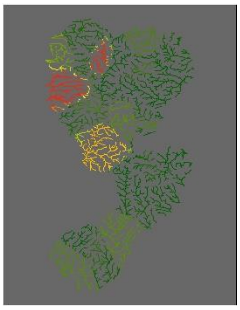


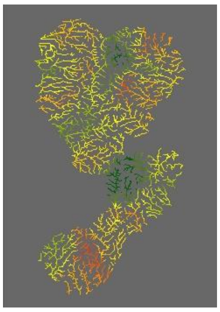




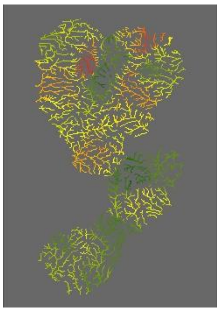
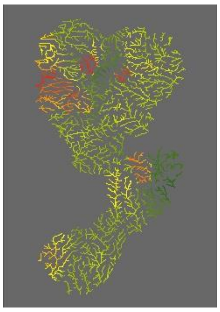
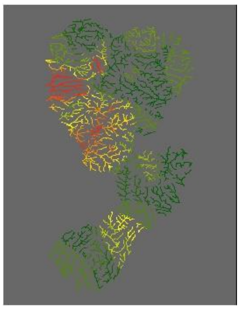
Au **As** **Ba**



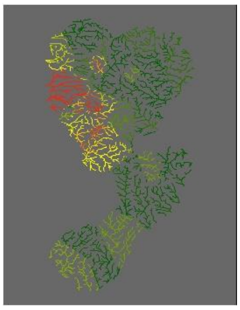
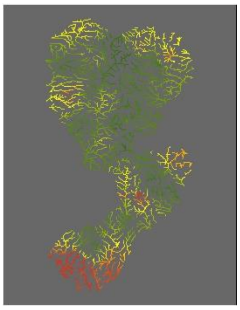
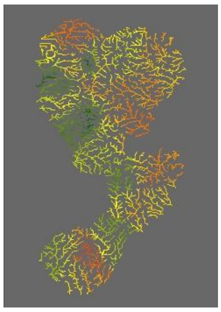




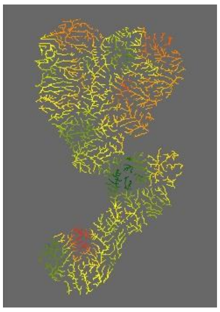
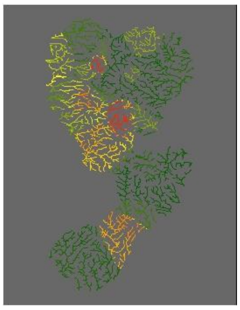
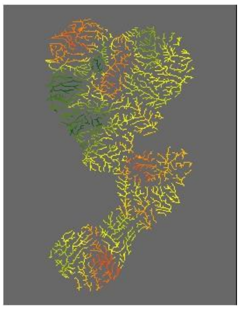
Ce Cu Fe

Mn Ni Pb

Sb Sn Ti

V Zn Zr

Legend Very Low Low Average High Very High

Figure 4. 12 Predicted stream surfaces of all the geochemical elements based on the lowest RMSPE of semivariogram models using Kriging

4.5.4. Potential Mineral Prospects

The multivariate overlay analysis for both IDW and Krigigng predicted surfaces were carried out in order to develop the Geochemical Accumulation Index (GAI) of the study area, as shown in Figure 4.13.

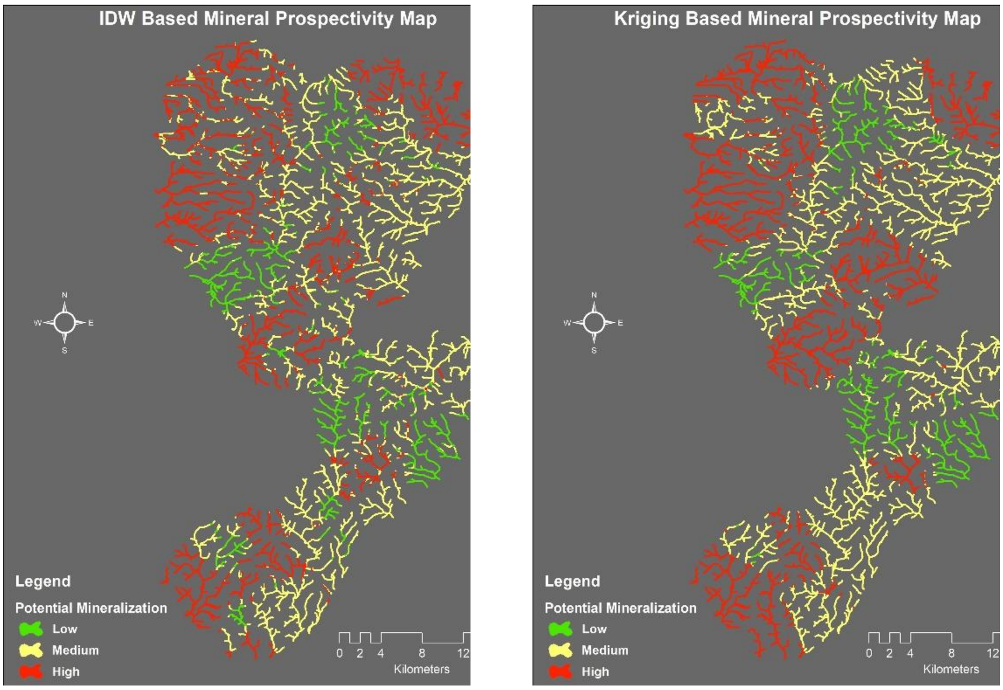


Figure 4. 13 Mineral enriched streams extracted through Geochemical Accumulation Index (GAI)

The higher values of GAI in a stream represents the higher chances of mineralization at the origin and surrounding regions and vice versa. The results showed that most of the mineral enriched streams were found in the North-West, North-East and Southern parts of the study area. By tracing these mineral enriched streams, the potential mineral prospects can also be mapped accordingly and with less time and cost. The potential mineral prospects, as identified in this research area, were also confirmed with the existing mining activities in the regions, as shown in Figure 4.14. This showed that the methodology adopted in this research study could successfully be applied to identify the potential mineral prospects in a region using stream sedimentation.

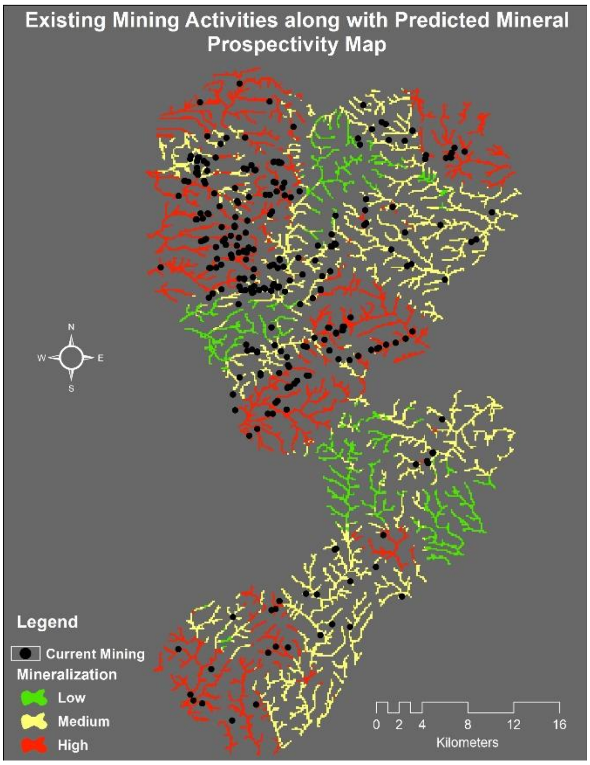


Figure 4. 14 Existing mining activities overlaid on the GAI based mineral enriched streams

4.5.5. Comparison of prediction models

In order to choose the best prediction surfaces, the root means square prediction error (RMSPE) of each geochemical element for both IDW and Kriging models were compared (Table 4.3).

Table 4. 3 Semivariogram models of all the geochemical elements along with different RMSPE

Geochemical Element	IDW	Kriging	Difference

As	7.599	7.474	0.125
Au	509.136	507.338	1.798
Ba	181.299	178.687	2.612
Ce	14657.941	14632.843	25.098
Cu	1624.36	1605.028	19.332
Fe	10103.139	10098.145	4.994
Mn	547.605	525.452	22.153
Ni	38.328	35.302	3.026
Pb	2305.09	2304.18	0.91
Sb	45.329	43.955	1.374
Sn	26.695	25.286	1.409
Ti	928.567	906.227	22.34
V	34.064	33.372	0.692
Zn	2047.331	2028.555	18.776
Zr	32.309	31.704	0.605

The highest and the lowest difference between IDW and Kriging model were found among Ce and As respectively. For Ce and As, the IDW model with power as 1 and number of nearest sample points as 5 gives the better prediction surface with the lowest RMSPE whereas as per Kriging, the spherical model generates the lowest RMSPE which is also lower than the IDW model. Overall, the Kriging model yields a low RMSPE because it incorporates the spatial correlation (autocorrelation) of the sampled streams geochemical values along with their locations. Several researchers also concluded that the Kriging model is better for spatial predictions than the IDW model. For example, Shahbeik et al. (2014) applied IDW and Kriging for mineral ore prediction, and it was concluded that the Kriging performs much better, and the results were reliable. Similarly, another study conducted by Mueller et al. (2004) showed that Kriging is more efficient in the prediction of soil fertility than IDW. Also, Goovaerts (2000) stated that Kriging yields more accurate rainfall predictions with the lowest root mean square error.

4.6. Conclusion

This study has shown a novel approach for prediction of geochemical properties of stream sediments using Geostatistical interpolation models, including IDW and Kriging. Instead of generating over-extended predicted surface models that most likely are outside the mineralization range from the stream samples, the new proposed geospatial characterizations were modelled strictly based on the geo-profile of each stream using digital elevation 3D models. The predicted geochemical value distribution at each stream showed an overall improved spatial characterization that can be used to accurately trace back the key minerals within each of the streams. The study showed that the two spatial prediction models used, IDW and Kriging methods should be used under a combinatory approach in order to predict the spatial distribution of geochemical elements better. Overall, spatial realizations using IDW with a low power function tended to produce more accurate results probably due to the relatively high variability among the geochemical sample data (Robinson & Metternicht, 2006) (Kravchenko & Bullock, 1999). The Kriging method also showed accurate results as well, specifically when using the spherical semivariogram model and then exponential models. The mineral prospectivity map based on the Geochemical Accumulation Index (GAI) showed that most of the mineral enriched streams were located in the North-West, North-East and Southern parts of Central Wales, the same was also confirmed with the existing mining activities in the region.

The statistics ‘root mean square prediction error’ (RMSPE) was used as the primary indicator to assess the estimation accuracy of both methods. The resulting models with the lower RMSPE values were considered for further analysis. One of the critical conclusions resulting from this study is that IDW and Kriging results should not be considered independently; instead, a combinatory model should be produced in the prediction of the geochemical elements for stream sediments. Further research should investigate if other geo-environmental factors such as proximity to the sea, structural geology, geophysical conditions, and terrain slope angle, can be incorporated into the characterization model.

5. AUTOMATIC EXTRACTION OF GEOLOGICAL LINEAMENTS IN ASSOCIATION WITH MINERALIZATION USING REMOTE SENSING AND GEOGRAPHICAL INFORMATION SYSTEM

This chapter is based on the following publication

Mahboob, M. A., Genc, B., & Atif, I. (2019, April). Automatic extraction of geological lineaments in association with mineralization using remote sensing and geographical information system. In *2019 Proceedings of the 26th International Mining Congress and Exhibition of Turkey, Antalya,* (pp. 285-289).

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

5.1. Introduction

In remotely sensed satellite images, generally, the lineaments are the linear or somewhat-curved structures that can appear as brighter or darker features than the surroundings. These lineaments are usually natural geological faults, fractures, joint zones, huge crustal fractures, geomorphological features, fold limbs as well as deep and buried faults (Vassilas et al., 2002). Identification and extraction of these lineaments are of great importance to get the geological information of any region particularly for mining and petroleum exploration (Marghany & Hashim, 2010; A. B. Pour & M. J. O. G. R. Hashim, 2015) . The more and forever increasing demand of several minerals is forcing the mineral industry to update and renew the exploration in the existing production areas, as well as to explore new, remotely and un-surveyed areas of the earth (Alvarez-Berríos & Aide, 2015; Hörfarter et al., 2019; O'leary et al., 1976). But, due to time and cost constraints, there is a dire need to develop new fast, robust and reliable exploration techniques to survey relatively large areas of earth usually hundreds of thousands of square kilometres at low cost and more confidence (Pour et al., 2017). In recent era of mineral exploration, a specific and strong relationship between the known mineral deposits and different geological lineaments, such as faults, fractures etc. has been developed and used several mineral exploration companies, especially for the regions where geology is known but data of mineralogical deposits is lacking (Govil et al., 2018). The lineaments data are usually used as the first-hand information and indirect indicator of mining potential of an area (Meshkani et al., 2013; Pour et al., 2018b; Solano-Acosta et al., 2007). Hence, in-depth research on lineament extraction is of great theoretical importance and practical implication.

The old classical methods of lineaments mapping are based on visual interpretations or semi-automatic approaches, but they are compromised on accuracy as well as time-consuming (Masoud & Koike, 2006). Recently, advances in the remote sensing processes and computer vision i.e., digital image processing, the extraction and characterization of lineaments from the earth surface through automatic extraction models are quite accurate and time-saving (Farahbakhsh et al., 2018; Yeomans et al., 2019). There are several algorithms developed by different researchers like Segment Tracing Algorithm (STA) (Koike et al., 1995), the Hough Transform (HT) (Kröger et al., 2016), hydrological analysis and automatic Lineament Extraction (LINE) model (Salui, 2018) etc. Usually the principal of most methods is the detection of edges based on various spatial filters. However, due to the presence of noise, cloud, the height of the satellite, illumination conditions, and others, some satellite images may give a “pseudo edges”. These falsely detected lineaments should be removed in accuracy assessment methods. In this research freely available Landsat – 8 satellite datasets were used to map the lineaments at Waziristan region situated in North-West of Pakistan. While previous research studies have utilised the Digital Elevation Model derived from the Shuttle Radar Topography Mission (SRTM) for the purpose of lineament extraction, it has been observed that Landsat-8 is also extremely effective in this regard. Landsat-8 and the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) serve different purposes, however, when considering the extraction of lineaments, Landsat-8 exhibits a significant advantage. The SRTM DEM dataset provides comprehensive topographical information, including accurate elevation measurements. On the other hand, Landsat-8 satellite imagery possesses valuable multispectral capabilities that are essential for distinguishing different surface materials and geological variations. The utilisation of a multispectral range facilitates a refined and detailed discernment of geological formations, including faults and fractures. Furthermore, the utilisation of Landsat-8's spatial resolution of 30 metres, in conjunction with its spectral bands, enables a more comprehensive and precise delineation of lineaments compared to relying solely on elevation data. Therefore, in terms of comprehensive lineament extraction, Landsat-8's capabilities are better suited and more effective compared to the topographical focus of SRTM DEM. Several research studies (Céleste et al., 2021; Es-Sabbar et al., 2020; Redouane et al., 2022) have utilised Landsat-8 satellite data for the purpose of extracting lineaments, resulting in highly effective and accurate results.

5.2. Materials and Methods

5.2.1. Study area

The research was conducted in the Waziristan tribal region, situated in the North-West of Pakistan. The region primarily consists of undulating topography, which is divided into two administrative districts known as North Waziristan and South Waziristan (Figure 5.1). The entire geology of the study is complex and consisted of rocks, which are generally deeply thrust, folded, faulted, fractured, brecciated, and granulated (Ahmad, 2016). Waziristan is a mineral-rich area of Pakistan and important from an economic point of view as huge resources of copper, manganese and chromite deposits found there.

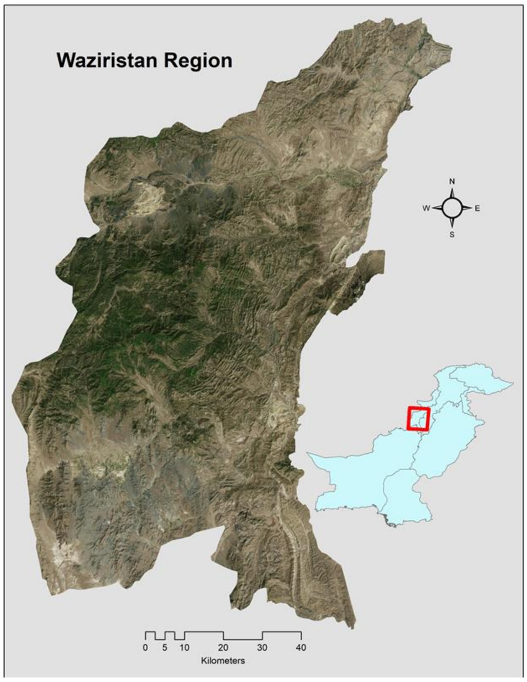


Figure 5. 1 The map of the study area

Besides its economic importance, this area of the world is little known to the world and had very poor information on geology and geography. Due to its complex topography and inaccessibility, mapping the geological structures through remotely sensed satellite data is the only available solution to get enhanced information of the region.

5.2.2. Methodological flowchart

The methodological flowchart of the study area is given in Figure 5.2.

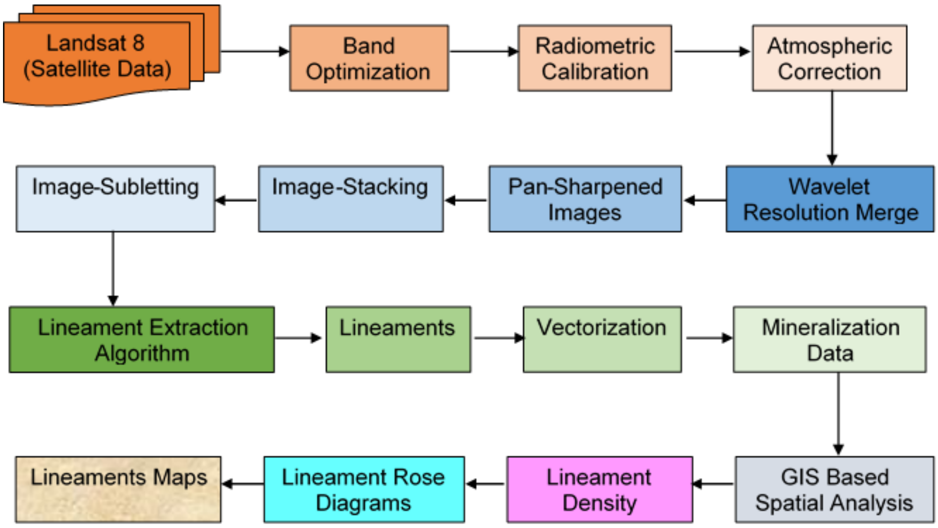


Figure 5. 2 The comprehensive flowchart of the research work, highlighting the important tasks applied

5.2.3. Satellite data

Open-source Landsat 8 satellite data was used in this research. Landsat 8 has 11 spectral bands with a spatial resolution of 30 and 100 meters (except the panchromatic band whose resolution is 15m). The division of bands is with respect to two sensors mounted of Landsat 8. The first sensor is Operational Land Imager (OLI) consisting of total 9 bands including 1 cirrus for clouds, 4 in visible, 1 in near infrared, 2 in middle infrared and 1 panchromatic band. The second sensor is a

Thermal Infrared Sensor (TIRS) including 2 bands in the thermal portion of the electromagnetic spectrum. Another important component of Landsat 8 data is its 16 bits of high radiometric resolution (Vermote et al., 2016).

The study area falls in two Landsat image tiles with the path-row specification of 152-037 and 152-038. The cloud-free and clear weather images of 27 December 2018 were downloaded and pre-processed for further analysis. First, the image stacking was done in ERDAS Imagine software whereas, spectral band 2 and 9 were not used in image stacking process because they are usually used for atmospheric aerosol, and cloud detections respectively (Adiri et al., 2017). As spatial resolution along with spectral bands play a key role in the extraction of lineaments (Mwaniki et al., 2015), so panchromatic band of 15 meters was used to enhance the spatial resolution of other bands using Wavelet Resolution Merge algorithm available in ERDAS Imagine software.

5.2.4. Wavelet resolution merge

Wavelet resolution merge is a powerful technique used to enhance the spatial resolution of multi-spectral satellite images. Usually low or medium spatial resolution data is improved by using a spatially referenced high resolution image (Pushparaj & Hegde, 2017). This algorithm employs a wavelet transformation, in which a relatively high-resolution satellite image is divided into high and low frequency bands. After transformation, the low-quality image is then substituted by the low-quality satellite bands respectively (Zhang & Mishra, 2012). This algorithm is very effective to enhance the resolution of Landsat and IKONOS satellite images. But, apparent modification of the spectral properties of the resultant images are an unavoidable problem of wavelet-based image enhancement techniques.

5.2.5. Lineament extraction

The automatic extraction of lineaments was done using computer vision techniques. This technique is a two-step approach: at first, the edges (contours) were detected from the satellite images based on spatial filtering. In a second step the abrupt changes in the region were mapped based on the values of neighbouring pixels, usually denoting the lines (linear features) (Corgne et al., 2010; Li, 2010; Nforba et al., 2019; Sharifi et al., 2018).

5.2.6. Edge detection

The edges present in the Landsat satellite image were detected using filter radius (RADI) for image pixels and the threshold value for edge gradient (GTHR) (Ahmed, 2014). As per literature, for RADI the values range from 3 to 8 were applied to minimize the noise factor. Similarly, the values from 10 to 70 were taken as the threshold in contours detection for the gradient.

5.2.7. Lines detection

For the detection of linear features, three different types of threshold values were applied. First LTHR which specifies the minimum length of the curve (in pixels), a value of 10 was applied in this study. Second FTHR which specifies the maximum error (in pixels) i.e., the error of fitting a line in pixel curve. Usually, the lower values of FTHR yields the better fitting but shorter sections. Third ATHR which is the threshold for angular differences, it is the maximum angle (in degrees) used to link the two lines as one (Adiri et al., 2017).

5.2.8. Slope

The slope of the area plays an important role in the lineament studies which can be helpful in the determination of their presence. An abrupt change in the values of the slope is usually key indicators of the existence of linear structures like lineaments (Bancalà et al., 2017). The slope was calculated from the digital elevation model (DEM) of the study area using ArcGIS spatial analyst extension.

5.3. Results and Discussion

In this research, the geological liniments of the study area were extracted using the most optimized enhanced resolution bands of Landsat 8 satellite data. The image pre-processing was done and automatic lineament extraction algorithm was applied. The results showed that the liniments mapped from the satellite image are more detailed and enhanced as compared to the filed data.

5.3.1. Mineralization zone and fault lines

The mineralization zone was mapped by the filed investigation. The high content of copper mineral is present in the study area, which is also being mined from the site. According to Malkani (2013) an estimate of about 120 million tons of copper reserves has been found in the study area.

Beside that the geological features mainly the fault lines were also mapped through the filed visit and the published geological reports as shown in Figure 5.3. The results showed that the most of the fault lines are in the orientation from North-East to South and South-West of the study area.

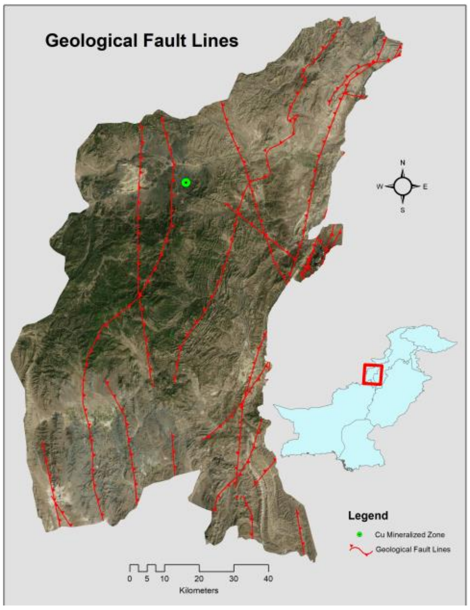


Figure 5. 3 The geological faults and the mineralized zone present in the study area (Source of Data: Geological Survey of Pakistan)

5.3.2. Hill shade modelling

The hill shade modelling of the study area was done using the ArcGIS spatial analyst extension and is very helpful in highlighting the potential linear features from the study area. The shaded relief map of the area was varied by considering different illumination source angles and shadow conditions. The azimuth value of 45 degrees was applied which highlights the maximum linear features present in an area (Biland et al., 2017). The default colour scheme for the shaded relief map is grey as shown in Figure 5.4.

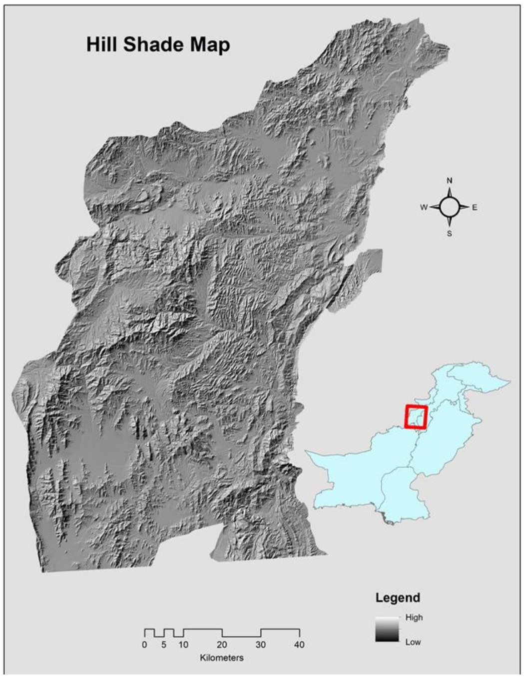


Figure 5. 4 The shaded relief map of extracted from digital elevation model of the study area

The high grey colour tones represent the area of abrupt change with respect to the topographical features. Usually, these are the areas where there is a high possibility of linear features to be situated in and around. A study conducted by Šilhavý et al. (2016) also found that hill shading is a very effective way to enhance the results of lineaments extraction from satellite data. Another study conducted by Abdullah et al. (2010) revealed that the hill shade of the digital elevation model improved the accuracy of the lineaments extraction.

5.3.3. Lineaments extraction

The lineaments were extracted using an automatic algorithm, the results (Figure 5.5) showed that the algorithm generated the detailed lineaments of the study area. A total of 3,980 small to large scale lineaments were identified in the study area, as there was no such definite pattern of the lineaments found but mostly, they were oriented in different direction. The lineaments were further overlaid on the geological fault lines and the mineralized zones of the study area.

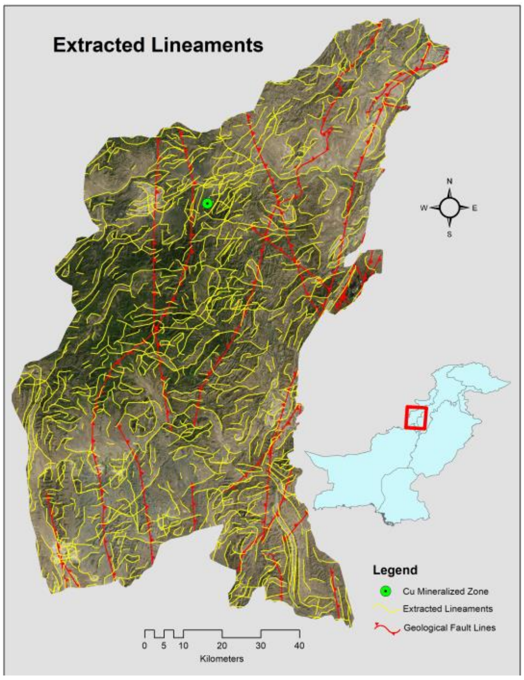


Figure 5. 5 Automatic extracted lineaments from remotely sensed Landsat satellite data as highlighted in yellow colour

For spatial variation and distribution pattern analysis of lineaments in relation to their association with regional structural anomalies, the azimuth based statistical tools were applied to the extracted lineaments (Ekneligoda et al., 2010). To further study the directional trends of extracted lineaments, the rose diagram type common tools were used (Ekneligoda et al., 2010; Masoud et al., 2017; Masoud & Koike, 2006; Zangerl et al., 2006). Figure 5.6 showed a rose diagram, in polar projected coordinates composed of azimuth length, orientation, and occurrence, of the strike plane at the distance of 10 degrees. It can be clearly seen from the Figure 5.6 that the main trends are North-East to South-West accompanied by East-West. The lineaments orientation was found to be in accordance with the other geological features as mapped during field investigation. Many hidden lineaments were identified which was very difficult and time taking to identify with naked eye and field investigation. The results showed that the lineaments extracted by the remotely sensed satellite images followed similar orientation, specifying that the geological features clearly control them.

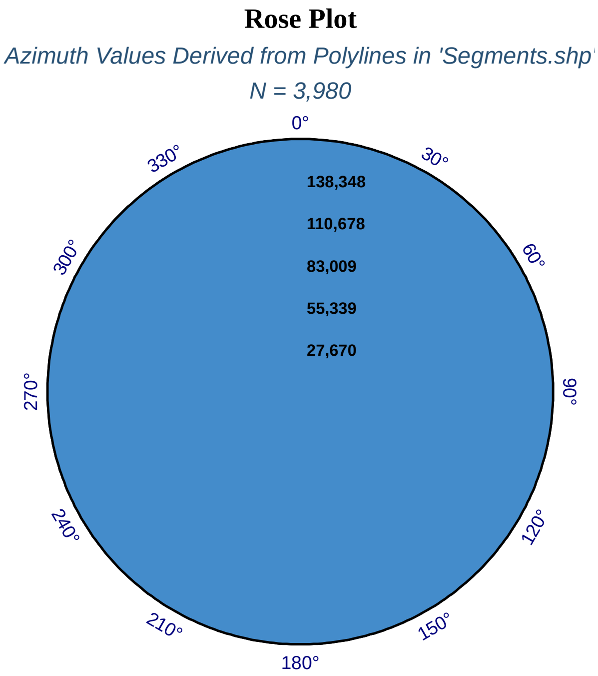


Figure 5. 6 The rose diagram of extracted lineaments highlighting the orientation

5.3.4. Lineament density and mineralization zone

The density analysis is very efficient statistical method for investigating the distribution of spatial density properties of lineaments, it can also reveal or provide information on deep complex geological features and mineralization (Corgne et al., 2010; Saepuloh et al., 2018; Sener et al., 2005). Generally, a highly dense region usually represents anomaly and can be considered as fault or fold expansion area, whereas a low dense region typically represents a comparatively stable tectonic block. High, medium, and low-density regions with a plane distribution have accordingly associations with certain mineralogical zones. The densest zones were found in the North of the study area very close to the actual copper mineralization zone as shown in Figure 5.7.

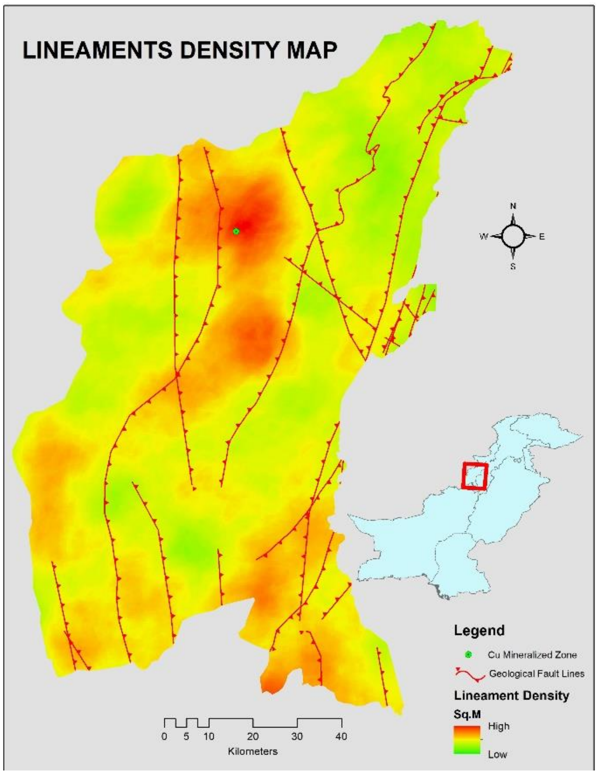


Figure 5. 7 The lineament density map in association with the mineralized rich zone *situated in the study area*

The results revealed that the copper zone may expand from North to the South and then shifted towards the West of the region. The same can also be observed by the orientation of the extracted liniments, as they follow similar trends. The other possible mineral-rich zone can be on the South-East of the study area as this area also has a high density of geological lineaments. The lineament density in the North side is higher than the Southside, and in the centre is higher than at ends. Furthermore, density analysis showed that the results of lineament extraction are in accordance with the geological phenomenon observed on site. Hence, the algorithm, for extraction of automatic liniments in accordance with mineralization, was proven appropriate for extracting lineaments in this region.

5.4. Conclusion

Based on Landsat 8 OLI remote sensing satellite images, fast extraction of geologic lineaments in a remote area Waziristan of Pakistan was successfully done using GIS based algorithm. To avoid unimportant satellite band noise disturbing the final lineament extraction results, the most appropriate remote sensing image bands were selected. The optimized bands were then enhanced in terms of their spatial resolution i.e., from 30 meter to 15 meters. The geographical information-based lineament extraction algorithm was applied and proved to be very effective for automatic extraction of lineaments. The orientation of the lineaments was in an accordance with the geosocial settings of the region and is mostly from North-East to South-West accompanied by East-West directions. The density map of the lineaments also highlights the significant orientation of potential mineralized zone in the region along with the identification of possible new mineralized pocket in the South-East of the region.

Hence, the techniques adopted in this research work has proven to be of great effectiveness for automatic extraction of lineaments and can be applicable in other areas where limited data is available. Beside the quick identification and mapping of possible mineralized zones in the region, the extracted lineaments information can also be used in further area of research like geological hazards, earthquakes, and groundwater recharge zones mapping. Simultaneously, the issues of lineament continuation and computational productivity, particularly the data processing power in large areas and with big amount of satellite image data, are also significant research area for further investigation.

6. MAPPING HYDROTHERMAL MINERALS USING REMOTELY SENSED REFLECTANCE SPECTROSCOPY DATA FROM LANDSAT

This chapter is based on the following publication

Mahboob, M. A., Genc, B., Celik, T., Ali, S., & Atif, I. (2019). Mapping hydrothermal minerals using remotely sensed reflectance spectroscopy data from Landsat. *Journal of the Southern African Institute of Mining and Metallurgy*, *119*(3), 279-289.

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

6.1. Introduction

Many developing nations depend on exploration and exploitation of mineral resources to sustain their economic growth. Usually, the traditional mineral exploration techniques require enormous finances, prolonged time, and tremendous manpower particularly in areas that are not easily reachable (Maduaka, 2014). Furthermore, mineral exploration required state of the art techniques and expertise along with geological, geochemical and geophysical datasets and most of the time these datasets are not easily available or may be missing where access is problematic (Bemis et al., 2014; Kaiser et al., 2002). Besides these, the modern remote sensing technology has also been proved to be one of the highly efficient and robust techniques used for mineral exploration. The use of remote sensing satellite images for geological mapping and mineral exploration usually involves studying the physicochemical properties of rocks and weathering soils, such as mineralogy, landforms, geochemical signatures and the spatially distributed lineaments (Bhattacharya et al., 2012). A fundamental principle for exploration of minerals is that it's quite possible that the undiscovered minerals will be located in the close vicinity of the discovered ones. For example, if mining is happening in one area, then similar minerals will be more likely found nearer to the discovered mineral, and as the distance increases, the likelihood of this will decrease. In that situation, before drilling exploratory boreholes at new locations, remote sensing images can be used effectively to identify regions with higher chances of minerals, mainly through multi- or hyper-spectral remote sensing images (Ciampalini, Garfagnoli, Antonielli, et al., 2013; Gholami et al., 2012). The use of reflectance spectroscopic information derived from remote sensing data allows effective localization of mineral exploration as well as reducing the cost and time spent during prospecting fieldwork including geological, geophysical, and geochemical studies of an area (Marjoribanks, 2010; Short & Lowman Jr, 1973; Tedesco, 2012). Several remote sensing studies for mineral exploration and lithological mapping have been done in arid and semi-arid regions (Hajaj et al., 2023; Lu et al., 2022; Tamani et al., 2019). With large exposures of geologic materials, satellites in space are capable of acquisition of spectral reflectance directly from rock or/and soils (Di Tommaso & Rubinstein, 2007; M. A. Mahboob et al., 2015; Pour & Hashim, 2012; Sabins, 1999; Zhang et al., 2007).

Hydrothermal alteration minerals with diagnostic spectral absorption properties in the visible and near-infrared through the shortwave length infrared regions can be identified by multispectral and hyperspectral remote sensing data as a tool for the initial stages of porphyry copper and epithermal gold exploration (Di Tommaso & Rubinstein, 2007; Ramakrishnan & Bharti, 2015; Zhang et al., 2007). In general, porphyry copper deposits are developed by the mineralogical alteration of the rocks caused by the hydrothermal fluids. These altered rocks can be identified through spectral characteristics of visible and infrared wavelengths (Pour & Hashim, 2012). Many minerals have unique and indicative spectral properties with a specific amount of electromagnetic (EM) energy reflected and/or absorbed at a particular wavelength, which can be used to identify them with high confidence. The portion of the electromagnetic spectrum from 0.4 to ~2.5 μm usually known as visible, near-infrared and shortwave-infrared is useful to sense the features with moderate and low-temperature properties because most of the sunlight is reflected from features in this portion (Mahboob et al., 2015) .

Usually, Iron oxides, oxyhydroxides, and ligands can be mapped very well in this range of the electromagnetic spectrum because of their high or low-temperature alterations (Clark et al., 1990). This portion of the spectrum can also be used for differentiation between silicate (clay) minerals from other features. This spectral differentiation of minerals has been the basis for the use of this technique in economic mineral exploration (Calvin et al., 2015). The thermal-infrared (TIR) portion of the spectrum usually from 7 to 14 µm, senses the emitted energy of Earth's surface. In addition to water, carbonate, and sulphate, this region of the electromagnetic spectrum is also sensitive to Si-O bonds in silicates (Manley, 2014; Repacholi, 2012; Salisbury et al., 1991; Udvardi et al., 2017). The spectral signatures of some typical minerals (Calcite, Orthoclase Feldspar, Kaolinite, Montmorillonite, and Hematite) that can be clearly and confidently mapped using reflectance spectroscopy of the satellite data i.e., hyperspectral data is shown in Figure 6.1.

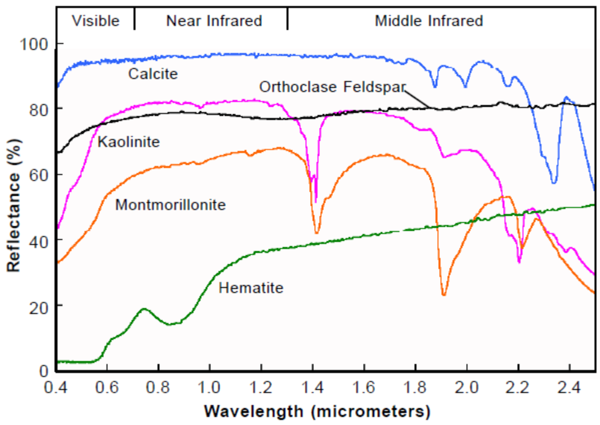


Figure 6. 1 Reflectance spectra of some typical minerals in visible, near-infrared and middle-infrared regions of the electromagnetic spectrum, Source: (Gale, 2017)

In this study, the identification of hydrothermally altered rocks and associated features with hydrothermal mineralization in Gauteng and Mpumalanga is examined using Landsat 8 (originally known as Landsat Data Continuity Mission (LDCM)) remote sensing satellite reflectance spectroscopy.

6.2. Material and Methods

6.2.1. Study area

The study area for this research was Roodepoort & Westonaria from Gauteng province and Witbank & Kriel from Mpumalanga province of South Africa as shown in Figure 6.2. Gauteng's name is derived from the Sotho, "Gauta" which means

"Gold" with the local suffix "-eng". "Gauta" has been taken from the Dutch word for gold, "goud". The main economic sectors are financial services, business services, logistics, communications, and mining.

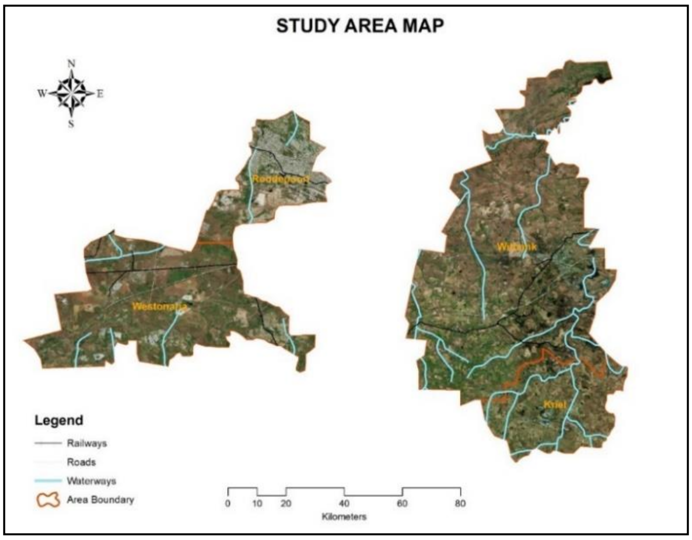


Figure 6. 2 Map of Gauteng and Mpumalanga provinces of South Africa

The general geology of Mpumalanga consists of mud-rock, siltstones, sandstones, conglomerate and several coal seams. However, in Gauteng, the most significant geological formation is Witwatersrand. Gold in this region was deposited by the greenstones (volcanic materials) and mountains of granites as the region was highly active geologically and erosion caused the formation and deformation of rocks. Due to these activates, the gold in the region became highly concentrated in the region. Gauteng was built upon the wealth of gold found deep underground, i.e. almost 40% of the world’s reserves (Durand, 2012). Whereas Mpumalanga (name changed from Eastern Transvaal on 24 August 1995) is another mineral-rich province of South Africa. Mpumalanga accounts for 83% of South Africa's coal production. 90% of South Africa's coal consumption is used for electricity generation and the synthetic fuel industry (Dabrowski et al., 2008).

6.2.2. Remote sensing data

Usually, ASTER is the most commonly used satellite data for hydrothermal mineral mapping and exploration. However, since 2008, its 6 x SWIR sensors are no more operational because of some malfunctioning (Wessels et al., 2013), the same has also been declared by NASA in 2009. On the other hand, Landsat is also free and used in mapping and exploration of hydrothermal altered minerals and rocks. In this research, the cloud-free level 1T (L1T) data of Landsat 8 satellite (path 170 / row 78) recorded in August 2017 was used. Landsat images are processed in units of absolute radiance using 32-bit floating-point calculations. These values are then converted to 16-bit integer values in the finished Level 1 product (Chander & Markham, 2003). Landsat 8 is equipped with the Operational Land Imager (OLI) and the Thermal-Infrared Sensor (TIRS) as their spatial and spectral characteristics are shown in Table 6.1. A single image of Landsat 8 satellite covers an area with extents of 170 km (north-south) by 183 km (east-west).

Table 6. 1 Landsat-8 satellite spectral and spatial details

Spectral band number	**Spectral band name**	**Wavelength (μm)**	**Spatial resolution (m)**
1	Coastal aerosol	0.43 - 0.45	30
2	Blue	0.45 - 0.51	30
3	Green	0.53 - 0.59	30
4	Red	0.64 - 0.67	30
5	Near-Infrared (NIR)	0.85 - 0.88	30
6	Shortwave-Infrared (SWIR) 1	1.57 - 1.65	30
7	Shortwave-Infrared (SWIR) 2	2.11 - 2.29	30
8	Panchromatic	0.50 - 0.68	15
9	Cirrus	1.36 - 1.38	30
10	Thermal-Infrared (TIRS) 1	10.60 - 11.19	100
11	Thermal-Infrared (TIRS) 2	11.50 - 12.51	100

6.2.3. Satellite image pre-processing

The pre-processing of raw satellite images is necessary to obtain geometrically and atmospherically corrected images so that the spectral information can be extracted and analysed. The raw satellite images of the study area are shown in Figure 6.3.

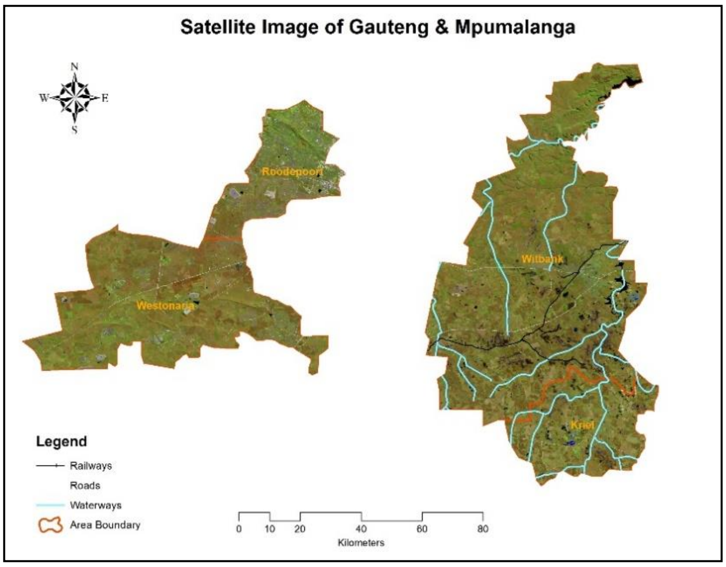


Figure 6. 3 Raw, unprocessed Landsat satellite image of the study area (False Colour Composite as Red: band NIR, Green: band RED and Blue: band Green)

The geometric correction is the georeferencing of satellite images with respect to the ground control points in order to obtain pixels with the same dimensions. Whereas for atmospheric corrections there are two different approaches: (1) relative normalization (Schroeder et al., 2006) and absolute correction (Chavez, 1996; Song & Woodcock, 2003). Relative normalization includes radiometric correction of the Landsat datasets with respect to a reference image based on the correlation between pseudo-invariant objects from multi-date images (Song et al., 2001). Absolute correction can be subdivided into two more categories: empirical and physical-based methods. In the empirical approach, the spectral properties of the ground features are used for transforming the spectral data from the sensor's radiance to ground reflectance. Empirical methods are simple but don't incorporate the pixel-to-pixel variation in atmospheric effects. Whereas the physical-based methods, such as Atmospheric/Topographic CORrection (ATCOR; (Richter, 1997), MODerate resolution atmospheric TRANsmission (MODTRAN; (Berk et al., 1998), and the Satellite Signal in the Solar Spectrum (6S) code (Vermote et al., 1997), incorporates the heterogeneity of the atmosphere, but required several complicated and manual procedures, which make them hard to process big amount satellite data. On the other hand, the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) software (Masek et al., 2006), which has implemented the 6S code, made an atmospheric correction for Landsat 4–7 fully automated. Recently, Vermote et al. (2016) made an improved atmospheric correction algorithm for Landsat 8 (L8SR), which has shown an improvement over the ad-hoc Landsat 5–7 LEDAPS product. The modified LEDAPS product was applied in this study for atmospheric corrections as shown in Figure 6.4.

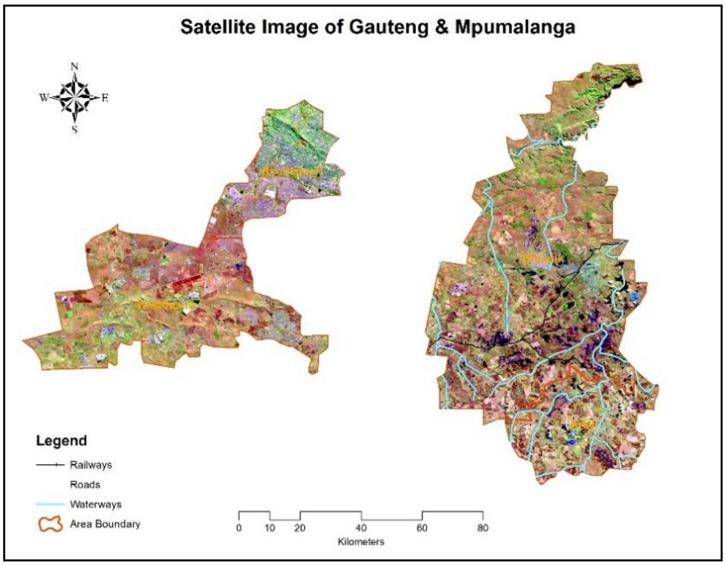


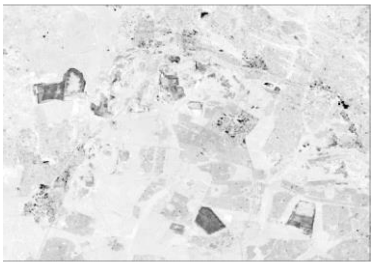
Figure 6. 4 Atmospherically corrected and enhanced Landsat satellite image of the study area using Near Infrared, Red, and Green spectral bands combination

6.2.4. Spectral sub-setting

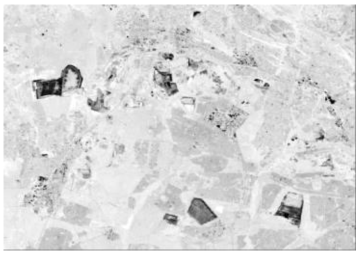
As OLI band 1 (coastal aerosol) is useful for imaging shallow waters and band 9 (cirrus) to detect the high-altitude clouds, and tracking fine particles like dust and smoke, these two bands were not included in further analysis. Moreover, according to the literature, for mineral exploration mapping, the most appropriate bands are located in the visible, NIR and SWIR regions. In this research study, OLI bands 2 – 7 and TIRS bands 10 – 11 were used for advanced processing. All these bands were stacked as a single image with spatial resolution of 30m using Layer Stacking digital image processing technique (Mahboob; et al., 2015).

6.2.5. Conversion of digital number to reflectance

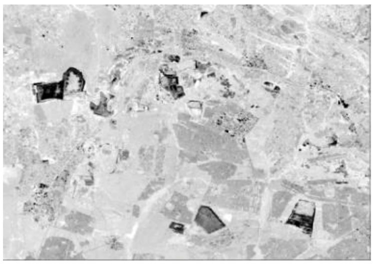
Usually, the L1T images of Landsat-8 consist of the Digital Numbers (DN), which cannot be used due to lack of physically meaningful information and should be converted to surface reflectance. This conversion is required for quantification of different features in remote sensing data as it incorporates solar conditions (geometry, illumination, and intensity) when the images were captured. In this study, the data was converted to the Top of Atmosphere (TOA) reflectance using radiometric coefficients as recommended by Roy et al. (2016), where DN were converted to reflectance representing the ratio of the radiation reflected off a surface to the radiation striking it (Han & Nelson, 2015) as shown in Figure 6.5.



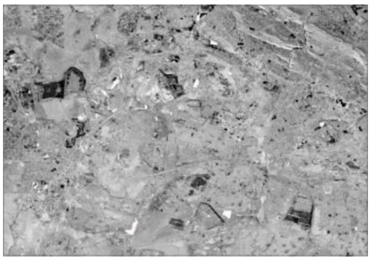
Band 2 (Blue)



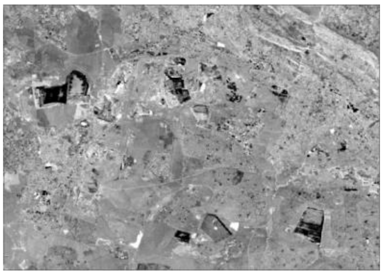
Band 3 (Green)



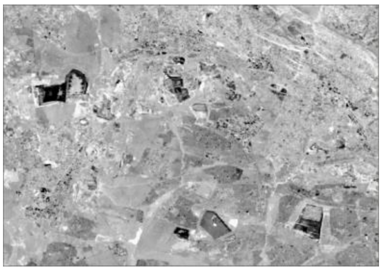
Band 4 (Red)



Band 5 (Near-Infrared)



Band 6 (Shortwave-Infrared-I)



Band 7 (Shortwave-Infrared-II)

Figure 6. 5 Spectral bands of Landsat satellite

Equation 6.1 was used to convert DN values to TOA reflectance (Zanter, 2016):

$$\rho_\lambda = \frac{M_\rho \times Q_{cal} + A_\rho}{\sin(\theta)} \quad \text{Equation 6. 1}$$

where:

ρ_λ = Top of Atmosphere Planetary Reflectance (dimensionless).

M_ρ = Reflectance multiplicative scaling factor for the band.

A_ρ = Reflectance additive scaling factor for the band.

Q_{cal} = Level 1 pixel value in DN.

θ = Solar elevation angle (degrees).

6.2.6. Reflectance spectroscopy of satellite image

The data captured by Landsat 8 remote sensing satellite represents the reflected and/or emitted spectral energy, which can be further used based on absorption characteristics of spectra to detect different materials and features of the earth. Some minerals and mineral groups in hydrothermally altered rocks have unique absorption characteristics in the EM spectrum. For example, alunite and some clay minerals (pyrophyllite and kaolinite group of minerals) have unique absorption features at ~2.1 µm, and their spectral responses are much higher at ~1.7 µm

(Sabins, 1999). Usually, the minerals like iron oxide and sulphate have low and high reflectance in the ultraviolet/blue and near-infrared portion of EM spectrum respectively (Johnson et al., 2016), because of which these minerals have a rusty shade in a natural colour image.

6.2.7. Band-ratioing of satellite image

Band-ratioing is a digital image-processing technique that enhances the contrast between features by dividing a measure of reflectance for the pixels in one band by the pixels in the other band of the same satellite image. This technique has already been widely used by several researchers for visualizing and mapping of hydrothermally altered rocks. For example, Han and Nelson (2015) efficiently used the image ratios of Landsat Thematic Mapper (TM) band 5 (1.55– 1.75 µm) over band 7 (2.09– 2.35 µm) to differentiate areas with high concentrations of alunite and clay, where pixels in the satellite image appear bright. Another study conducted by Van Der Meer (2004) used the ratio image of band 3 (0.63– 0.69 µm) over band 1 (0.45 –0.515 µm) to highlight areas with rich iron ores. In current research work different band ratios, as shown in Table 6.2 were developed and applied in order to enhance hydrothermally altered rocks and lithological units. The selection of bands is related to the spectral reflectance and position of the absorption bands of the mineral or assemblage of minerals to be mapped.

Table 6. 2 Band ratios for enhancement of hydrothermally altered rocks and ore bodies

Sr. No	Spectral Band Name	Wavelength (nm)	Type of Enhanced Mineral
1	$\frac{\text{Red}}{\text{Blue}}$	$\frac{640\ -\ 670}{450\ -\ 510}$	*Iron oxide*
2	$\frac{\text{Shortwave} - \text{Infrared (SWIR) 1}}{\text{Shortwave} - \text{Infrared (SWIR) 2}}$	$\frac{1570\ -\ 1650}{2110\ -\ 2290}$	*Hydroxyl bearing rock*
3	$\frac{\text{Shortwave} - \text{Infrared (SWIR) 2}}{\text{Near} - \text{Infrared (NIR)}}$	$\frac{2110\ -\ 2290}{850\ -\ 880}$	*Clay minerals*

4	$\frac{\text{Shortwave} - \text{Infrared (SWIR) 1}}{\text{Near} - \text{Infrared (NIR)}}$	$\frac{1570\ -\ 1650}{850\ -\ 880}$	*Ferrous mineral*

Even though the band-ratioing technique works very well for visualization, it is not capable of mapping or quantification of the land features. One of the reasons for this limitation is that most of the optical multi-spectral remote sensing instruments used the bandwidths > 0.05 µm, which is too wide to explicitly differentiate the unique spectral absorptions related to specific alteration minerals (Van Der Meer, 2004). Furthermore, many band-ratioing techniques use only two or sometimes three bands whereas multispectral remote sensing instruments offer many more bands than that. Based on these limitations, there is a need to adopt another advanced digital satellite image analysis approach which can utilize all available satellite bands together as highlighted in the next sections.

6.2.8. Spectral band combination

Spectral band combination technique is also known as Red Green Blue (RGB) combination is a very useful image enhancement technique which offered powerful means to visually interpret a multispectral satellite data (Novak & Soulakellis, 2000). The band combinations of satellite data can be true (natural) or false-colour composites (i.e., FCC) using individual bands or band ratios. For this purpose, several band ratios and bands combinations had been developed over the time to differentiate lithology in a satellite image as few examples are given in Table 6.3.

Table 6. 3 Landsat TM spectral band combinations, RGB combinations for enhancement of hydrothermally altered rocks and ore bodies

Sr. No	Analysis Type	Spectral Bands	Type of Enhanced Mineral	References
1	$Kaufmann\ ratio$	$\frac{7}{4}:\frac{4}{3}:\frac{5}{7}$	Red color represents minerals containing iron ions; Green represents vegetated zones, and Blue	(Kaufmann, 1988)

			represent hydroxyl-bearing minerals.	
2	$Chica\text{–}\ Olma\ ratio$	$\frac{5}{7} : \frac{5}{4} : \frac{3}{1}$	Red color depicts Clay; Green color represents the Iron-ions; whereas Ferrous Oxide represented in blue color.	(Mia; & Fujimitsu, 2012)
3	$Sabin's\ ratio$	$\frac{5}{7} : \frac{3}{1} : \frac{3}{5}$	Yellow color represents hydrothermal alteration areas; Black identifies water; the Dark tones of Green indicates vegetation whereas the Lighter tones of Green signifies the clay-rich rocks; Blue shows sand; Red, Pink or Magenta indicates some mineral rocks-iron oxides.	(Sabins, 2007)
4	$Sultan's\ ratio$	$\frac{5}{7} : \frac{5}{1} : \frac{5}{4} \times \frac{3}{4}$	Deep violet color represents the hydroxyl minerals; Green will present the ferric, and Blue color highlights the ferrous iron oxides.	(Gad & Kusky, 2006)
5	$Abrams\ ratio$	$\frac{5}{7} : \frac{3}{1} : \frac{4}{5}$	The hydrothermally altered iron-oxide will be represented as Green and clay minerals as Red color.	(Pour & Hashim, 2012)

In this research, the Kaufmann and Sabin's ratios were applied to highlight the hydrothermally altered rocks associated with the minerals present in the study area. Most of the research studies have applied these two ratios for the identification of altered areas and their associated minerals. For example, Mia and Fujimitsu (2012) and Abhary et al. (2016) applied the Kaufmann ratio for the mapping of minerals containing hydroxyl and iron ions using Landsat satellite data. Similarly, da Cunha Frutuoso (2015) used Sabin's ratio for the identification of sulphide deposits associated with the alteration areas of iron oxides.

6.2.9. Principal component analysis

Principal component analysis (PCA) is an advanced information extraction technique and is frequently used in the earth sciences, e.g., (Cheng et al., 2011; El-Makky, 2011). PCA is a famous multivariate statistical method and has been generally used to study associations between variables. By orthogonal transformation, several correlated variables can be transformed into uncorrelated combinations (eigenvector loadings) of principal components (PCs) based on their covariance or correlation matrix (Horel, 1984; Loughlin, 1991). Generally, the first few PCs highlights the most variability in the original dataset (Panahi et al., 2004). Therefore, PCA reduced the dimensionality and redundancy of datasets and is commonly applied to enhance information interpretability (Cheng et al., 2011; Horel, 1984; Jolliffe, 2002). According to the algorithm of PCA, PCs are linear combinations of the original variables, whereas each PC incorporates input variable uniquely and signifies only limited information within the complete dataset (Abdi & Williams, 2010). Due to its handling of multivariate datasets PCA has been extensively used in remote sensing for mapping of geological bodies like ores, igneous rocks, strata, etc. (Grunsky et al., 2014). Generally, the spectral properties of different features present in the area i.e. vegetation, rocks and soils are responsible for the statistical variance mapped into each PC which becomes the basis of the Crosta technique (Tangestani & Moore, 2001). In this study, the same technique has also been applied based on highly variable non-correlated satellite bands for hydrothermally altered rocks.

6.3. Results and Discussion

The effectiveness of the remote sensing based hydrothermal minerals identification and mapping depend on clear differentiation of the reflected spectra of altered bedrock from those of the other objects. The true colour composite of bands 3, 2 and 1 as Red, Green, and Blue respectively highlighted the textural characteristics of the igneous rocks which could be separated from those of sedimentary rocks. Pournamdari et al. (2014) tested the same satellite band combinations and found them to be effective for extraction of igneous rocks in south Iran. The false colour composite was assigned to bands 4, 3 and 2 as Red, Green, and Blue respectively and shown in Figure 6.6 to analyse the reflected satellite spectroscopy. Whereas false colour composite is important to enhance the regional geological and geomorphological features as also reported by Bedini (2009). Vegetation appeared in red shades because of near-infrared (0.7 to 1.2 μm) band which was highlighted with red colour and vegetation reflects the maximum in the near-infrared.

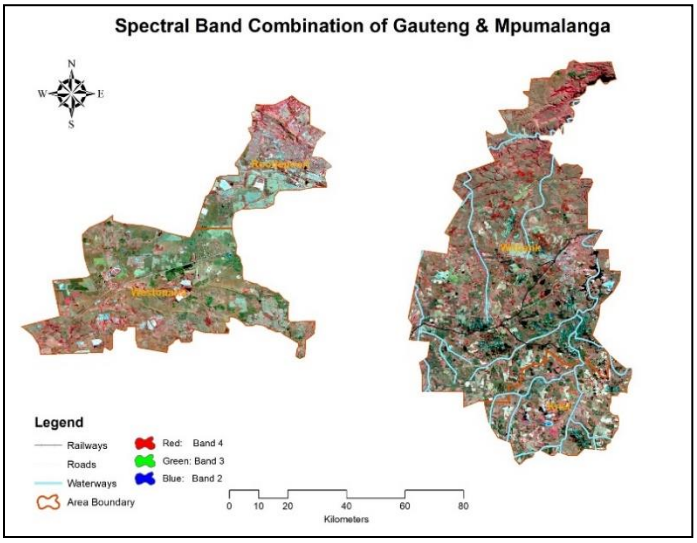


Figure 6. 6 RGB combination for bands 4, 3, 2. Enhanced image where outcrops are represented in shades of green and vegetation in dark red

Hydrothermally altered clay and carbonate minerals are recognizable as a yellow area of crystalline igneous rocks in the Gauteng and Mpumalanga district as shown in Figure 6.7, this may be due to clay and carbonate minerals have absorption at 2.1–2.4 μm (band 7 of Landsat-8) and reflectance at 1.55–1.75 μm (band 6 of Landsat-8) properties. Van Der Meer (2004) also reported the same absorption and reflectance bands in the clay minerals using NASA’s Airborne Visible/Infrared

Imaging Spectrometer (AVIRIS) data. Another study conducted by Zaini et al. (2016) concluded that clay and carbonates have the same absorption and reflectance bands and these can be effectively used for their mapping through reflectance spectroscopy.

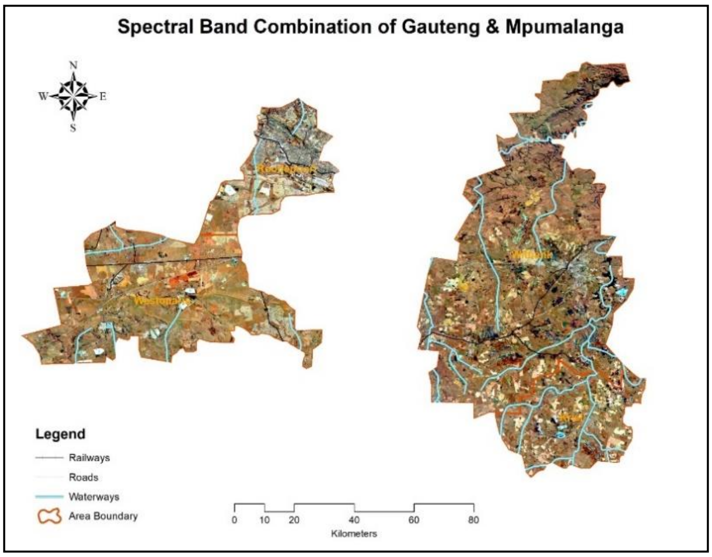


Figure 6. 7 RGB combination for bands 6, 3, 1. Enhanced image where outcrops are represented in shades of yellow

The iron oxides present within the study area were highlighted by the spectral band ratio of Red and Blue bands as discussed in Table 6.2. The soils rich in iron oxides reflects more in the red band of the spectrum, i.e. 0.64 - 0.67 μm and it has absorption characteristics at blue band i.e. 0.45 - 0.51 μm (Schwertmann, 1993). The typical reflectance curve of iron oxide-rich soils is shown in Figure 6.8. A. B. Pour and M. Hashim (2015) had also shown the importance of the red and blue bands of Landsat satellite data for mapping of iron oxides in Iran region. In 2017, Pour et al. also extracted the iron-rich mineralized zones of remote Antarctica from Landsat satellite imagery by using spectral band ratio techniques.

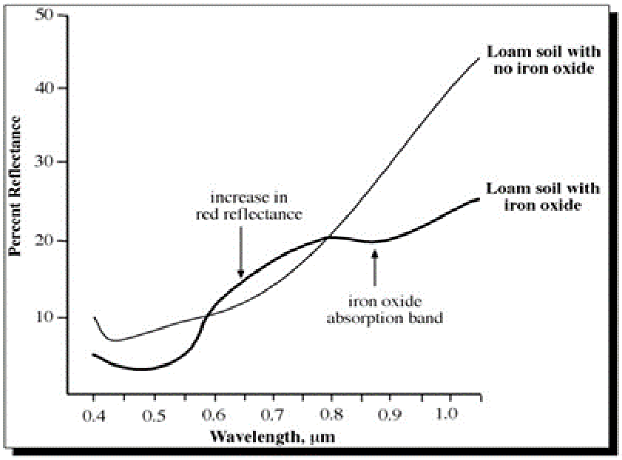


Figure 6. 8 Typical spectral reflectance curve of iron oxide in loam soils, source (Rathod et al., 2016)

The map of potential iron oxides presents in the study area derived from band ratio of the red and blue wavelengths of the spectrum is shown in Figure 6.9. This map of the iron oxides in comparison with the Council for Geoscience (CGS) published a map of iron deposits of South Africa showed a good agreement as highlighted in Figure 6.10. The area circled as dark blue represents the iron deposits in Gauteng and Mpumalanga. In addition, the research conducted by Holmes and Lu (2015) supported the results of this study and highlights the potential of iron deposits in the Gauteng and Mpumalanga.

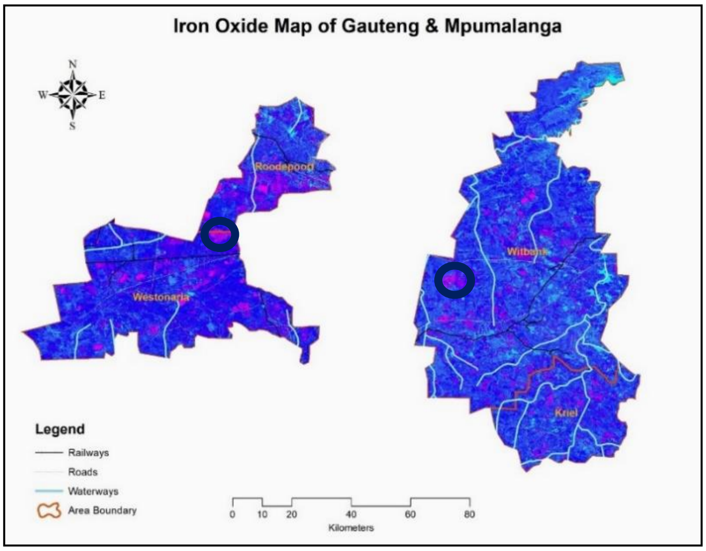


Figure 6. 9 Band ratio of red and blue band enhanced soils with potential iron oxides as represented in shades of purple

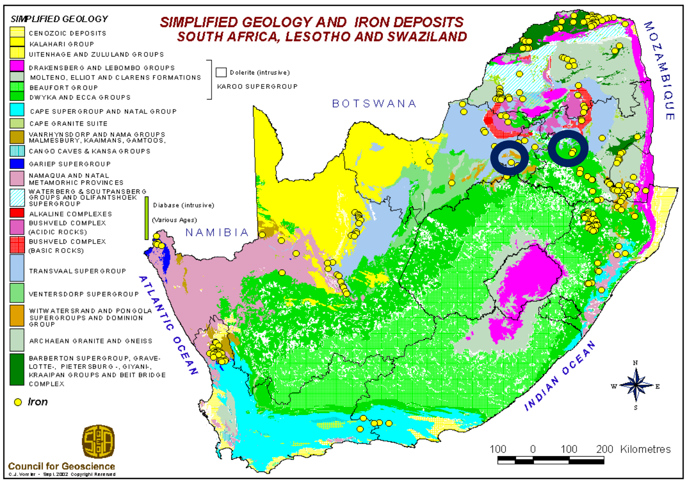


Figure 6. 10 Iron deposits of South Africa, Source: (Vorster, 2012b)

Gold cannot be “seen” directly in any remotely sensed satellite image, however, the presence of this precious metal can be mapped with the presence of several other associated minerals based on their spectral reflectance (Kotnise & Chennabasappa, 2015). The group of minerals present in the alteration zones related to gold deposits is generally referred to as clay minerals including illite-1M, illite-$2M_1$, Dioctahedral smectite and Kaolinite etc. (Drews-Armitage et al., 1996). These minerals have indicative spectral signatures mostly in the shortwave-infrared portion of the electromagnetic spectrum. These spectral signatures can be used to map the potential sites that are most favourable for the occurrence of gold deposits, which is very cost and time effective for the mineral industry for their exploration programs. The band combination of shortwave-infrared 2 with spectral band 2.11 - 2.29 μm and near-infrared with spectral band 0.85 - 0.88 μm shows the areas in shades of red to yellow which have the potential of clay minerals as shown in Figure 6.11. L. Liu et al. (2018) applied various image processing techniques, including false colour composite, band ratios, and matched filtering, to process ASTER satellite data and map the distribution of hydrothermal minerals associated with the gold deposition. They concluded that the SWIR and NIR bands are most effective to map the regions associated with the gold deposits.

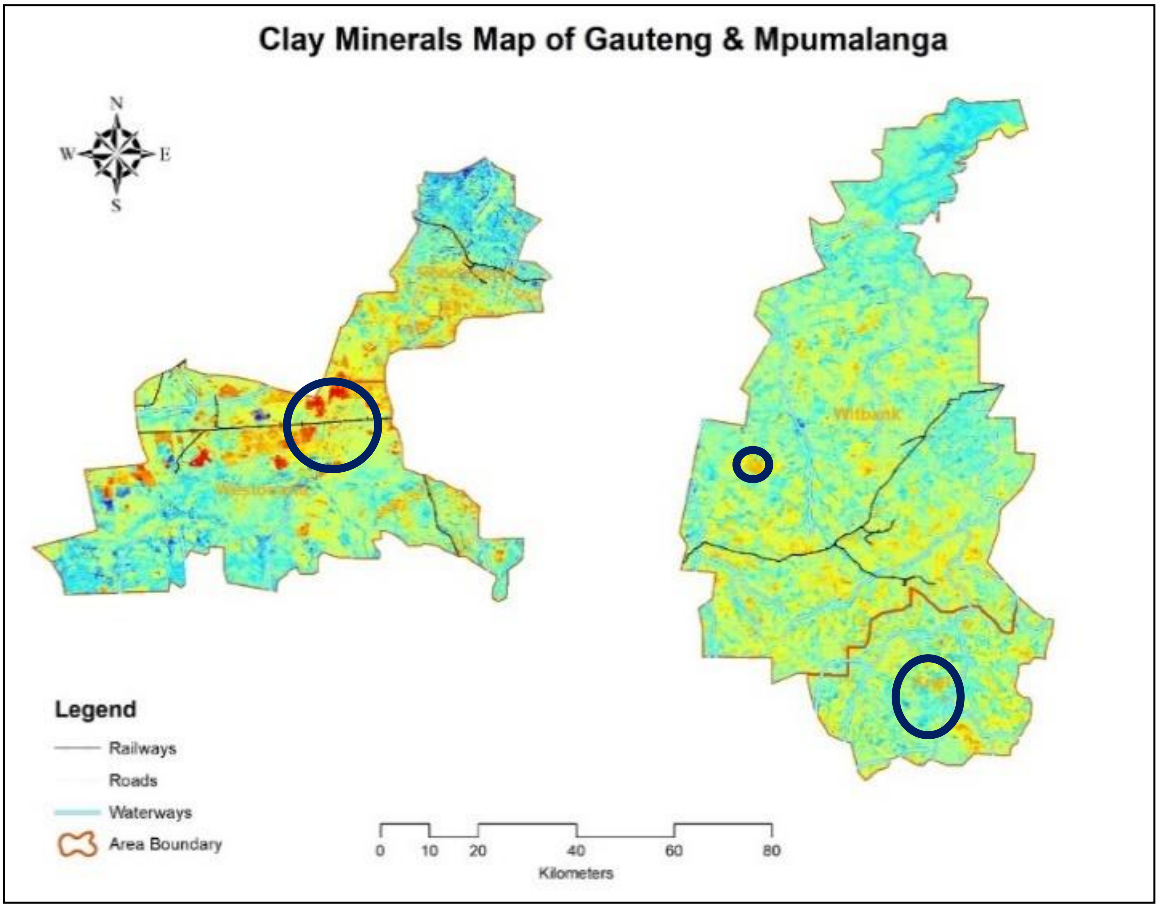


Figure 6. 11 Band ratio for enhancement of potential clay minerals as represented in shades of red to yellow

In general, Gauteng province has more clay minerals which are associated with the gold mineralization potential as compared to the Mpumalanga province (Durand, 2012; Sutton et al., 2006; Sutton, 2013). The approach that was applied to map potential gold in this study has also been applied by several other researchers i.e. Safari et al. (2017) used a similar approach to identify gold mineralization and concluded that integration of remote sensing data with other information sources led to the definition of locations possibly suitable for hosting Sn-W and Au-Ag mineral occurrences. Another study conducted by Gabr et al. (2010) utilized the shortwave-infrared and near-infrared bands of ASTER satellite to map high-potential gold mineralization in Egypt. However, through an optical sensor satellite, it is quite difficult to directly map any mineral located at the depth of 100 meters or more in the ground as in the case of Gauteng. But there are always indirect measurements from the surroundings such as altered rocks or soils on the surface which gives an indirect indication for the presence of these minerals such as gold. In this study, the identified potential gold areas (Figure 6.11) can be associated with the gold mining tailings, because no direct and indirect measurement can be made to map surface gold potential in Gauteng due to the geological settings of the region. The map of gold deposits developed by the Council for Geosciences (CGS); South Africa (Figure 6.12) was used in order to verify/support the potential gold (tailings) identification in the area.

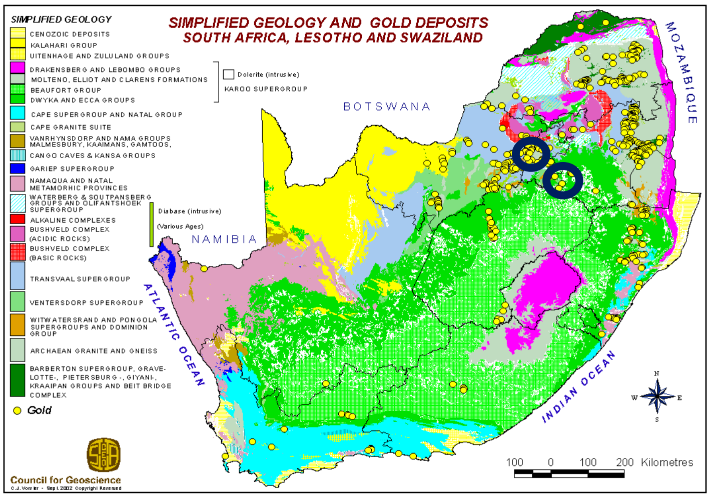


Figure 6. 12 Gold deposits of South Africa, Source: (Vorster, 2012a)

The map showed that there are several gold deposits in Gauteng province but no or minimum deposits in Mpumalanga, which supports the results of the gold (probably gold tailings) map (Figure 6.11) generated during this study.

The other important band ratio comprised of shortwave-infrared and near-infrared was applied for identification of ferrous minerals. The results showed that this band ratio technique produced good results of ferrous minerals in Gauteng but overestimated in Mpumalanga. The soils with ferrous minerals reflect mostly in shortwave-infrared 1 with spectral band 1.57 - 1.65 μm and near-infrared with spectral band 0.85 - 0.88 μm (Ducart et al., 2016) and are shown in Figure 6.13 in shades of orange to yellow.

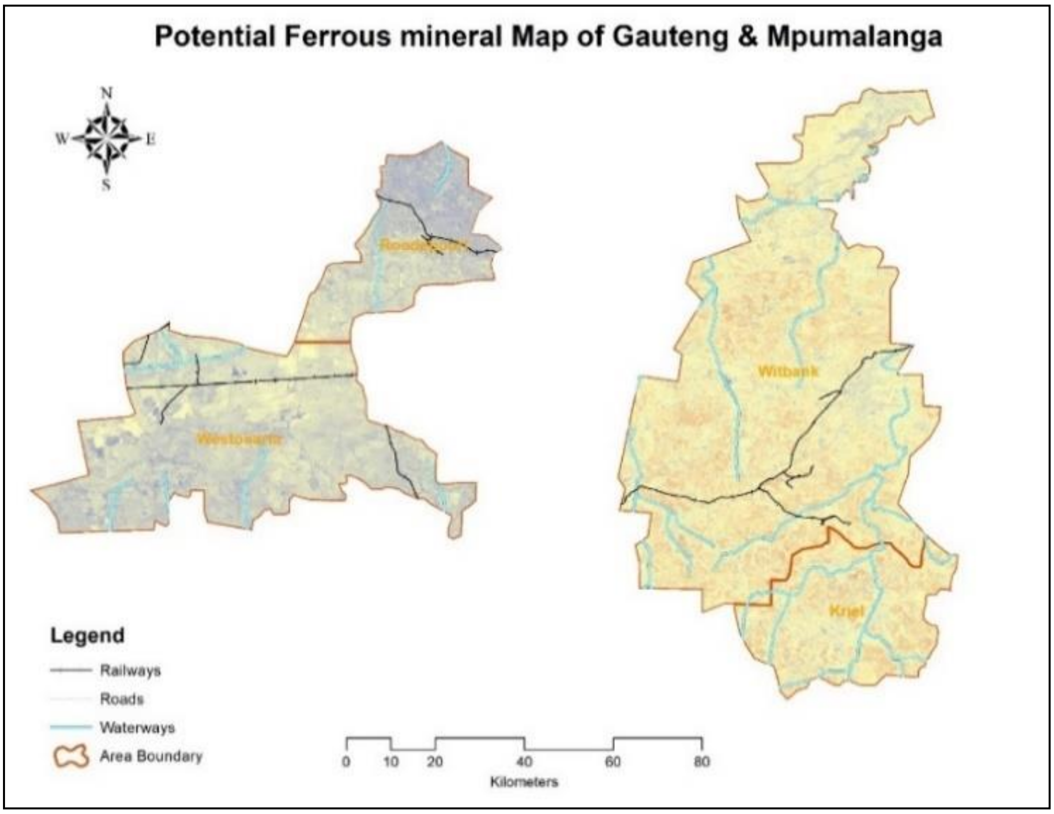


Figure 6. 13 Band ratio for enhancement of potential ferrous minerals in soils as represented in shades of orange to yellow

The Kaufmann and Sabin's ratios were developed by combining spectral reflections of satellite bands ratios, and their results were proved to be promising for mapping of hydrothermal minerals (Mahandani et al., 2018). da Cunha Frutuoso (2015) also applied the same ratios to map the hydrothermal gold mineralization in Portugal and found them to be very effective and accurate. Figure 6.14 and 6.15 show that results of these ratios both for Gauteng and Mpumalanga province respectively. In the Kaufmann ratio map, the red colour represents minerals containing iron ions; green represents vegetated zones, and yellow represents hydroxyl-bearing minerals. Whereas, in Sabin's ratio map the shade of green colour represents hydrothermal alteration areas; black identifies water; the dark tones of green indicate vegetation whereas the lighter tones of greenish-yellow signify the clay-rich rocks; blue shows sand; red, pink or magenta indicates some mineral rocks-iron oxides. These colour shades for several earth features and minerals are in accordance with the research done by Sabins (2007) and showed the effectiveness of the Sabin's ratio in mineral mapping.

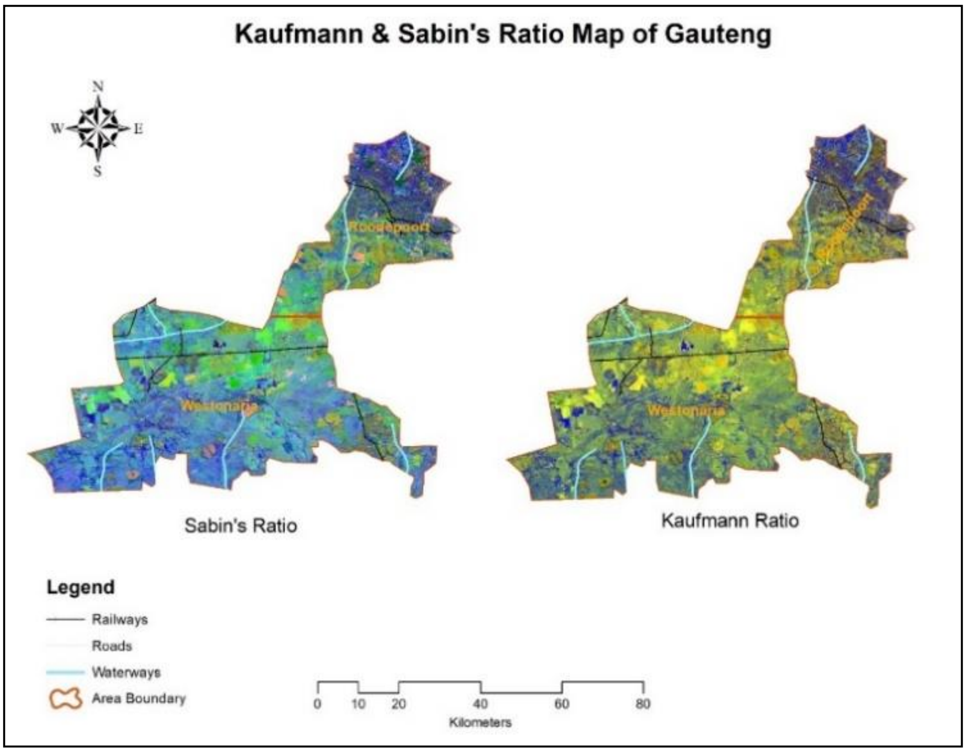


Figure 6. 14 Kaufmann and Sabin's ratios representing the hydrothermal minerals in different shades of colours for Gauteng region

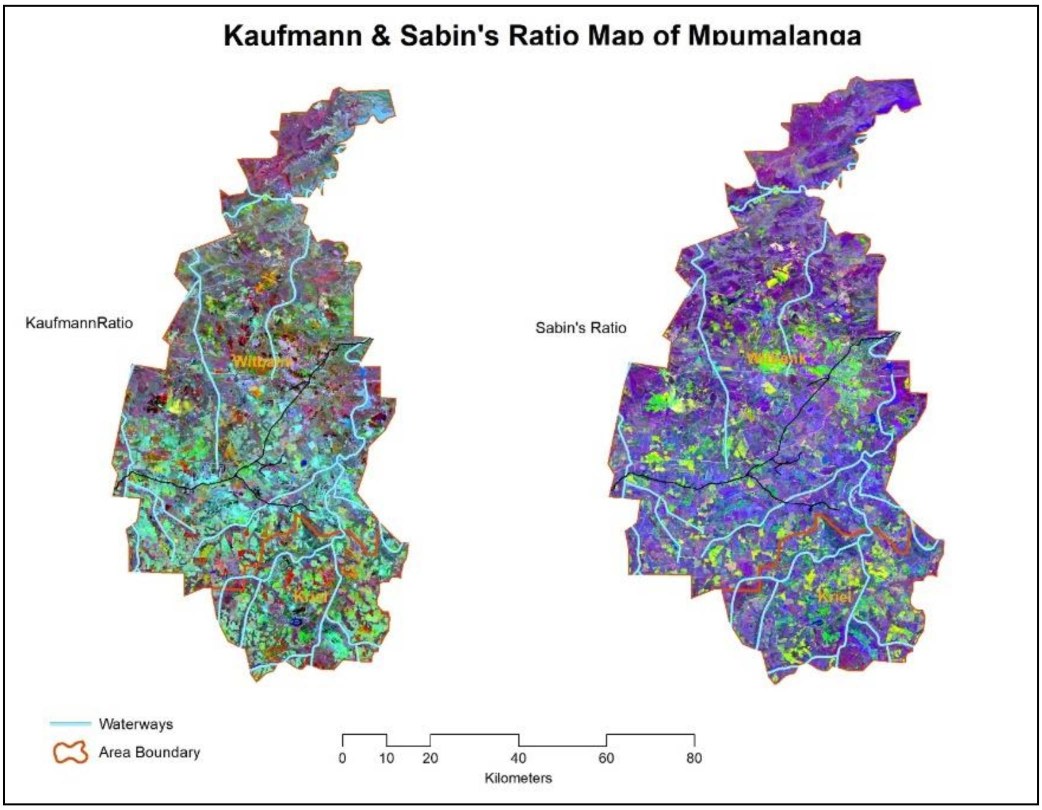


Figure 6. 15 Kaufmann and Sabin’s ratios representing the hydrothermal minerals in different shades of colours for Mpumalanga

Similarly, the outputs of principal component analysis also revealed very good results in terms of hydrothermal mapping of minerals present in the study area. In our study, the principal component transformation specifies that the first principal component (PC1) is composed of a negative weighing of all total bands as shown in Table 6.4.

Table 6. 4 Principal component analysis of Landsat satellite bands of Gauteng and Mpumalanga area

Input Bands	PC1	PC2	PC3	PC4	PC5	PC6
Band 2	0.360613	-0.149047	-0.473757	0.642169	-0.156241	-0.431860

Band 3	0.346576	0.127976	0.409188	0.052900	-0.093199	0.827397
Band 4	-0.352093	-0.273256	0.602391	0.608503	0.463980	-0.352654
Band 5	-0.489315	0.846570	-0.302391	-0.155955	-0.127694	0.055763
Band 6	0.484542	0.143013	0.471431	0.321111	0.435239	-0.037414
Band 7	-0.389437	0.386786	-0.256332	0.295141	0.738839	0.005991
Eigenvalues (%)	94.53	7.93	3.67	0.08	0.04	0.001

The PC1 is about 94.53% of the eigenvalue of the total variance for un-stretched data of principal component analysis. The eigenvector loadings for PC3 indicates that PC3 is dominated by vegetation, which is highly reflective in Band 4; the positive loading of Band 4 in this PC (0.602391) also indicates that strongly vegetated pixels will be bright in this PC image. The similar results were also found by Cheng et al. (2011), as they have vegetated dominated pixels in band 4 loadings for PC3.

Because the eigenvalues for Band 2 and 4 in PC6 of Table 6.4 are also opposite in sign, which highlights that iron oxides will be distinguished by bright pixels in PC6. PC5 can be used to map the Hydroxyl-bearing minerals because of the positive and negative eigenvalues of band 7 and 5 respectively and appeared as dark. These selected PCs were further used in Crosta technique, the hydroxyl (H) and iron-oxide (F) images are combined with selected PCs to produce a map revealing the pixels with abnormal concentrations of both hydroxyls and iron-oxides. Whereas the principal components having positive eigenvalues from both input bands is selected for the analysis. Figure 6.16 shows the results of Crosta images of hydroxyl (H), hydroxyl+iron-oxide (H+F) and iron-oxide (F) as RGB composite. This band combination returns a dark bluish colour composite image on which alteration zones are highlighted as bright regions as also discussed by Ciampalini, Garfagnoli, Antonielli, et al. (2013).

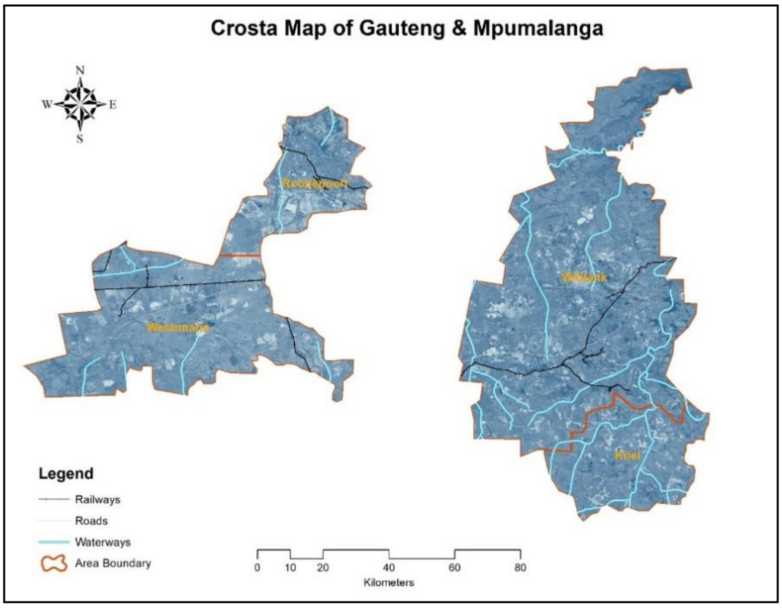


Figure 6. 16 Hydrothermal minerals map through Crosta technique as light shades of blue

6.3.1. Validation of the results

The most common method to evaluate the results of this sort of mineral-mapping is the application of spectrometer and/or spectral testing of samples in the lab. But according to Clark (1999), there are two types of methods for authentication of information extracted from remotely sensed imagery i.e. 1) virtual and 2) in situ. If the spatial and/or spectral resolution of remotely sensed satellite data is fine and accurate then virtual verification can be done by inspecting the remote sensing data directly and comparing it with the already published reports or data. In this paper, virtual verification, i.e., visual interpretation of absorption of spectral bands and comparison with the published maps of Council for Geoscience (CGS) were used to evaluate the results of hydrothermal mineral exploration along with the similar research studies done by several researchers. The good qualitative agreement was observed for the results of principal components and the spectral reflectance of the satellite data.

6.4. Conclusions

This study confirms that reflectance spectroscopy of remotely sensed Landsat satellite data can be used easily and efficiently to map hydrothermal minerals. In this study different digital image processing algorithms and techniques were applied to assess their significance for hydrothermal mineral exploration. The Principal Component Analyses (PCA) based Crosta technique and band ratio techniques like Kaufmann and Sabin’s ratios are proved to be more significant and efficient for hydrothermal mineral exploration. Several researchers also concluded that the advanced image processing techniques like applied and tested in this current study are quite efficient in terms of mineral-mapping through remote sensing data (L. Liu et al., 2018; Manuel et al., 2017). The maps produced in this study are not only appropriate for any spatial queries and analysis but also for environmental modelling such as assessing the impacts of mining activities on environmental features such as the forest, farmlands, urban areas, etc. The hydrothermal minerals identified in this study are only the estimation based on literature cited proved algorithms for remotely sensed data, and they should be treated as the preliminary assessment of the study area. The other important point is about the resolution of the satellite data used in this study i.e., Landsat 8 which has a 30 – 100 meter spatial and moderate spectral resolution. The datasets with this kind of resolutions may be enough for regional work but are insufficient for much detailed minerals mapping. Using satellite data with the higher spatial and spectral resolutions, such as data of Worldview 3 satellite (Kruse, 2015) or aerial drones, may be more suitable for much detailed minerals mapping. So, the maps developed in this research are a valuable data source for the comprehensive studies to be conducted in the future. The methodology developed in this research study is cost-effective and time-saving which can be applied to the inaccessible and/or new areas with limited ground knowledge where reliable and updated minerals information is desired.

7. EVALUATION OF ISODATA CLUSTERING ALGORITHM FOR SURFACE GOLD MINING USING SATELLITE DATA

This chapter is based on the following publication

Mahboob, M. A., & Genc, B. (2019, July). Evaluation of ISODATA Clustering Algorithm for Surface Gold Mining Using Satellite Data. In *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)* (pp. 1-6). IEEE.

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

7.1. Introduction

Surface mining activities such as mineral exploration, exploitation, and dumping of overburden cause huge environmental changes, if not done in a proper way. These changes become worse in the case of poorly planned and/or illegal mining scenarios (Thomas, 2019). Therefore, it is very important to monitor the mining activities in any region on a regular basis. The monitoring of a large mining area with remote sensing satellite technologies is not only practical but also cost-effective. Several research studies have been applied remote sensing data in the mining industry from environmental monitoring (Karan et al., 2016; Mahboob et al., 2019; Yu et al., 2018; Zeng et al., 2017) to post mine closure land management (CHEN et al., 2004; Yang, 2008).

Moreover, the satellite data is multi-dimensional in terms of spatial, spectral and radiometric aspects. Due to this multi-dimensionality, the satellite images are usually of big size and are difficult to process in order to convert it into useful information. It is always recommended to perform image processing on the raw satellite images before any further analysis (Jensen, 2015). Due to the advancement in computer technologies, now there are several digital image classification schemes ranging from supervised to unsupervised, hybrid to parametric and pixel-based to generic classification, which can be used not only to identify certain land covers but also to quantify them (Mather & Tso, 2016). All these classification schemes have their own strengths and weaknesses depending on the underlying conditions and selection of appropriate algorithms in each case. The choice of appropriate algorithms for objects identification and mapping can highly affect the accuracy of the results and thus the decision making. Several research studies (Munyai, 2017; Obodai et al., 2019; Schueler et al., 2011) have tested different classification schemes for surface gold mining mapping. A study conducted by Hemalatha and Anouncia (2017) showed that ISODATA clustering algorithms are not only time saving but are efficient in terms of accuracy. However, the accuracy highly depends on the selection of parameters used to tune the model. Another study conducted by Wang et al. (2014) used the ISODATA clustering algorithm for classification of hyperspectral remote sensing satellite images and the results showed that the improved ISODATA algorithm performs much better than the Spectral Angle Mapper is a supervised classification method as it employs user-defined reference spectra as training data for image classification. Similarly Sunuprapto et al. (2016) used the ISODATA cluster algorithm for identification of surface gold mining along with several other digital image processing techniques and concluded that, the ISODATA is a robust unsupervised image classification technique but is sensitive to the choice of the number of initial clusters used for classification. Besides that, very little research has been conducted in terms of the selection of proper parameters on which the accuracy of the ISODATA clustering algorithm is highly depended for gold mining regions mapping. In this research, an unsupervised image classification algorithm ISODATA clustering has been applied on satellite remote sensing data for identification of gold mining areas in the image.

7.2. Materials and Methods

7.2.1. Study area

The study area for this research was the Obuasi mining town of Ashanti Region of southern Ghana as shown in Figure 7.1. The town is known for its gold mining, and according to CNCrusher (2018) it is one of the nine largest gold mines on the earth, and mining activities in the region are as old as since 17th century. Due to the presence of high gold content in the region, recently there is a boom in legal as well as illegal gold mining in the area. The unplanned and unsustainable gold exploration and exploitation in the region are causing several devastating impacts on the surrounding environment such as land and forest degradation, soil, water, and air pollution and several other health-related issues.

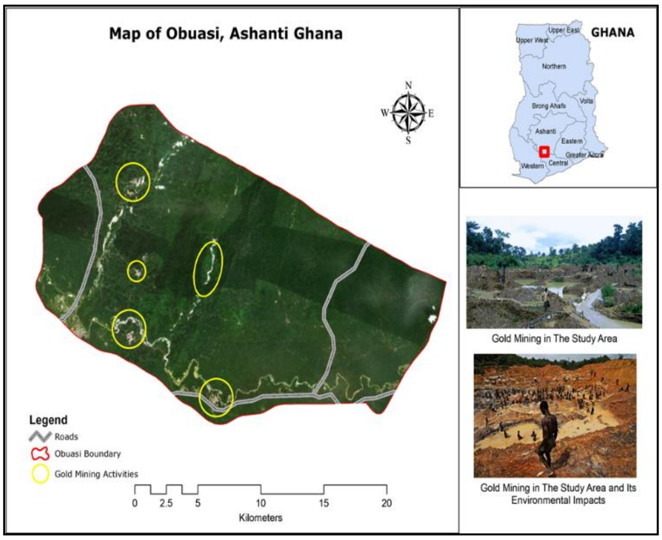


Figure 7. 1 Map of the study area

7.2.2. Satellite data

The satellite data used in this research was acquired from sentinel satellite recently launched by European Space Agency in 2015. The sentinel satellite provides a fine scale spatial and multispectral data with good time frequency. This data has been used in several research studies for environmental monitoring of mining-related activities but the application of different computer-based classification algorithms is limited to this dataset. The most recent available satellite image of the year 2019 with a spatial resolution of 10m and 4 spectral bands including Blue, Green, Red, and Near-Infrared was processed.

7.2.3. Satellite Image Pre-processing

Usually, the Sentinel satellite image is of 100 × 100 m in width and is a large data set to process as a single image. Hence the study area was masked out from the large image to a comparatively smaller image, which is also easy to process. The satellite data is usually composed of raw Digital Numbers (DN) values in each pixel, it is highly recommended to convert the DN numbers into Top-Of-Atmosphere (TOA) sensor reflectance. The Sentinel satellite data is available in both DN and TOA forms, so in this study, the TOA-based data was acquired and further image processing was done. The Histogram Equalization technique was adopted in order to adjust and enhance the contrast of the satellite images by varying the intensity distribution of the spectral bands. The result of the pre-processed and enhanced satellite image was used for further application of the ISODATA clustering algorithm.

7.2.4. ISODATA clustering algorithm

Iterative Self-organizing Data Analysis Techniques Algorithm (ISODATA), is an unsupervised semi-automated classification scheme which has been widely used in several research studies in order to map certain landcover classes in satellite images (Abburu & Golla, 2015; Jensen, 2015; Verma et al., 2017). ISODATA clustering algorithm used the minimum spectral distance formula to form the clusters of the pixels in the form of the unique classes. The initialization of the algorithm can be done with either an arbitrary mean of clusters or with means of existing class signatures. As it is an iterative process, each time the clustering is done, the mean of the developed class is shifted from old to a new location. The algorithm will keep generating the classes until one of the two conditions met, either the maximum allowed number of iterations or a maximum allowed percentage of unchanged pixels has been achieved between the two successive iterations.

The adjustable parameters in the ISODATA classification scheme were, 1) the minimum and maximum number of classes to be generated; 2) minimum percentage size of a class, if a cluster has less number of pixels than this specific limit then the cluster will be deleted at the end of iteration; 3) maximum standard deviation which will allow the algorithm to divide the clusters into two if the value increases than this; 4) minimum distance parameter is used for merging of the clusters based on the Euclidean distance; 5) maximum merges which allow the number of clusters that can be merged together at the end of each iteration; 6) the number of iterations which determines the specific number of times the algorithm can re-cluster the pixels, this parameter actually prevents the algorithm to run too long and from getting stuck in a particular class in order to identify the unique number of pixels and 7) the convergence threshold which specifies the number of pixels that should remain unchanged at the end of each iteration.

In this research study, we have changed the number of iterations and convergence threshold by keeping all other parameters constant which resulted in 16 different scenarios as given in Table 7.1.

Table 7. 1 ISODATA clustering algorithm scenarios

No. of Scenarios	Maximum Iterations	Convergence Threshold			
Scenario 1 – 4	5	25	50	75	100
Scenario 5 – 8	10	25	50	75	100
Scenario 9 – 12	15	25	50	75	100
Scenario 13 – 16	20	25	50	75	100

The Maximum Iterations (MI) were varied from 5 to 20 with a successive increase of 5, whereas the Convergence Threshold (CT) ranged from 25 to 100 percent with a successive increase of 25%. The other parameters were kept constant as mentioned in Table 7.2.

Table 7. 2 Values of constant parameters

Sr. No	Parameter	Value
1	Minimum no. of classes	2
2	Maximum no. of classes	5
3	Minimum size percent	0.01
4	Maximum standard deviation	5

5	Minimum distance	4
6	Maximum merge	1

7.2.5. Accuracy assessment

The accuracy of each classified map generated from each scenario was assessed by using 100 random generated points all over the study area (Figure 7.2). The classified map was overlaid on the original satellite image and through visual interpretation, the total number of true classified points and a total number of misclassified points were calculated in terms of their percentages. The higher the number of percentages of truly classified points indicate the most suitable ISODATA clustering algorithm scenario and vice versa.

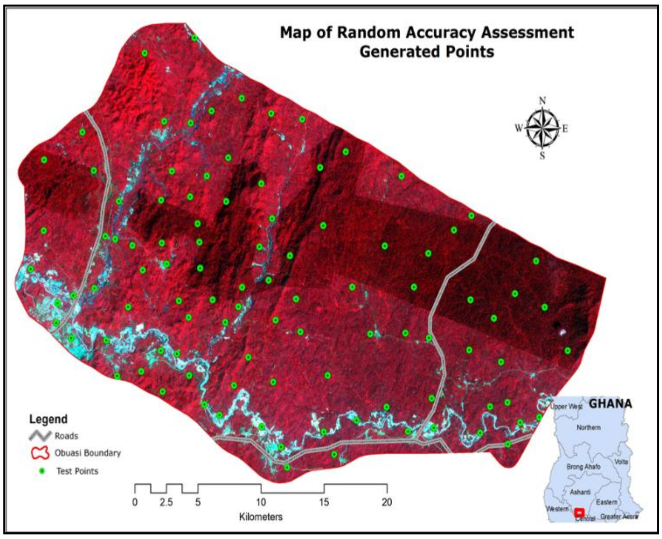


Figure 7. 2 The random generated 100 points all over the study area

7.3. Results and Discussion

The results of Histogram Equalization processing showed that there was a significant enhancement in the satellite image as shown in Figure 7.3 (A) before and (B) after. The three spectral bands Near Infrared, Red, and Blue were highlighted as Red, Green, and Blue respectively.

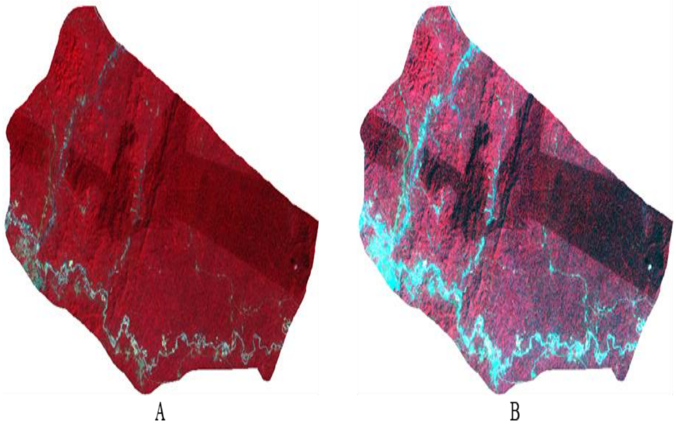


Figure 7. 3 Effect of histogram equalization before (A) and after (B) on the satellite image

The light red colour represents the forest area whereas the dark red patch represents the dense forest cover. The cyan colour represents the disturbed land cover in terms of built-up and gold mining activities.

The feature space analysis also revealed that the spectral bands Blue and Near-Infrared showed the maximum and Blue and Green showed the minimum variations as highlighted in Figure 7.4. This is a clear indication that for ISODATA cluster algorithm the number of spectral bands available in the satellite image was most suitable for its application.

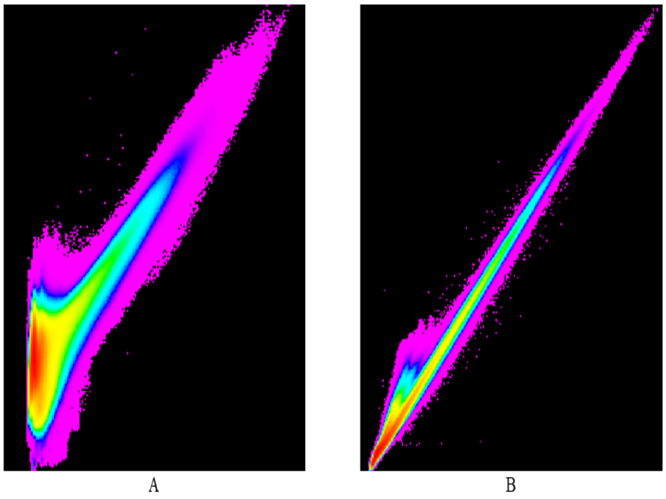


Figure 7. 4 Feature space of Blue – Near-Infrared (A) and Blue – Green (B) spectral bands

The ISODATA clustering results showed that the lower value of Maximum Iterations (MI) and higher value of Convergence Threshold (CT) yield the most suitable classification and hence mapping of gold mining regions. The mapped gold mining regions and classification results of all ISODATA clustering algorithms scenarios are given in Figure 7.5. As per the results the scenarios-1, 5, 9 and 13 resulted in only two classes, hence generated a general classified map of the area with no identification of gold mining regions.

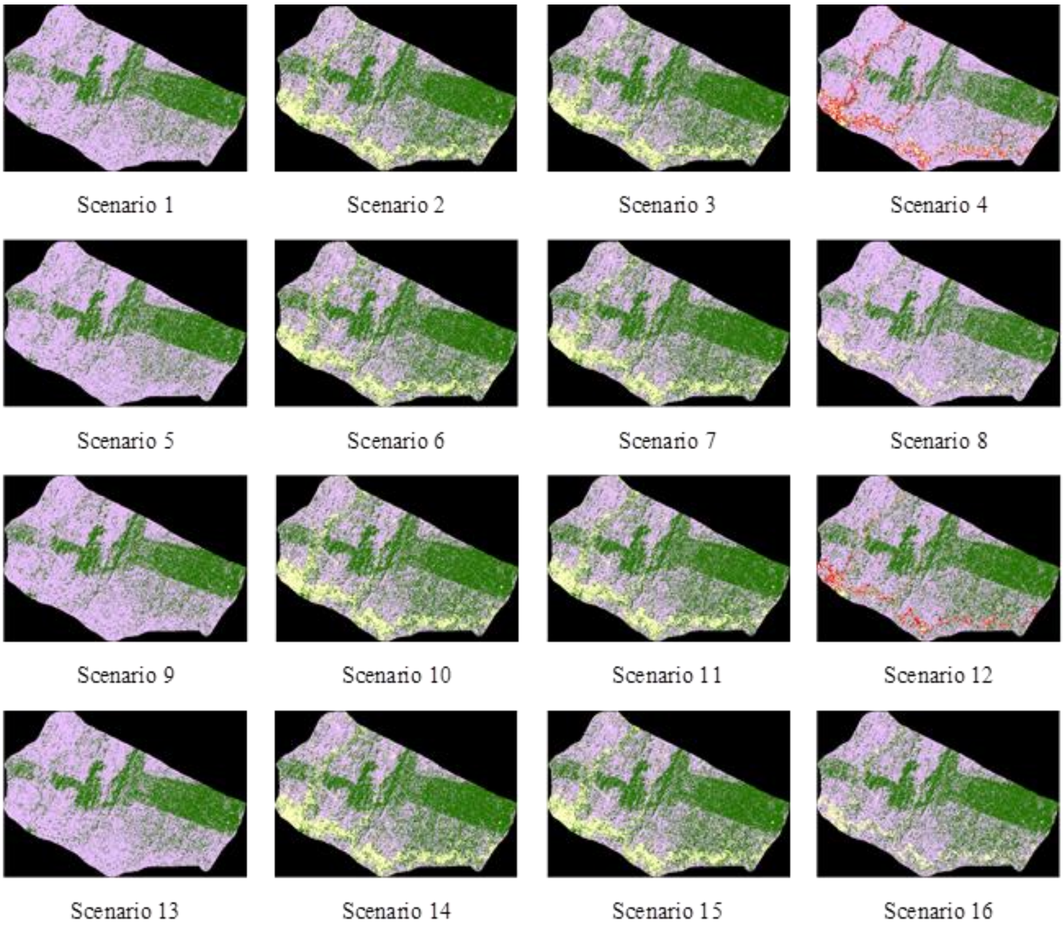


Figure 7. 5 Results of ISODATA Cluster Algorithm based gold mining region along with other classes. Whereas Fir-Green colour represents the dense forest cover, Lepidolete Lilac (purple shades) represents the forests, Poinsettia Red

colour represents the infrastructure and Yucca Yellow represents the gold mining areas.

In the scenarios-1, 5, 9 and 13 the MI was 5, 10, 15 and 15 respectively whereas the CT was the same as 25%. The lower CT value with all MI values not only resulted in the general classified maps but also not suitable for mapping of surface gold mines in the study area. Whereas the scenarios-2-3, 6-7, 10-11 and 14-15 resulted in almost similar kind of results in the form of three classes as gold mining areas, dense forest cover and light vegetation. The MI for all these scenarios were 10 and 15 whereas the CT were 50% and 75%. This means that after the development of clusters only 50% and 75% pixels can remain the same in each cluster as pure pixels with iterations from 10 to15 times. Scenarios 8 and 16 also resulted in three classes as gold mining areas, dense forest, and light vegetation. The MI for these were 20 with a CT value of 100%.

The most suitable scenarios for mapping of surface gold mining activities along with other land cover classes were 4 and 12. The MI and CT for scenario-4 was 5 and 100%, whereas for scenario-12 was 15 and 100%. In both of the scenarios the CT was 100% which means that after generation of clusters, the pixels were not allowed to shift from one cluster to another. There was a total of four classes including gold mining areas, dense forest, light vegetation and infrastructures present in the study area. Even among these two, the scenario-4 performs much better than 12 on the basis of accuracy assessment. The research conducted by Swenson et al. (2011) used the similar ISODATA clustering algorithm for mapping of gold mining activities in the study area and with similar parametrization. Another research study conducted by Sunuprapto et al. (2016) also used the same parameters to map the gold mining in the region.

Several other research studies also showed that the ISODATA Clustering Algorithm generate better results with low to medium MI and higher CT values. The research studies conducted by Ding et al. (2018), Verma et al. (2017) and Mahboob; et al. (2015) and Sunuprapto et al. (2016) also support the findings of this study in terms of resulted maps based on ISODATA Clustering classification algorithms.

The accuracy of all the ISODATA Clustering Algorithm scenarios is given in graph Figure 7.6. As per the analysis, only 10 out of 13 scenarios was successful for the identification of gold mining regions. Among these 10, only 2 scenarios i.e., scenario-4 and 12 have identified the gold mining regions most accurately, particularly the former one. For other land covers like dense forest cover, other vegetation, and infrastructure, scenario-13, scenario-1 and scenario-12 were the most suitable scenarios respectively, with highest correctly classified pixels in each class.

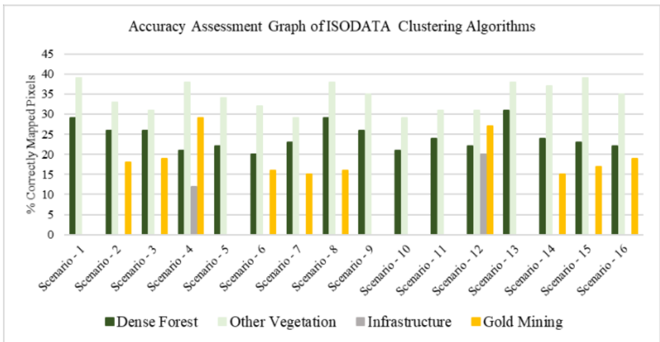


Figure 7. 6 Percentage of correctly classified/mapped pixels in each class for all the ISODATA Clustering Algorithm scenarios

7.4. Conclusion

Sixteen (16) different ISODATA Clustering Algorithm scenarios were designed with variations in Maximum Iterations (MI) and Convergence Threshold (CT) parameter values, in order to correctly map the surface gold mining areas. The Sentinel satellite data was pre-processed and enhanced using histogram equalization techniques and were used as the base image. Total of 100 random generated points was used for assessment of correctly classified pixels of each class in each scenario. The results showed that the lower values of MI (5) and higher values of CT (100%) yielded the most correctly classified surface gold mining map in scenario 4. The dense forest cover, other vegetation, and infrastructure present in the study area was also correctly classified in scenario-13, scenario-1 and scenario-12 respectively.

The variations in the MI and CT resulted in different degree of accuracy, which can be further enhanced by micro-tuning of other constant parameters used in this research. The evaluation of ISODATA Clustering Algorithms was performed in order to assess their capability to detect and identify gold mining activities and related other land covers classes in the study area. The methodology developed in this research is not only robust but also cost and time effectively by using satellite remote sensing images in order to monitor and map the surface gold mining activities in the regions particularly in the dense forest covered areas.

8. HARNESSING THE POWER OF MACHINE LEARNING FOR MAPPING OF IRON OXIDE MINERALIZATION USING SENTINEL SATELLITE SPECTRAL IMAGERY

This chapter is based on the following publication

Mahboob, M. A., Genc, B., & Atif, I. (2022, August). Harnessing the Power of Machine Learning for Mapping of Iron Oxide Mineralization using Sentinel Satellite Spectral Imagery. In *2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (pp. 285-289). IEEE

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

8.1. Introduction

Exploration of earth's natural resources through remotely sensed satellite datasets has been practised and verified by various research studies (Adiri et al., 2020b; Mahboob et al., 2019; Pendock & McKay, 2019). The ever-increasing demand for iron mineral makes its regional exploration more vital, particularly after the Brumadinho tailings dam disaster which occurred on 25 January 2019 at the Córrego do Feijão iron ore mine (Atif et al., 2020). However, due to time and cost constraints, the classical mineral exploration techniques cannot be used for vast area discoveries. Remote sensing satellite images can be used effectively and robustly for mineral explorations with iron oxides (Adiri et al., 2020b; Booysen et al., 2019), particularly for larger areas. Several research studies have been successfully applied remote sensing techniques for the mapping of iron mineral, e.g. Ducart et al. (2016) used Landsat-8 satellite data for mapping of iron oxide using multisource classification technique which proved to be an adequate alternative for mapping iron oxides in vegetated shrub areas of Brazil. Similarly, another research study conducted by Bersi et al. (2016) used the remotely sensed satellite data for mapping of iron ore in Algeria and concluded that remote sensing is very cost and time-effective technique for this purpose.

Recent advancements in computer vision have made the image processing techniques very suitable to process satellite datasets. Many researchers (Farahbakhsh et al., 2020; Farahbakhsh et al., 2018; Vassilas et al., 2002) have developed computer vision-based frameworks for mineral explorations and geological features mapping using the satellite datasets. However, to apply the computer vision-based techniques effectively, there is a requirement of high computational power due to processing of big datasets (Fink, 2018). This problem can be overcome by using Graphical Processing Units (GPU) clusters instead of Central Processing Units (CPU). The research study conducted by Liu et al. (2019) used the multilevel parallel processing for time-series quantitative assessment of satellite data using the GPU based clusters and concluded that using GPUs increases the efficiency of the models and the results.

This research study aims to analyse three computationally intensive digital image processing techniques for efficient mapping of iron oxide deposits in the Northern Cape region of South Africa using multispectral and multi-spatial Sentinel-2 satellite data. Three advance computer vision techniques, i.e., Spectral Band Combination (SBC), Spectral Band Ratio (SBR) and Spectral Angle Mapper (SAM) were applied to map the iron deposits in the study area.

8.2. Materials and Method

8.2.1. Satellite data

In this research, the raw Sentinel-2 satellite data was acquired and pre-processed for further analysis. The multispectral channels of Sentinel-2 satellite are suitable for mapping of different environmental features on the surface of the earth, including several minerals like iron oxides. The multispectral channels were stacked in three groups resulted in three images based on spatial resolutions, as shown in Table 8.1. The initial pre-processing of data, i.e., histogram equalization, was done to enhance the images in terms of spectral properties for each spectral channel and the detailed methodological framework is given in Figure 8.1.

Table 8. 1 Spectral and spatial characteristics of sentinel-2 satellite data

Sr. no of Image	Spectral Resolution	Type of Spectral channels	Spatial Resolution
Image – 1	Four channels	1 × Blue, 1 × Green, 1 × Red, 1 × NIR	10 m
Image – 2	Six channels	3 × Vegetation Red Edge, 1 × Narrow NIR, 2 × SWIR i and ii	20 m
Image – 3	Three channels	1 × Coastal Aerosol, 1 × Water Vapor, 1 × SWIR-Cirrus	30 m

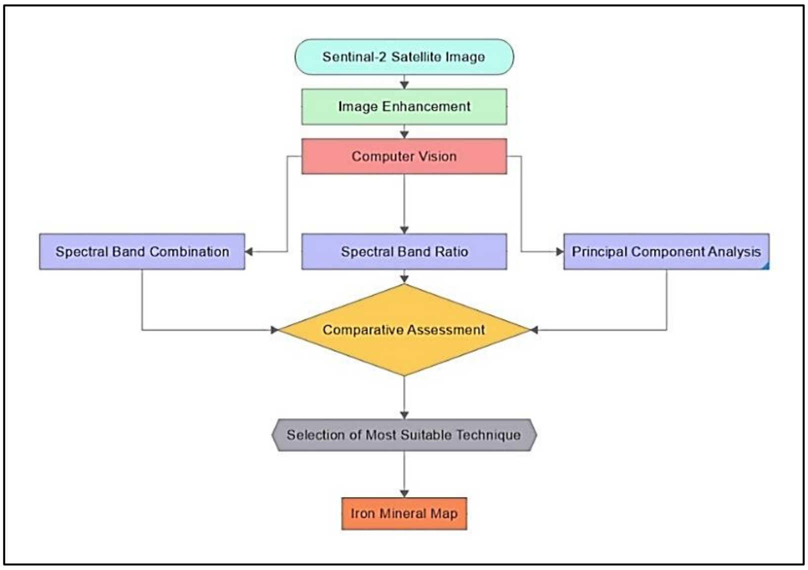


Figure 8. 1 Detailed methodological framework of the research study

8.2.2. Spectral band combination (SBC)

The SBC is the selection of most appropriate channels and then their representation in Red-Green-Blue (RGB) as shown in Figure 8.2. The selection of most appropriate channels can be made through the optimum band index depending on the absorption/reflectance characteristics along with least spectral correlation (Ducart et al., 2016; Shirazi et al., 2018). Except for the image-1, the other two images with 20m and 30m resolution produced the false colour composites because of the absence of visible spectral channels.

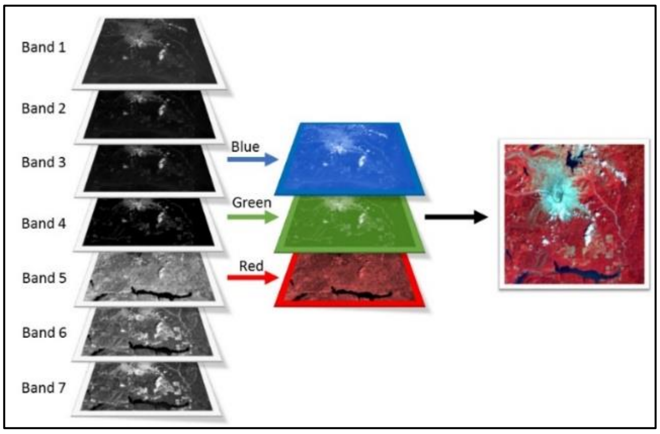


Figure 8. 2 The architecture of band combination technique as highlighted with Red, Green and Blue, Source: Montgomery (2016)

8.2.3. Spectral band ratio (SBR)

The SBR is another advance computer vision technique in which the pixel values of one channel is divided by the other channel and normalize sometimes (Crippen, 1990). The resultant new image is usually in the greyscale ranging between -1 (dark) to +1 (bright), as iron oxide deposits. The new greyscale images can then be combined as RGB and represent the enhanced identified spatial features.

8.2.4. Spectral angle mapper (SAM)

The SAM is an image classification technique, where the spectral profile of the target feature from satellite data is compared to the reference spectral profile. The

SAM compares the similarity between the target and references spectra by considering them as a vector separated by a spectral angle, as shown in Figure 8.3. The angle (cosine ɑ) between the two vectors varies from 0 to 1, where zero or close to zero means, the target feature is precisely the same as the reference spectra feature (Girouard et al., 2004).

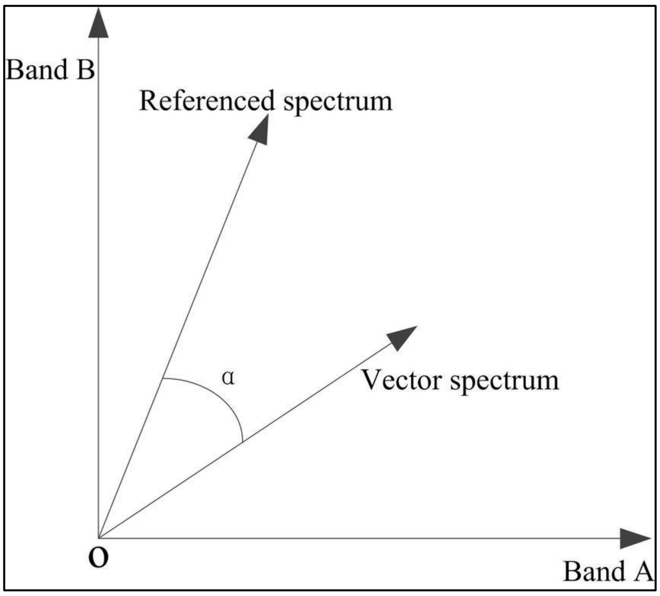


Figure 8. 3 The architecture of spectral angle mapper where the vector spectrum is separated by an angle ɑ regarding the reference spectrum. Source: Yan and Wang (2016)

The ɑ value can be calculated using the Equation 8.1, where, nB is the number of channels, t_i is the reflectance of band i of the satellite image spectrum, and r_i is the reflectance of band i of the reference spectrum.

$$\alpha = \cos^{-1} \frac{\sum_{i=1}^{nB} t_i r_i}{\sqrt{\sum_{i=1}^{nB} t_i^2}\ \sqrt{\sum_{i=1}^{nB} r_i^2}} \qquad \text{Equation 8. 1}$$

8.3. Results and Discussion

The histogram equalization enhanced the satellite image, particularly the iron ore deposits in the shades of dark to light purple, as shown in Figure 8.4.

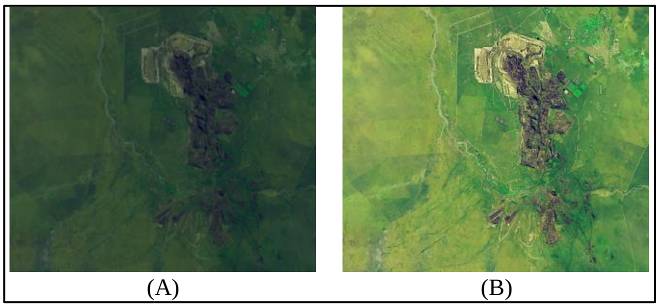


Figure 8. 4 Effect of pre-processing of satellite image as given raw data (A) and processed data (B)

The first computer vision technique, spectral band combination was applied, and nine combinations was generated (Table 8.2) as a set of three against each image, as shown in Figure 8.5 (A-I).

Table 8. 2 Spectral band combinations as applied against red, green and blue

Image Number	**Red**	**Green**	**Blue**
A	NIR	RED	GREEN
B	GREEN	BLUE	NIR
C	RED	NIR	BLUE
D	NARROW NIR	SWIR-1	SWIR-2
E	SWIR-1	RED EDGE	NARROW NIR
F	SWIR-2	SWIR-1	RED EDGE
G	SWIR-CIRRUS	COASTAL AEROSOL	WATER VAPOR

H	COASTAL AEROSOL	WATER VAPOR	SWIR-CIRRUS
I	WATER VAPOR	SWIR-CIRRUS	COASTAL AEROSOL

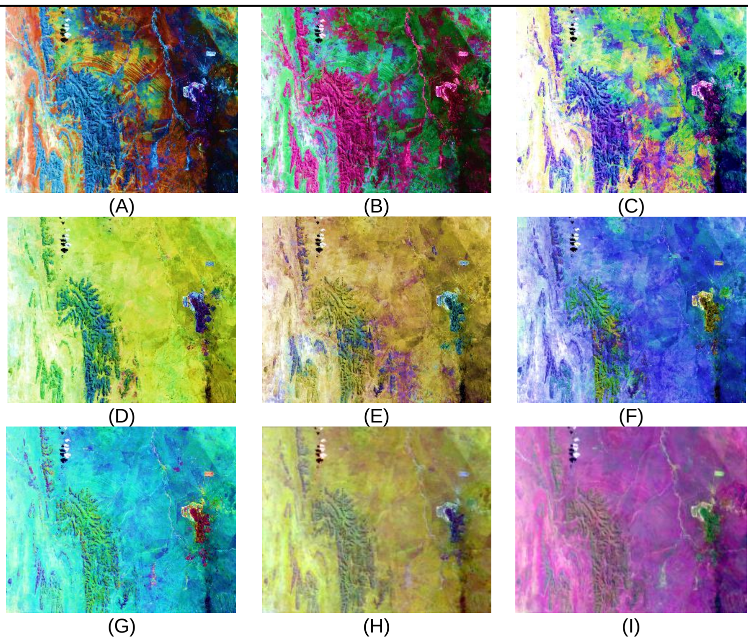


Figure 8. 5 Iron oxide enriched areas highlighted because of different spectral band combinations

From Figure 8.5, the results of SBC for image-1 (A-C) was not satisfactory because of the minimum-reflectance of iron ore in the visible and NIR portion of the electromagnetic spectrum (Ciampalini, Garfagnoli, Del Ventisette, et al., 2013). However, the results of SBC for image-2 (D-F) and image-3 (G-I) were more reasonable, whereas the image-2 was more significant. SWIR spectral channels in these two images can be the primary contributor for iron ore identification. Although image2 and 3 produced more enhanced results for iron ore deposits, their spatial resolutions were coarser than the image-1. Similarly, several SBR was developed and applied on all the three images, as shown in Figure 8.6 (J-R).

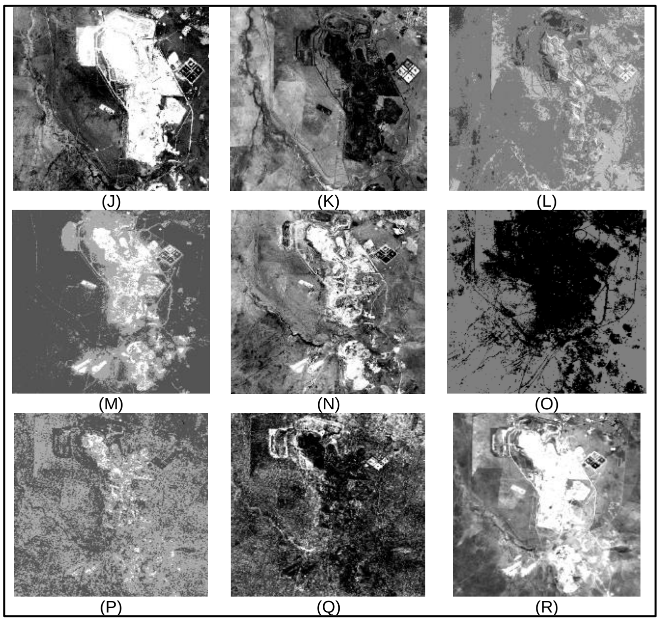


Figure 8. 6 Iron oxide enriched areas highlighted because of different spectral band ratios

The three-band ratios of image-1 with low spectral and high spatial resolution revealed that Red/NIR (J) and NIR/Green (K) ratio images have a high potential for mapping of iron oxide mineralized zones. For image-2, once again the SWIR spectral region revealed the high potential to iron oxide mineralization as only one band ratio SWIR-1/SWIR-2 (N) highlight the iron deposits as bright tones. In image-3 band ratios, only one ratio image containing the SWIR band (R) highlight the iron deposits.

For spectral angle mapper, the reference spectral library used for SAM analysis was acquired from global mineral spectral profiles (Rossel et al., 2016). Figure 8.7 showed hematite and goethite spectral profiles related to iron oxide along with the spectral profile of Sentinel-2 satellite data.

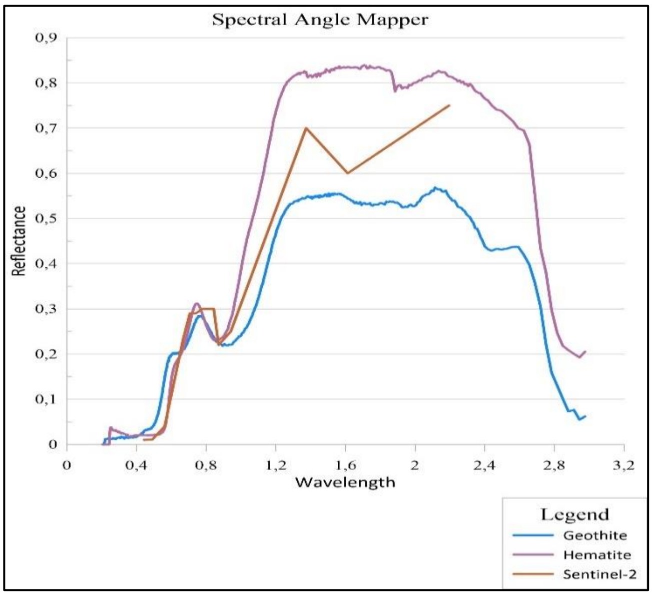


Figure 8. 7 Spectral reflectance of laboratory and satellite data from visible to shortwave infrared wavelengths (Source: USGS Spectral Library)

The maximum similarity index between the reference and satellite spectra were 0.15 and based on this similarity index; the potential iron ore deposits were mapped from Sentinel-2 satellite data and shown in electron gold colour in Figure 8.8.

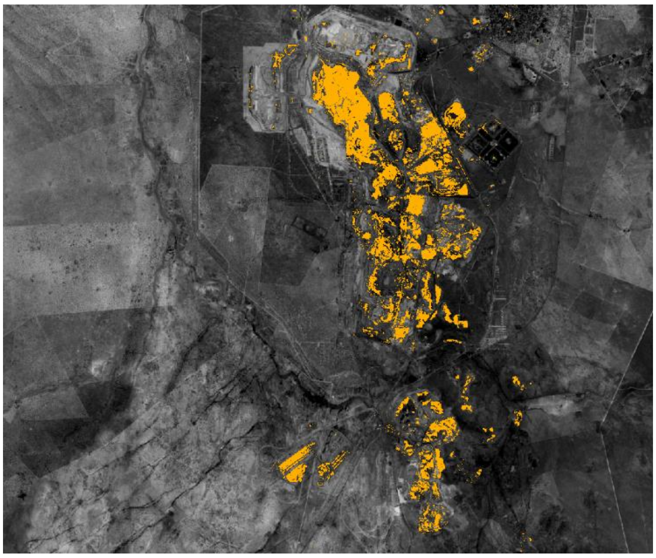


Figure 8. 8 Iron oxide mineral enriched areas mapped through spectral angle mapper

The high spatial resolution 10m spectral channels of Sentinel-2 satellite was not proved as significant for iron oxide deposit mapping in any of computer vision techniques as compared to those with 20m and 30m resolution. This can be due to the absence of SWIR spectral channel in image-1, which is one of the crucial spectral band for iron mineral identification (Soe et al., 2005). In SBC technique, the spectral combinations result in Figure 8.5 E, G and I were the most appropriate for iron ore deposits generated from image-2 and image-3 with the presence of SWIR spectral channel. For SBR image processing technique, the iron ore deposits were highlighted as brighter pixel values in Figure 8.6 (J and N) with values greater than 0.71. Overall, the performance of SAM proved more promising for identification and mapping of iron ore deposits present in the study area with an overall accuracy of 81.62%. However, the spectral information of Sentinel-2 satellite was not sufficient as compare to reference global mineral spectra profiles, because of the presence of a smaller number of spectral channels. The studies conducted by Ciampalini, Garfagnoli, Del Ventisette, et al. (2013) and Shirazi et al. (2018) also concluded that SAM is the efficient and robust image processing technique in computer vision which can be used for mapping of iron mineral as compare to the other techniques. The parallel processing of datasets with graphical processing units helps to accelerate the process and results were generated in all computer vision techniques.

8.4. Conclusion

The three advanced computer vision techniques (SBC, SBR, and SAM) were applied on Sentinel-2 satellite data. Overall, SAM performed more efficient in terms of accurately mapping of iron ore deposit as compare to others. The SWIR spectral band is the most important for iron mineral prospectively zonation besides other spectral channels. The cost-effective methodology described in this research can be applied in other regions with the same geographical and geological characteristics to explore iron mineral resources.

9. PREDICTIVE MODELLING OF MINERAL PROSPECTIVITY USING SATELLITE REMOTE SENSING AND MACHINE LEARNING ALGORITHMS

This chapter is based on the following publication

Mahboob, M. A., Celik, T., & Genc, B. (2023). Predictive Modelling of Mineral Prospectivity Using Satellite Remote Sensing and Machine Learning Algorithms. Mining, (under review).

The author of this thesis conceptualized the work, performed the experimental measurements, analysed the data and wrote majority of the paper.

9.1. Introduction

Mineral prospectivity mapping or modelling (MPM) is a technique for identifying and ranking areas which are likely to contain undiscovered mineral deposits of a particular type. This is accomplished by establishing a correlation between geological features (input variables) and the presence of the target mineral deposits (output variables) (Yin & Li, 2022). This technique typically involves these steps: (i) obtaining geospatial information from various geoscience data sources; (ii) determining exploration criteria that effectively depict processes essential for the formation of the desired deposit type; (iii) creating predictor maps from spatial datasets based on the chosen exploration criteria; and (iv) combining predictor maps to generate a predictive model for locating exploration targets in unexplored regions (Zuo, 2020). Geospatial information is often gathered through field surveys, which are not only costly but also time intensive. Another important issue is the lack of access to unreachable locations and limited field samples. Satellite remote sensing, on the other hand, has proved its effectiveness and importance in identifying mineral deposits and associations by detecting spectral anomalies (Benaissi et al., 2022; Fu et al., 2023; Mahboob et al., 2019; Rajan Girija & Mayappan, 2019; Rajesh, 2004; Shirmard et al., 2022). Satellite remote sensing is a powerful technology for geologists and other mining professionals to identify hydrothermally altered rocks, structures, lineaments, lithological units, vegetation, and other valuable data. Generally, the MPM models can be distinguished based on the approach taken to assign evidential weights: (i) knowledge-driven models that use an expert-based approach to assign weights to the evidential maps, and (ii) data-driven models which use the spatial relationship between the evidential maps and known mineral deposits to calculate the weights empirically (Carranza, 2008). Over the last few decades, various data-driven modelling approaches have enabled progress in the field of MPM. Weights of evidence and logistic regression are two probabilistic methods that have become widely used due to their easy to understand and interpretation approach. Additionally, in the past few years, ML algorithms, developed mainly by computer experts for solving multi-field and multi-dimensional problems of classification and pattern recognition, have been seen as promising tools for creating predictive mineral prospectivity modelling and mapping (Jung & Choi, 2021; Köhler et al., 2021; Mahboob et al., 2022; Rodriguez-Galiano et al., 2015; Shirmard et al., 2022; Sun et al., 2020). ML algorithms such as SVM and RF have demonstrated superior predictive performance when compared to traditional statistical techniques or empirical models, especially when dealing with complexly distributed input features and nonlinear mineralization associations (Geranian et al., 2016; Santos et al., 2022; Sothe et al., 2020). Recently, deep learning, a powerful branch of ML, has been widely successful across many scientific domains. By learning hierarchical representations of input data, deep learning can recognize complex patterns and accurately classify them. Several geoscience applications, such as land subsidence susceptibility mapping, geochemical mapping, meteorological modelling, and environmental degradation modelling have been using these advantageous algorithms (Mahboob et al., 2022). The research conducted by Xiong and Zuo (2020) applied the combination of SVM and deep learning models for accurate mineral exploration using geochemical field data. Another research conducted by Kong et al. (2022) applied Student Teacher Ore-induced Anomaly Detection (STOAD) deep learning model for mineral predictions mapping. The study concluded that the model performed exceptionally well and accurately predicted the location of target deposits. The research conducted by da Silva et al. (2022) showed the benefits of applying machine learning model RF on the drillhole data for accurate gold predictions in unexplored regions.

However, deep learning has been rarely implemented in the domain of MPM. Numerous algorithms have been proposed for MPM, but none of them is superior in all situations. Thus, a comparative analysis of multiple predictive algorithms is also necessary for an effective data-driven MPM.

The North Waziristan district, part of the Khyber Pakhtunkhwa province of Pakistan, was selected as a case study area for the ML based data-driven MPM. The region, which once comprised the Federally Administrated Tribal Areas (FATA), is a mountainous territory bordered with Afghanistan and located south and west of Islamabad, Pakistan. It is well-known for its natural resources and minerals, specifically its abundant Cu deposits (Mehsud, 2012). Since 2001, the region has been facing a severe law and order situation mainly due to geopolitical instability (Spychała-Kij, 2020). In recent years the country's geological survey department undertook geological mapping along with historical records, and it was discovered that huge amounts of Cu are located at Shinkai and Degan in North Waziristan. Estimates of these Cu ore-reserves are 122.71 million tons with a Cu content ranging from 0.3865% to a maximum of 2% (Khan, 2000). However, due to the poor law and order situation in the region, there were very limited efforts made to fully explore and extract the mineral resources. A comprehensive and reliable map of potential mineralization in the area is still needed to help the government achieve their exploration goals through drilling.

9.2. Materials and Methods

9.2.1. Study Area

The study area for this research is in the Miran Shah tehsil of North Waziristan region of Pakistan and near the Afghanistan border (38 km), as shown in Figure 9.1. The region was previously known as the FATA of Pakistan and in May 2018 it has become part of the Khyber Pakhtunkhwa Province of Pakistan. In the past limited research or field surveys were conducted in the Miran Shah due to geopolitical instability in the region(Khan et al., 2022). Due to these geopolitical challenges, there is very limited information available on the geographical, geological, and economical aspects of features located in the region. After military operations, the government rebuilt the region and provided economic means to the people of Waziristan.

Figure 9.1 The map of the study area

The area is enriched with several precious metals and minerals due to the geological conditions, but the details of these mineralization were unknown. In 2015, the geological filed visits were conducted to map the geology of the study area. The region is part of the larger geological framework of the Indian Plate's collision with the Eurasian Plate, which has resulted in a complicated tectonic history and a wide range of rock formations. The geology of Miran Shah comprises mainly complex ophiolite igneous rocks which are intensely folded, faulted, fractured and brecciated in places (Figure 9.2). The ophiolite present in the study area is the part of the disjointed Tethyan ophiolites which is formed as back-arc basin and differs from mid-ocean ridges settings. The belt of the igneous rocks present in the study area extends south-west to north-east and is mainly composed of ultramafic, harzburgite, pyroxenites, quartz-diorites, micro-quartz diorites, granodiorites, dolerites and gabbros (Sothe et al., 2020). The Sulaiman Fold Belt is distinguished by substantial folding and thrusting of sedimentary strata. The deformation zone where the Indian Plate collided with and was forced over the

Eurasian Plate is represented by this belt. The volcanic fine-grained porphyritic pillow basalts and andesites with secondary breccias and agglomerates are also present. The variety of sedimentary rocks that can be found, mainly consists of Jurassic to Eocene shallow marine continental and unfolded types. Sandstone, shale, limestone, and conglomerate are examples of these rocks. North Waziristan contains sedimentary sequences such as the Datta Formation, Lockhart Limestone, Lumshiwal Formation, and Tochi Formation. In addition to sedimentary rocks, the area has metamorphic rocks formed by tectonic activity's strong pressure and heat. Phyllite, schist, and gneiss are examples of metamorphic rocks. Some of the widely distributed sedimentary rocks are quite favourable for hosting economic Cu orebodies in the region and the most prominent outcropping structures are the North-South trending geological folds (Khan et al., 2007). Large porphyry intrusions formed in the early Cenozoic (Eocene-Oligocene) zones as a result of the convergence of the Waziristan Thrust and the subsequent subduction due to geotectonic settings. Significant porphyry copper and associated hydrothermal mineral deposits are found within these porphyry intrusions. The study area experienced mineralization processes during the Jurassic to Tertiary phases of magmatism, resulting in the formation of diverse porphyry Cu and Cu-Au deposits, low and high sulfidation epithermal Au deposits, and Cu-Pb-Zn vein-type deposits. The porphyry copper occurrences show hydrothermal alteration, which comprises of various regions including the potassium silicate alteration zone (K-zone), the propylitic alteration zones, as well as irregular Argillic and regular Phyllic alteration zones (Malkani et al., 2017).

The faults and folds in this area can be grouped as North-South and North-West trending basement faults, and North-East trending cap rock faults and folds.

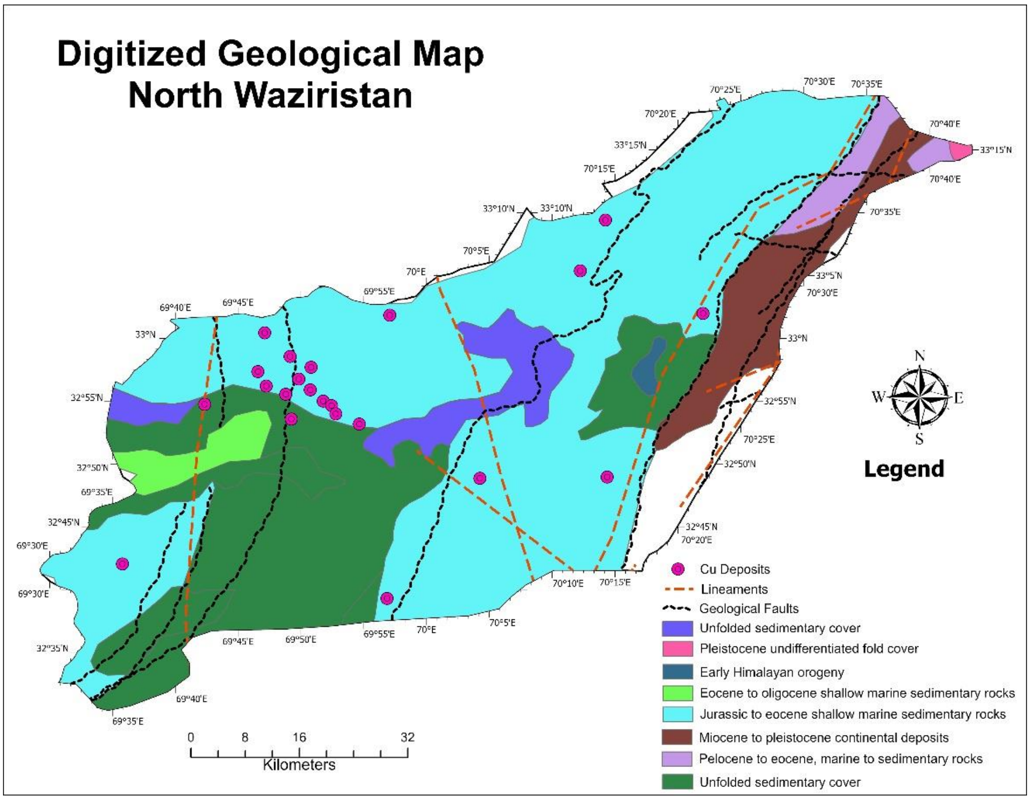


Figure 9.2 Simplified Geological map of the study area along with lithological and structural features

9.2.2. Support Vector Machine (SVM)

The SVM is a powerful ML model used to solve the classification problems. The SVM was initially introduced by Cortes and Vapnik (1995), and because of its great predictive capability, adaptability, and robustness, it has been proven useful for identifying hidden patterns in data (Lim et al., 2002). Several researchers have applied SVM in geosciences, e.g., for land cover classification (Xu et al., 2019a), Geological 3D modelling using limited data from a variety of sources (Smirnoff et al., 2008), porosity prediction in a heterogeneous reservoir (Al-Anazi & Gates, 2010), consistency analysis of geophysical datasets (Turlapaty et al., 2010) and for mineral exploration and prospectivity mapping (Abedi et al., 2012). In the SVM algorithm, all data is expressed as points in an n-dimensional space, where n is the total number of attributes in the data. The aim is to identify the hyperplane that clearly differentiates the two feature groups by classifying the n-dimensional space, while maximizing the margin between classes.

For a given dataset of training which can be separated linearly, SVM categorizes the data α in the input space S into a high dimension space D—$\alpha \in \mathbb{R}^S \mapsto \Psi(\alpha) \in \mathbb{R}^D$ with a kernel function $\Psi(\alpha)$ to search for a separating hyperplane. SVM was originally established for binary classification problems, but later on it has been applied to multiclass problems. In multiclass, SVM starts with a single class and separates from the remaining classes one at a time. Each SVM is trained to differentiate all occurrences of a single class from the occurrences of all other classes. In testing, the class label y of a class pattern x is given by equation 9.1:

$$y = \begin{cases} m, & if\ d_m(x) + t_h > 0 \\ 0, & if\ d_m(x) + t_h \leq 0 \end{cases} \qquad \textit{Equation 9.1}$$

Where $d_m(x) = max\,\{d_c(x)\}_{c=1}^{N_h}$, $d_c(x)$ is the distance from x to the hyperplane corresponding to the class c, range from 1 to N_h and t_h is the classification threshold. The two main parameters that should be considered for the optimization of SVM results are regularization and gamma (Namdeo & Singh, 2021). The regularization usually denoted by C is the factor that decides the SVM optimization regarding the avoiding of misclassification in a class. The higher value of C will cause the optimizer to select a hyperplane with a small margin and higher misclassification, whereas a low value of C will result in a high-margin hyperplane with low misclassification. The gamma factor decides the influence of single class values on the selection of hyperplane (Tuba et al., 2017). Usually, the low gamma caused the data points far away from possible hyperplane are considered, whereas the high gamma means the points close to the possible hyperplane line are considered in the calculation.

9.2.3. Convolutional Neural Network (CNN)

A CNN is a form of deep learning neural network comprising of an input layer, one or multiple convolutional layers, pooling layers, fully connected layers, and an output layer. Input data, $X = (x_1, x_2, \ldots, x_n)$, is fed into a convolutional layer, which applies a randomly initialized filter (or kernel) to execute a convolution operation.

Figure 9.3 demonstrates a general convolution, where a filter is slid across the input data, calculating the dot products between the filter and the input, which results in a new feature dataset (a feature map).

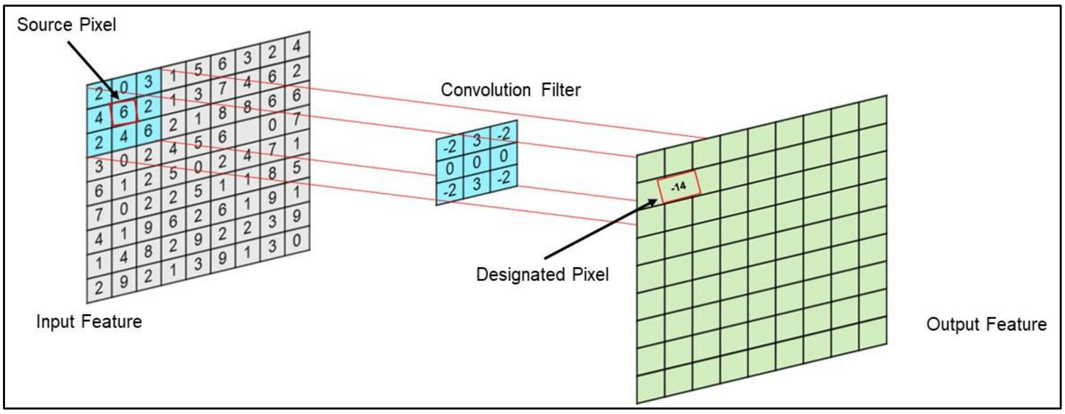


Figure 9.3 General simplified architecture of convolutional neural network (conceptualized from source Robinson).

Applying different $\sum i$ representing kernels yields multiple convolutional outputs, resulting in a set of feature maps that represent the input vectors in a hierarchy as shown by equation 9.2 derived by Imamverdiyev and Sukhostat (2019).

$$C_i = (c_1, c_2, \dots, c_n) = k(X \cdot \Sigma_i) \qquad \text{Equation 9.2}$$

Whereas C_i represents the ith convulational output with $c_1, c_2, \dots, c_n$ as this elements. X is the input data and k is a scaling factor applied to the convolution operation. Using the activation function k, nonlinear amplification of convolutional results is achieved (Ghorbanzadeh et al., 2019). The most widely used activation function in CNNs is the Rectified Linear Unit (ReLU), expressed as equation 9.3.

$$k(x) = \max(0, x) \qquad \text{Equation 9.3}$$

A pooling layer receives the outputs of one or more convolutional layers. Pooling is a filter-based technique for obtaining the maximum (max-pooling) or average (average-pooling) value from feature maps. This research uses max-pooling to downsample the data and reduce network dimensions, hence limiting overfitting and lowering computing costs. Features are extracted after the convolutional and pooling layers have processed the input data. Ultimately, the output of the fully connected layers is transmitted to the final layer, which employs a softmax activation function to compute the probability of each class according to the formula given in equation 9.4 (Adem, 2022).

$$P(l = m|x) = \frac{\exp(w_m x + b_m)}{\sum_{i=1}^{2} \exp(w_i x + b_i)} \qquad \text{Equation 9.4}$$

The predicted class l is determined by the weights and bias terms w and b, respectively. The m represents the class lables and x is the input data. The input data of ML in MPM can be considered as an image, with each pixel representing a measurement of a certain attribute. Consequently, each grid cell is represented by a column vector whose length is equal to the number of features discernible, as a result, CNN with one-dimensional data representation was also used in this research (Krupski et al., 2021). The final layers of a CNN are fully connected layers that can transform feature maps into a column vector and classify them into categories or use them as a feature vector for further processing.

9.2.4. Random Forest (RF)

Breiman (2001) developed RF ML algorithm which makes repeated predictions of a single phenomenon by combining multiple decision trees. Typically, the RF generates a huge number of decision trees that are made up of several subsets of the original training dataset based on the bootstrap aggregating (or bagging) sampling method (Wang et al., 2016). After that, the selection and implementation of evidential layers will usually be done at each node of the tree to diversify and keep the forest growing. In this research, the root node was divided into several leaf nodes in each decision tree and finds the best one which contributes to the most accurate of the resultant trees. Several parameters can be considered in RF to measure the accuracy of the resultant trees, the most common ones are Gini index (I_G), gain ratio and Chi-square (Wijaya et al., 2022). In this research, the I_G was applied to calculate the purity of the leaf nodes compared to their root nodes as per the following equation 9.5.

$$I_G f = \sum_{i=1}^{n} f_i \,(1 - f_i) \qquad \text{Equation 9.5}$$

Where f_i can be described as the probability of class i at node n, and can be calculated as equation 9.6:

$$f_i = \frac{m_j}{m} \qquad \text{Equation 9.6}$$

Where the m_j is the number of samples related to class j, and m is the number of total samples present in a particular node.

9.2.5. Predictor Maps

The predictor maps were developed from the spatial data collected through the field survey, published reports, maps and remotely sensed satellite imagery. A total of nine predictor maps were developed as the input datasets to ML models. The details are given in Table 9.1.

Table 9.1 The evidential features applied in the research for the mineral prospectivity mapping

Serial. No	**Evidential features**	**Description**
1	Geological Map	Delineation of suitable lithological contacts
2	Geological Lineaments	Proximity to faults and folds
3	Intrusive Contacts	Proximity to inferred intrusive contacts identified based on the resistivity anomalies

4	Iron-oxide Alteration	Proximity to iron-oxide anomalies delineated from Sentinal-2 Satellite data
5	Argillic Alteration	Proximity to argillic anomalies delineated from Sentinal-2 Satellite data
6	Phyllic Alteration	Proximity to phyllic anomalies delineated from Sentinal-2 Satellite data
7	Propylitic Alteration	Proximity to propylitic anomalies delineated from Sentinal-2 Satellite data
8	Topographic Elevation	Surface height of the deposits from the mean sea level extracted from space based digital elevation model
9	Topographic Slope	Steepness of deposits relative to the horizontal plane delineated through space based digital elevation model

An understanding of complex mineral systems was used for the generation of predictive maps based on publicly available datasets of geologic features and satellite-derived hydrothermal alteration. These datasets are used as spatial approximations to represent the source, transport, sink, and depositional processes essential to Cu ore formation. The geology of the area (1:2,000,000 and 1:2,040,000 geological maps collected by the Geological Survey of Pakistan), is used to study various geological features such as structural zones, bedrock faults, lineaments formation and blanket fault intersections as shown in Figure 9.2.

A multizone buffer analysis was performed to obtain numerical representations of the proximity of known Cu deposits to intrusive contacts (Figure 9.4).

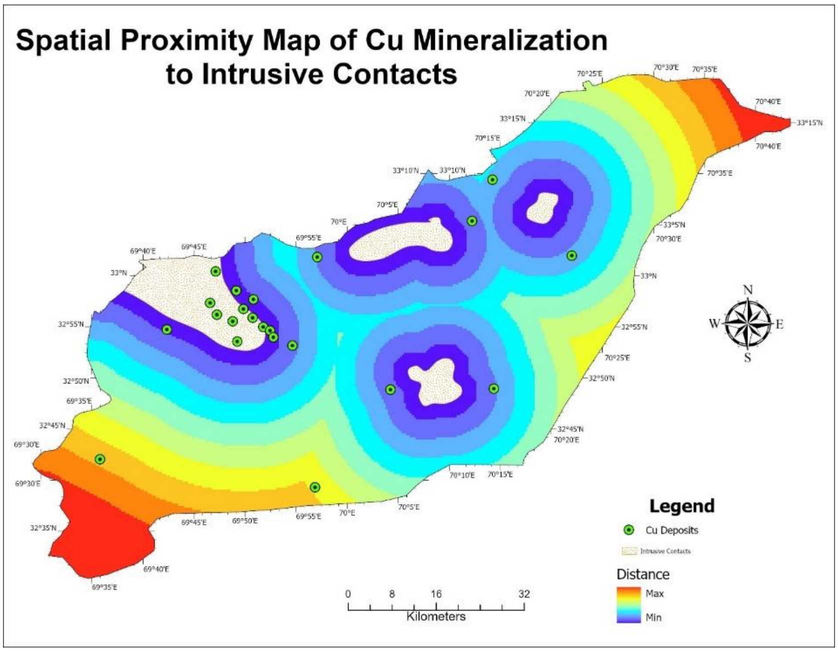


Figure 9.4 Spatial proximity map of Cu mineralization to intrusive contacts

Figure 9.5 shows a fault density map that was produced to highlight the wide dispersion of caprock faults, North-South trending faults and lineaments intersections.

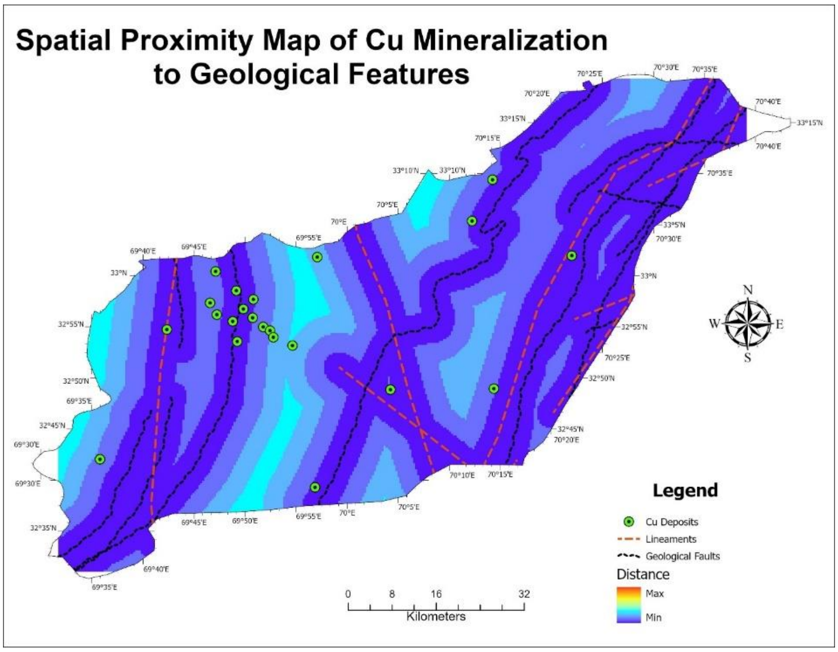


Figure 9.5 Spatial proximity map of Cu mineralization to geological faults and folds

In addition to field geological data, remote sensing satellite imagery from Landsat-8 (acquired in October 2019) and Sentinel-2 (acquired in October 2019) were utilized to identify and map hydrothermal alteration zones (Sekandari et al., 2020; Tompolidi et al., 2020; Van der Werff & Van der Meer, 2016) and lithological components related to Cu mineralization in Waziristan, on regional, local, and district scales. The ASTER satellite has become known as a prominent tool in the field of satellite remote sensing for mineral exploration, particularly in the domain of hydrothermal mineral mapping and exploitation. However, the SWIR sensors on ASTER had been dysfunctional since 2008 due to a fault in the cryocooler (Shimoda & Kimura, 2017). This particular situation has necessitated the emergence of alternate remote sensing data sources for mineral exploration, which are Landsat and Sentinel satellite series. The technical characteristics of the Landsat-8 and Sentinel-2 sensors are shown in Table 9.2. Both the Landsat-8 and Sentinel-2 satellites have prospective applications and benefits for mineral exploration research (Blandine et al., 2023; Chen et al., 2022; Mahboob et al., 2019). This relates mostly to their good multispectral imaging capabilities, sensitivity to many minerals, including copper minerals, reasonably high spatial resolution, worldwide coverage, and open data policy. They provide essential data for the detection and mapping of various mineral resources, including copper. Specific sections of the electromagnetic spectrum, such as (0.64 to 0.68 μm), are sensitive to the reflectance characteristics of iron oxide minerals including hematite or goethite, which have a strong association with copper presence. The near infrared band (0.84 – 0.881 μm) can also be used to differentiate between different rock types associated with copper deposits. Because of their mineralogical compositions, copper-bearing rocks frequently show distinct spectral signatures in the NIR region. Minerals that contain copper, such as bornite or chalcopyrite, indicate unique absorption characteristics within the short-wave infrared (SWIR), which covers about 1.57 to 2.323 μm range. By analysing the reflectance values within a specific range of the electromagnetic spectrum, geologists are able to identify regions exhibiting increased absorption, which serves as an indicator of the existence of copper minerals (van Leeuwen, 2009).

Table 9.2 Spectral and spatial characteristics of Landsat-8 and Sentinel-2 satellite imagery

LANDSAT – 8 Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS)				**SENTINEL-2 Multispectral Instrument (MSI)**			
Band No	Band Name	Spectral Range (μm)	Spatial Resolution	Band No	Band Name	Spectral Range (μm)	Spatial Resolution
B1	Coastal aerosol	0.43-0.45	30	B1	Coastal aerosol	0.433-0.453	60
B2	Blue	0.45-0.51	30	B2	Blue	0.458-0.523	10
B3	Green	0.53-0.59	30	B3	Green	0.543-0.578	10
B4	Red	0.64-0.67	30	B4	Red	0.650-0.680	10
				B5	Vegetation red edge	0.698-0.713	20
				B6	Vegetation red edge	0.733-0.748	20
				B7	Vegetation red edge	0.773-0.908	20
B5	Near Infrared (NIR)	0.85-0.88		B8	Near Infrared (NIR)	0.848-0.881	10

			30	B8a	Narrow Near Infrared (NIR)	0.931-0.958	20
				B9	Water vapor	1.338-1.414	60
B9	Cirrus	1.36-1.38	30	B10	Cirrus	1.539-1.681	60
B6	Short wave infrared (SWIR 1)	1.57-1.65	30	B11	Short wave infrared (SWIR 1)	2.072-2.312	20
B7	Short wave infrared (SWIR 2)	2.11-2.29	30	B12	Short wave infrared (SWIR 2)	2.081-2.323	20
B8	Panchromatic	0.50-0.68	15				
B10	Thermal Infrared (TIRS) 1	10.6-11.19	100				
B11	Thermal Infrared (TIRS) 2	11.50-12.51	100				

This is because remote sensing data has been widely and efficiently used for mineral exploration (Adiri et al., 2020a; Benaissi et al., 2022; Mahboob et al., 2019; Rajan Girija & Mayappan, 2019; Soydan et al., 2021). Analysing satellite data to determine the geographical locations of hydrothermally altered rocks can help identify the main outflow zones of hydrothermal systems, potentially leading to the discovery of mineral deposits. Although Cu or any other metallic mineralization can be rarely identified directly through any existing remote sensing techniques, certain minerals such as iron oxides and clays detected through their spectral signatures in the visible and near/shortwave/thermal infrared portion of the electromagnetic spectrum can be used to indicate the presence of hydrothermal alteration zones, which are associated with Cu occurrence (Atwizukye, 2022; Pour & Hashim, 2012).

The Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) technique, together with the sub-arctic summer (SAS) and Maritime aerosol models, were used to eliminate atmospheric attenuation interference on satellite data, converting it from sensor radiance to surface reflectance using PCI Geomatics code. No atmospheric adjustment was applied on Landsat-8 TIR data and hence used with the original brightness values. Following that, the pre- and post-processing results of Landsat-8 and Sentinel-2 satellite imagery were compared as shown in Figure 9.6.

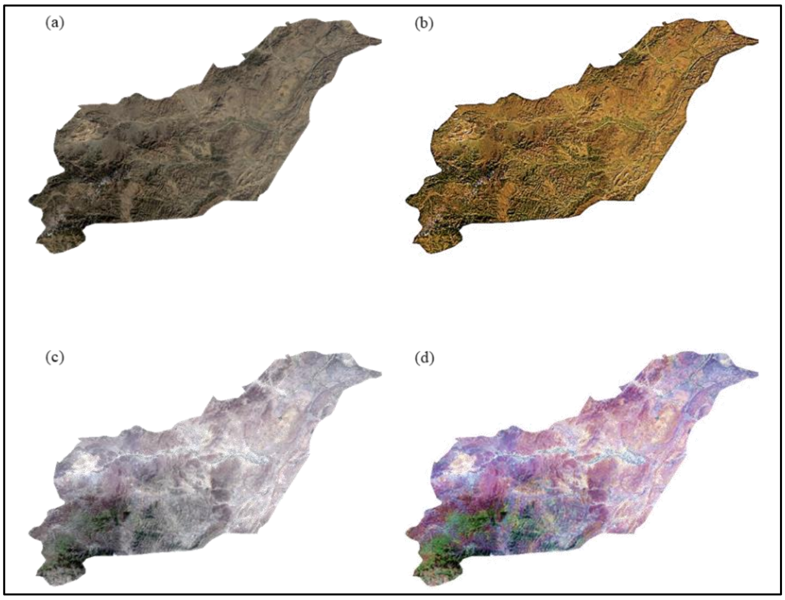


Figure 9.6 Digital image processing of satellite imagery (a) raw Landsat-8 image; (b) pre-processed Landsat-8 image; (c) raw Sentinel-2 image; (d) pre-processed Sentinel-2 image. The illustration is for False Colour Composite (FCC) highlighting Band 3 as Red, Band 4 as Green and Band 2 as Blue.

The VNIR (Visible Near Infrared) and SWIR (Short wave infrared) spectral regions have already been highlighted in terms of their importance and high potential for mapping alteration materials in considerable detail by several researchers (Atwizukye, 2022; Benaissi et al., 2022; Mahboob et al., 2019; Pour & Hashim, 2012; Rajan Girija & Mayappan, 2019). In the VNIR wavelength range, the Sentinel-2 data includes nine bands, while the Landsat-8 data has four bands. However, in the SWIR wavelength range, both data sets have two bands. As a result, Sentinel-2 and Landsat-8 multispectral data have the potential to effectively map various hydrothermally altered minerals.

Several hydrothermal mineral alterations associated with Cu mineralization such as Iron-oxide alteration (Bauer et al., 2022), Argillic alteration (Nasab & Agah, 2023), Phyllic alteration (Soloviev et al., 2019) and Propylitic alteration (H.-h. Yang et al., 2022) was mapped using the band rationing, principal component analysis and Intensity- Hue-Saturation (IHS) transformation from remotely sensed satellite data.

The terrain can also influence the spatial distribution of mineralization, which can be considered as an indirect sign of the subsurface geology (Carranza, 2009; Keykhay-Hosseinpoor et al., 2020). The study area is characterized by a rugged mountainous landscape, where the spatial attributes and land cover are significantly influenced by topographic variables such as surface elevation and slope. The guidelines developed by Grunsky (1996) for mineral exploration also included the topographic elevation as an important parameter for effective mineral exploration and included in the list of must require datasets. Thus, topographic factors such as surface elevation and slope are taken into consideration when studying the spatial distribution of Cu mineralization in the study area as shown in Figures 9.7 and 9.8 respectively. The topographic predictor maps were derived and processed from the Space Shuttle Radar Topography Mission (SRTM) based 1-arc second Digital Elevation Model (DEM) with a spatial resolution of about 30 meters as it has been used in several research studies.

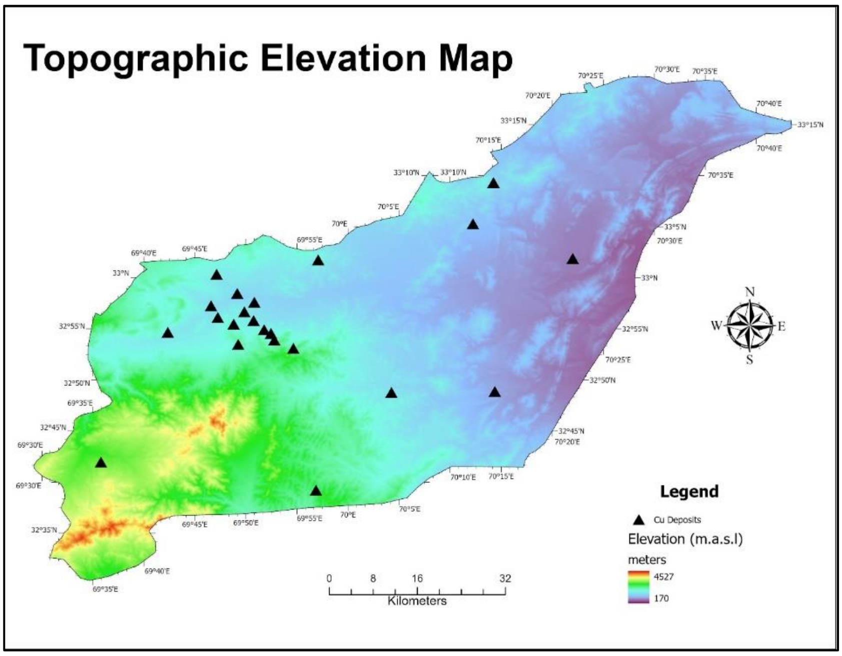


Figure 9.7 The topographic elevation of the study area emphasises the predominantly difficult mountainous environment

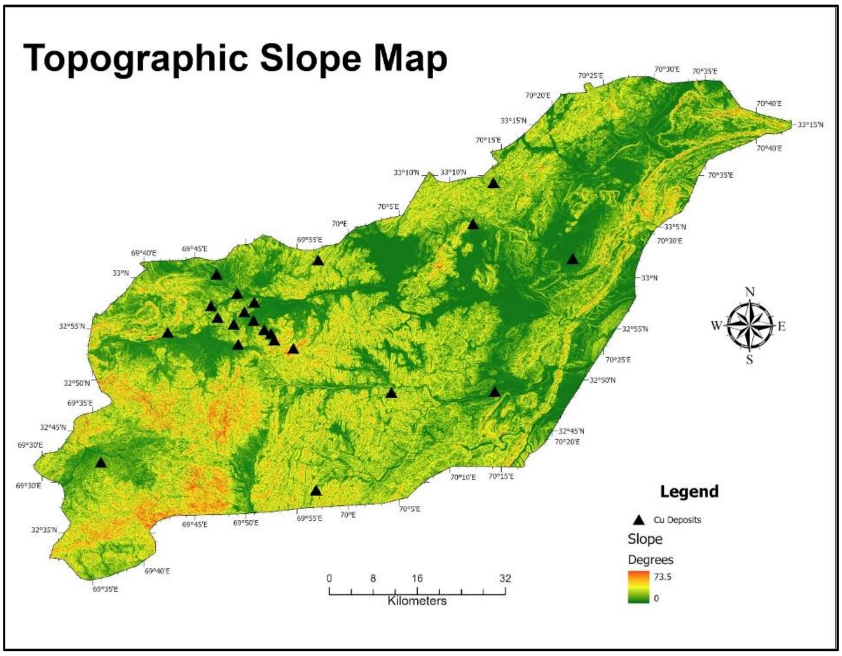


Figure 9.8 The topographic slope of the study area, highlighting the western high slopes

9.2.6. Target variable

During training and testing, the occurrence of mineral deposits is represented as a binary variable, with a value of 1 if the deposit exists and 0 if it does not. The 22 known Cu deposits were used as positive samples for the presence of mineralization. The following criteria were used to select sample locations for non-deposit occurrences; first to ensure adequate training data and an equitable balance of positive and negative samples, the quantity of non-deposit and deposit samples must be equal. This will allow for an accurate assessment of the ML algorithms' prediction capability. Second, non-deposit locations should be far away from any known deposits, since those near to known deposits may exhibit similar characteristics to those of the Cu mineralization and thus have a high chance of new deposit discoveries. Point pattern analysis was applied to assess the adequacy of the distance between deposit locations. The distances between each deposit as well as its closest adjacent deposit were computed, and the findings were statistically assessed and shown in Figure 9.9. It is evident that all the deposits are within 8247m of each other, indicating a 100% likelihood of another deposit being present within that range for any given deposit. As deposits are unlikely to be found beyond 8247 meters, a buffer distance of 5200 meters, with an 80% chance of finding deposits near any identified deposit, was established.

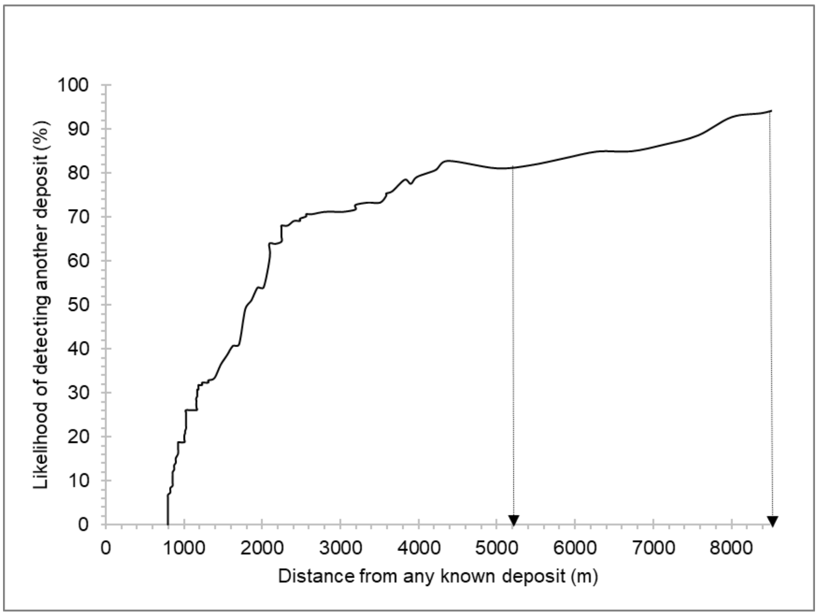


Figure 9.9 Point pattern analysis indicating the likelihood of discovering another Cu deposit relative to known Cu deposits

The non-deposit samples must be chosen outside of known deposit areas. Favourable regions within a 3500 m radius of intrusive rocks were identified through the distance distribution analysis conducted in the area. Several buffer intrusions appear to coincide with delineated areas of known deposit sites, indicating areas of extensive Cu mineralization for further exploration. The other buffered zones of intrusions outside of known deposits may provide favourable indicators in underexplored locations. Hence, the locations of non-deposit should not only be selected outside of any buffer zone, but also distant from known deposit areas. Non-deposit locations should be randomly dispersed, as opposed to mineral deposits, which are the outcome of rare events and non-random mineralization processes that tend to cluster spatially. As target samples, 22 non-deposit locations were chosen at random using the criteria in Figure 9.10.

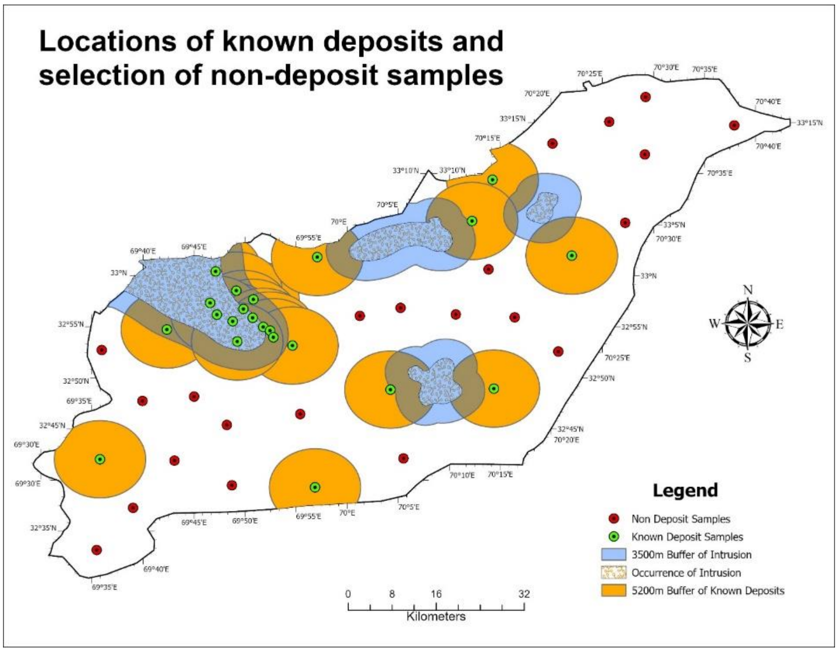


Figure 9.10 Point pattern analysis indicating the likelihood of discovering another Cu deposit relative to known Cu deposits

9.2.7. Input dataset

Maps of evidentiary features need to be converted into raster data with numerical values representing the evidence in each cell before applying prospectivity modelling. A methodology proposed by Carranza and Laborte (2015) was used to objectively select the cell size. According to the point pattern analysis, the shortest distance between two deposits is 900 m, indicating that using a cell size larger than 900 m will likely involve multiple deposits. Thus, a suitable cell size should be within the range of 900 m, as the upper limit. The minimum cell size can be determined based on the map scale of the spatial information. The best readable spatial resolution (S_R) can be calculated through equation 9.7.

$$S_R = MS \times 0.00025 \qquad \text{Equation 9.7}$$

The MS (scale of the map) determines the minimum cell size, which is 250 m. Finally, the raster maps were generated with 40,500 pixels, by using a cell size of 400 m within the appropriate size range. Notably, the original geographic data was collected using separate surveys at a lower scale, yielding in raster images that may not provide a single sample per cell. The spatial interpolation methods of Inverse Distance Weighted were utilized in the rasterizing processes, which influenced the accuracy of feature delineation. In most Geographical Information Systems (GIS) based mineral prospectivity modelling applications, the predictor maps were combined into a single grid data instead of being used directly as an input for creating prospectivity models (Rodriguez-Galiano et al., 2015; Zuo et al., 2021). The nine predictor maps were integrated into a single integral map, along with an associated feature dataset, using the digital overlay approach in GIS. This database contains 40,500 records, each containing a nine-dimensional input vector for training. After the training processes were done, the results were assigned to a new field in the same table.

From the 40,500 cells, 44 cells having occurrence sites (deposit and non-deposit) were extracted to create a labelled dataset. While 66% of the extracted data was utilized for training the CNN, RF, and SVM models, 34% was retained as a test dataset to evaluate the accuracy of these models.

9.2.8. Model training

Choosing the right criteria to train a model is a critical aspect of predictive modelling, necessary to ensure reliable predictions. Nonetheless, it can be challenging to determine the best configuration for achieving the desired accuracy, as there is no defined method for determining the appropriate parameters for every given case. This study uses a 10-fold cross-validation technique to assess the predictions made by different combinations of parameters. While empirical terms are helpful, finding the ideal arrangement requires a very subjective trial-and-error approach (Sun et al., 2020). The input training dataset is divided into ten equal-sized parts, one of which is selected as the validation dataset and the other nine are selected as the training dataset. This step was repeated ten times to ensure that all subsets were only used once as the validation data set (Raschka, 2018). The mean squared error was used to assess cross-validation results, which can be expressed as equation 9.8.

$$MSE = \frac{1}{D_N}\sum_{i=1}^{D_N}(\hat{P}_i - P_i)^2$$ Equation 9.8

For a validation dataset containing D_N data points, $\hat{P}_i$ represents the predicted class value (deposit as 1, non-deposit as 0) and P_i is the actual class value of each target data. The optimal configuration was established by identifying the model with the least Mean Squared Error (MSE). Table 9.3 lists the parameters of each ML model, their characteristics, as well as the recommended range of parameter values indicated by different authors (Adem, 2022; Ghorbanzadeh et al., 2019; Imamverdiyev & Sukhostat, 2019; Sothe et al., 2020) for CNN; (Imran et al., 2022; Sothe et al., 2020; Tuba et al., 2017; Xu et al., 2019b) for SVM and (Agrawal et al., 2022; Carranza & Laborte, 2015; Yin & Li, 2022) for RF.

Table 9.3 Training parameters for machine learning models

Model	**Parameters**	**Description**	**Reference Range**
RF	Number of trees	Number of trees in a random forest	10–500
	Number of features	Number of features utilized to construct each tree	1–8
	Maximum depth	Maximum number of iterations to split	2-20
	Minimum leaf size	Minimum sample size in each leaf node	1-20

SVM	Gamma	RBF width parameter determining the affecting range of each support vector	0.1–1
	Cost	Penalty for incorrect classification	0.1–50
CNN	Number of feature maps	Convolutional filters utilized to create new feature maps	8–64
	Number of neurons	The total neurons present in the connected layer	8–64

Information Gain (IG) was utilized to evaluate the impact of input features on the trained model (N. Yang et al., 2022). The formula in equation 9.9 can compute the value of IG for a feature E_f associated with output class C (deposit or non-deposit).

$$IG\left(C, E_f\right) = H(C) - H\left(C \middle| E_f\right) \qquad \text{Equation 9.9}$$

Where $H(C)$ denote the entropy value of C, and $H\left(C \middle| E_f\right)$ be the entropy value of C when evidentiary features E_f have been assigned values.

9.2.9. Model evaluation

To evaluate the ML model's efficacy thoroughly, the research employed a confusion matrix, predictive accuracy metrics, a ROC curve, and a success-rate curve. The confusion matrix can effectively summarize the predictions of the model. It shows four outcomes of classification, i.e., TP (true positive; deposit sample identified correctly as a deposit), FN (false negative; deposit sample misclassified as a non-deposit), FP (false positive; non-deposit sample incorrectly classified as a deposit) and TN (true negative; non-deposit sample accurately identified as such) (Sun et al., 2019). The confusion matrix was used to measure the predictive performance of all the models using a range of statistical indices. These indices were formulated as given in equation 9.10 (Beguería, 2006; Maria Navin & Pankaja, 2016).

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Positive\ predictive\ value = \frac{TP}{TP + FP} \qquad \text{Equation 9.10}$$

$$Negative\ predictive\ value = \frac{TN}{TN + FN}$$

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

The effectiveness of ML models can be assessed through ROC and success-rate curves. A ROC curve depicts how a binary classification system performs when the discrimination threshold is changed. It illustrates how well the system can distinguish between two classes. The curve is produced by plotting the sensitivity (TPR) on the y-axis and the false positive rate (FPR, 1-specificity) on the x-axis with various thresholds. The threshold establishes the criterion for distinguishing the predictive outcomes. Cells with values higher or less than the limit were labelled as "deposit" or "non-deposit" respectively. A set of data points (TPR, FPR) were collected by varying the threshold in order to generate the ROC curve. In general, the closer the ROC curve is to the upper left corner, the better the model's performance will be. The Area Under the Curve (AUC) was used to quantify the performance of various predictive models. This metric ranges from 0 to 1, with a score of 1 indicating a perfect accuracy, i.e., a sensitivity of 1 and a specificity of 0 whereas a value of 0.5 would signify a completely random model. The success rate curve is derived by comparing the percentage of deposits in the target area with the total number of deposits at different thresholds.9

9.3. Results and Discussion

9.3.1. Satellite Derived Hydrothermal Alterations

The hydrothermal alterations associated with the potential identification of Cu mineralization contain critical diagnostic information at the RGB, NIR and SWIR regions of the electromagnetic spectrum. The spectral bands of Landsat-8 and Sentinel-2 data demonstrate a strong capability in detecting hydrothermal alterations. The results of Iron-oxide alteration, Argillic alteration, Phyllic alteration and Propylitic alteration based on Sentiel-2 are shown in Figure 9.11. Iron-oxide alteration is a key geological process in developing and exploitation of copper deposits. Iron-oxide alteration is distinguished by the presence of iron oxides such as hematite and magnetite and are formed by the reaction of hot fluids with iron-bearing minerals in rocks. Other minerals, such as feldspar, may also be replaced by iron oxides as a result of the process. Because the fluids that induce the alteration frequently contain large amounts of copper, there is a strong relationship between iron-oxide alteration and copper mineralization (Khaleghi et al., 2020). As these fluids interact with the rocks, copper minerals such as chalcopyrite, bornite, and chalcocite are deposited. These copper minerals are frequently discovered in the altered rocks that surround the iron-oxide alteration zone, which is highlighted in tone of purple colours of Map of Iron-Oxide alteration in Figure 9.11. Argillic alteration is a significant geological process that contributes to the development of copper deposits and serves as a crucial indicator for geologists in their search of discovering new copper deposits (Beygi et al., 2021). Argillic alteration is a geological process that develops when hydrothermal fluids come into contact with rocks that comprise minerals such as feldspar, mica, and others. The minerals present in the rocks undergo alteration due to the previously mentioned interaction, leading to the formation of a type of clay mineral known as kaolinite. This phenomenon has the potential to spread across a vast area and possibly extend to considerable depths below the surface. Figure 9.11 depicts the argillic alteration, which is visually distinguished through chromatic variations ranging from cyan to pink shades. Phyllic alteration is another hydrothermal alteration process that is frequently related with the formation of porphyry copper deposits. This process transforms primary minerals into a suite of minerals that commonly includes quartz, sericite, and pyrite. The existence of phyllic alteration zones is often an indication of the possible occurrence of porphyry copper mineralization. The phyllic alteration is highlighted in Figure 9.11 as light orange to yellowish in colour. Another alteration of rocks produced by hydrothermal fluids is referred to as propylitic alteration. These fluids contain a high concentration of water, carbon dioxide, and other minerals that react with the host rock to generate new minerals. Propylitic anomalies are zones of propylitic alteration seen in rocks around copper deposits

(Parcutela et al., 2022). These anomalies are distinguished by the presence of minerals indicative of hydrothermal alteration, such as chlorite, epidote, and actinolite. The existence of these minerals indicates that the rocks were changed by hydrothermal fluids migrating from the copper deposit (Parcutela et al., 2022). The propylitic alteration is highlighted in Figure 9.11 because it is greenish-grey in colour and has a fine-grained texture on the ground.

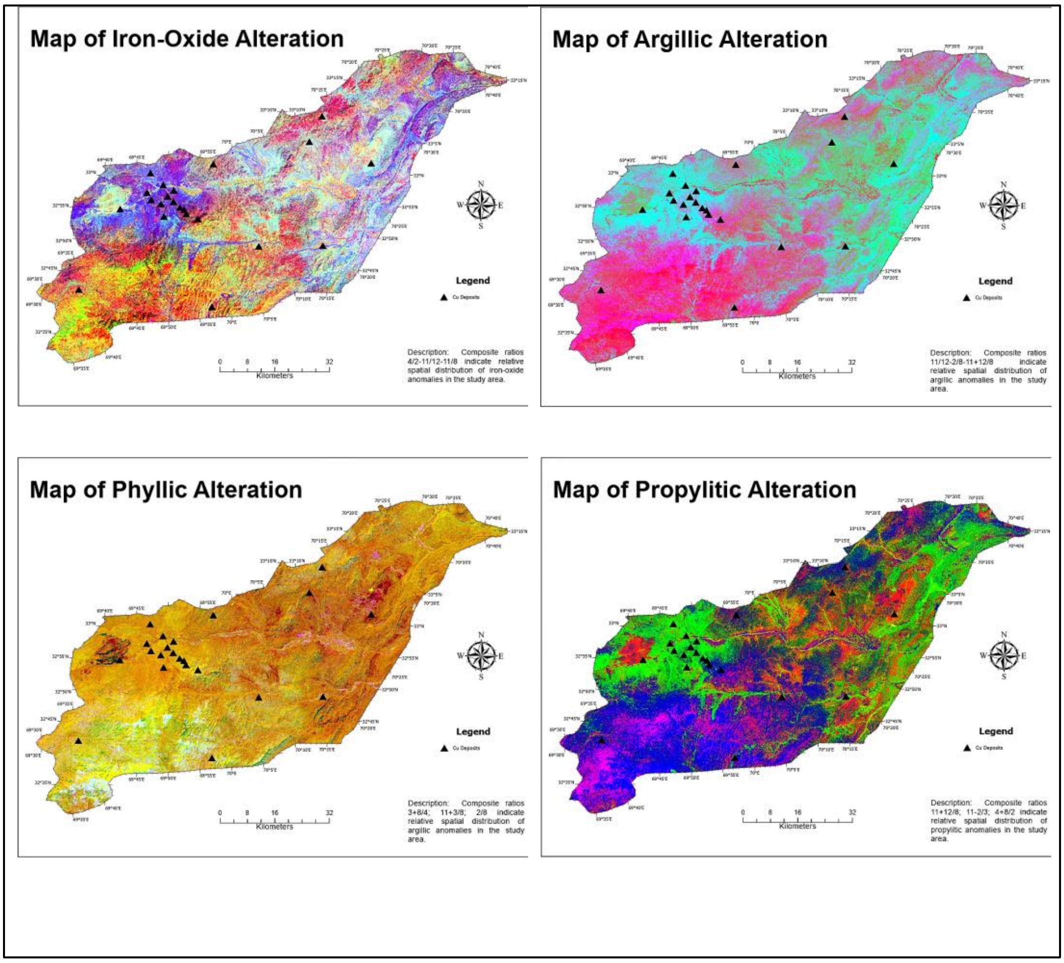


Figure 9.11 Hydrothermal maps of the study area used to map hydrothermal alternation zones associated with Cu mineralization based on Sentinel-2 satellite data

9.3.2. Sensitivity analysis

Parameters configured for model training can have a major effect on the reliability and generalizability of ML techniques, which significantly influences the predictions accuracy. The 10-fold cross-validation results revealed significant variations in the misclassification rates of the three models when different parameter settings are used which can be seen through overall accuracy results as shown in Table 9.4.

Table 9.4 Analysis of 10-fold cross-validation showing Overall Accuracy results.

Model	**Overall Accuracy**			
	Minimum	Maximum	Mean	Standard deviation
CNN	0.79	0.89	0.86	± 0.03
SVM	0.70	0.79	0.73	± 0.02
RF	0.79	0.85	0.82	± 0.02

The different parameters yield different levels of classification precision when training individual ML models. However, RF and CNN models show similar variation patterns of MSE in response to changes in the specific parameter, i.e., no significant change was observed in the MSE of both models. It does not indicate that the MSE values were comparable. Generally, these two models demonstrate more accurate and stable performance with overall accuracy values greater than 0.8 and are less sensitive to parameter variations. On the other hand, SVM models produce inadequate results with an average overall accuracy of 0.73. The complexity of a model can be determined by the number of ML parameters associated with its architecture, such as the number of trees in a Random Forest or the number of feature maps and neurons in a Convolutional Neural Network. Many research studies (Agrawal et al., 2022; Caruana et al., 2015; Jung & Choi, 2021; Shirmard et al., 2022; Sun et al., 2019) have established that more precise predictions are often produced by complex ML models. This study investigated the influence of architecture-related parameters on the MSE results of prospectivity modelling. The result revealed that increasing the complexities of the model did not increase the efficacy of prospectivity modelling in the area. As the parameters increase, the statistical indices used to quantify classification error, such as the lowest and average MSE, have not decreased significantly. Moreover, the use of a greater number of trees, neurons, and feature maps increases computation power. Consequently, it is suggested that the modelling processes discussed here do not necessarily require complex ML model architectures. These findings are also consistent with those from earlier research (Rodriguez-Galiano et al., 2015; Sun et al., 2019) which showed that complex models did not necessarily lead to improved accuracy of predictions. The limited size of training datasets, typically consisting of only a few deposit and non-deposit places, may explain the achieved accuracy in these results. Simple architectures and an appropriate amount of training can help avoid over-fitting errors that could be caused by using more complex models or extensive training.

9.3.3. Selection of Prospectivity Map

The optimized models will provide a probability score for each cell, assigned as a floating-point value ranging from 0 to 1, indicating the likelihood of a mineral presence. ML algorithms classify cells with probability values greater than 0.5 as areas with potential mineralization, while those below are marked as barren and without prospecting likelihood.

The classification accuracies of the three ML models are represented by the confusion matrices in the training and test datasets. Table 9.5 displays the confusion matrices for the three ML models, which are used to quantitatively measure the accuracy of binary classification through various statistical indices listed in Table 9.6.

Table 9.5 Confusion matrix of machine learning algorithms in test procedures

	RF		**CNN**		**SVM**	
	Actual Deposit	Actual Non-Deposit	Actual Deposit	Actual Non-Deposit	Actual Deposit	Actual Non-Deposit
Predicted deposit	6	2	7	1	3	5
Predicted non-deposit	1	6	1	6	4	3

Table 9.6 Predictive efficiency of machine learning models.

	RF	**CNN**	**SVM**
Sensitivity	85%	87%	75%
Specificity	75%	85%	71%
Positive predictive value	75%	85%	38%
Negative predictive value	85%	85%	60%
Accuracy	80%	80%	60%

The CNN model has overall accuracy greater than the RF and SVM models while predicting positive and negative samples. Notably, SVM produces poorer predictions in both training and testing, particularly in the testing process that misclassifies most of the deposits. The CNN model was found to be the most sensitive by correctly identifying 87% of deposit locations, followed by the RF model (85%). The CNN model shows a specificity of 85%, which indicates that 85% of cells without mineralization being correctly identified as non-deposits. Similarly, the RF and SVM models also achieved reasonable specificities of 75% and 71%, respectively. The CNN and RF models successfully predicted Cu occurrences in the majority of the predicted cells (> 80%), demonstrating a high positive predictive value. The CNN and RF models had the highest negative predictive rate (85%), indicating that 85% of predicted non-deposit cells are true non-deposit samples followed by SVM (60%). In terms of overall accuracy, the CNN model achieved the best value of 80%, indicating that 80% of all samples were correctly classified, and the same pattern was also followed by RF (80%).

The ROC curve was used to assess the predictive capability of high-probability zones. The discriminative thresholds were changed from high to low probability values for evaluation purposes. The ROC curve with a y-value of 1 and an area under the curve (AUC) of 1 means the probability of occurrence samples is greater than that of non-occurrence samples. Any deviation from this line, whether due to a non-occurrence cell with a high probability or an occurrence cell with a low probability, reduces the AUC value.

The RF model has the ROC curve closest to the upper left corner of the graph, indicating the highest predictive capability with AUC values greater than 0.95. The AUC values of the CNN model are similarly good, while the SVM model shows the least accurate performance in terms of ROC curves as shown in Figure 9.12.

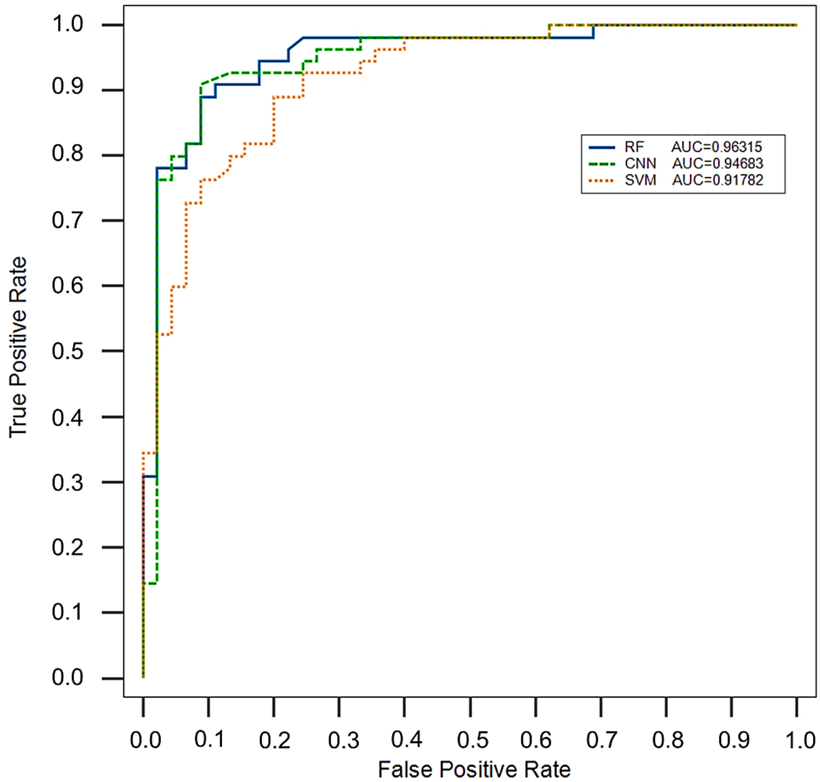


Figure 9.12 The ROC curve and AUC analysis of RF, CNN and SVM models

The majority of ML models can accurately predict >80% of known events under the initial classification parameters. Nonetheless, exploring all these "prospective areas" is too costly for "real world" mineral exploration, thus making target regions with high confidence and minimal area the ideal option. Hence it is vital for such applications to evaluate the demarcation of these zones as well as the effectiveness of the models. The success rate curve analysis highlights the performance of models in making predictions, with a steeper slope indicating a higher level of efficiency. Identifying the cut-off limits indicated by the three regression lines, which signify high, moderate and low probability for finding new Cu deposits, can be used to carry out prospectivity zoning. This enables delineated areas to capture more known mineral occurrences. Areas with a success rate curve slope more than five are considered high potential and are suitable for exploration. Areas with a slope equal to or less than the natural distributed density (for example, 1% of the area containing 1% of mineral occurrences) are not considered prospective. The success-rate curves show that the RF model has the highest slope, making it the most efficient in predicting outcomes. In the RF model, 8.99% of the study area accounted for 63.46% of the known deposits, compared to 5.98% for the CNN model containing 57.67% of the deposits and 3.14% for the SVM model containing 53.85%. The RF model identified 18 known deposits (81.81%) within 14.79% of the study area, including very-high and high potential zones. CNN and SVM generated similar results, with 16 and 14 deposits respectively, but covered a larger area (26.18%).

Based on the threshold values for the predictive models, the area was classified into potential zones for discovering new deposits as shown in Figure 9.13.

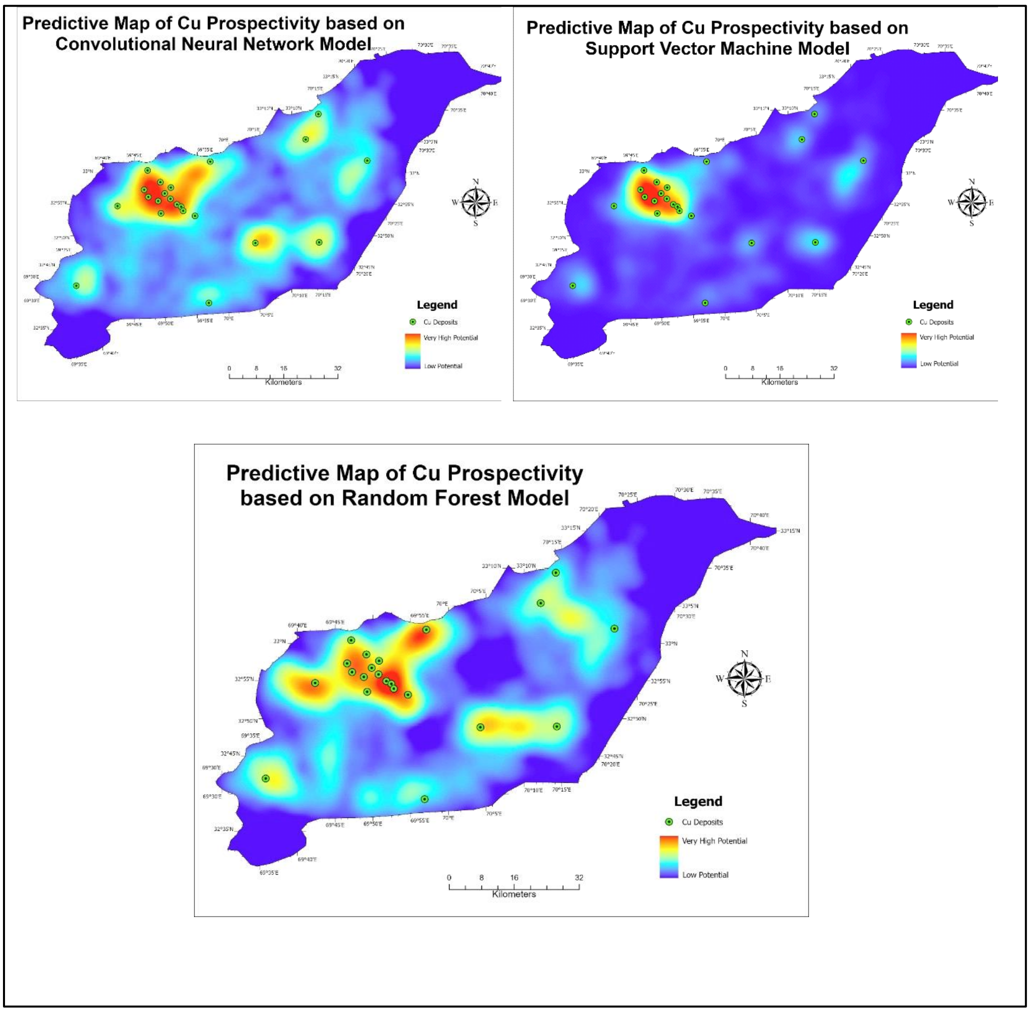


Figure 9.13 Predictive maps of CNN, SVM, and RF models highlighting low to very high potential areas delineated based on threshold values

The CNN model was found to be the best to correctly identify the Cu deposits in the study area, however it is not the best predictor. One of the possible explanations of this can be associated with the availability of limited field dataset. The RF model, on the other hand, is both a good identifier and the best predictor of Cu deposits, making it the optimal choice for exploration targeting since predictive accuracy is of great importance for further mineral exploration decisions. The areas with very high potential of new Cu deposits were also in the close vicinity of the geological lineaments and mostly present in the suitable hosting geological rocks. The predicted zones are also aligned with the hydrothermal alterations as extracted by processing of satellite remote sensing data. The ongoing mining in the region also falls under the very high potential zones towards the North-West side as predicted under this research study (Figure 9.14). Hence the predictive zones highlighted by the RF model were used as the final prospectivity map. The RF model showed the most impressive results in comparison to other ML based Mineral Prospectivity Mapping or Modelling (MPM) applications. Several research studies (Jung & Choi, 2021; Rodriguez-Galiano et al., 2015; Shirmard et al., 2022; Woodhead & Landry, 2021) has also reported that the RF model has outperformed other ML models in the application of machine learning for minerals exploration and mapping. This can be due to the random sampling conditions employed by the RF training, which provides a high-level of diversity with limited data and prevents over-fitting.

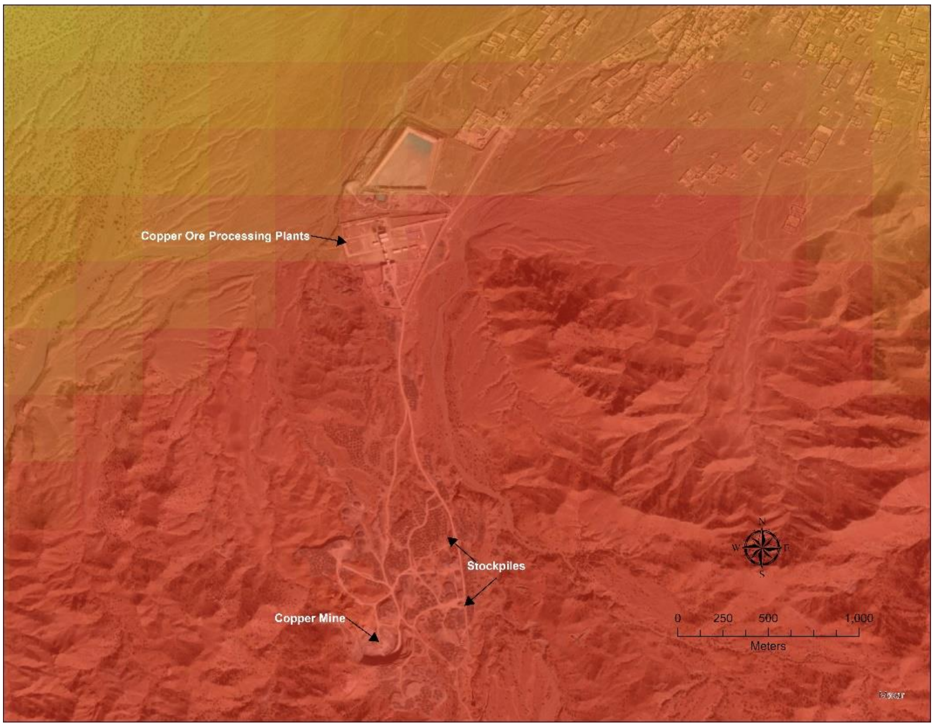


Figure 9.14 Current copper mining in the study area lies within the very high potential mineralization zone, as predicted by the RF model

Although the hydrothermal alteration maps were produced from satellite data, field-based investigations for the verification of host rocks are extremely important and should be explored for future research in the region. To further enhance the identified potential Cu mineralization prospects, more detailed geological information, particularly data on the geochemistry of rocks and soil, needs to be collected and analysed.

9.4. Conclusions

In this study, a data science approach was applied based on ML algorithms including CNN, RF, and SVM. A set of ML models, including RF, SVM and CNN were trained based on nine predictor maps using different parameters. During the training and testing process, both the CNN and RF models demonstrated high accuracy and consistency despite any changes to their parameters. According to the sensitivity analysis, model design with complex structure does not necessarily improve the prediction accuracy of MPM in the study area. In terms of classification accuracy, the CNN model achieves the highest level of accuracy, with 92.38% of labelled samples correctly identified, while RF and SVM trail closely behind. The success rate curves showed that the RF model can detect a greater number of Cu deposits in smaller, well-defined regions, resulting in the most promising predictive accuracy. Given the model evaluation measurements obtained, the RF model stands out as the best ML model for developing prospectivity maps for further research in the study area. The study reveals that satellite-derived hydrothermal alteration maps, proximity to intrusive contacts, and the number and density of geological fault intersections are the most influential criteria for Cu prospectivity predictions. The findings suggest that Cu hydrothermal alterations in the host rocks along with ML based models represent an underutilized exploration technique, providing important information on the origin of Cu mineralization.

The recent discovery of a known Cu deposit through field mineral exploration in the study area validates the predictive efficiency of the ML models applied in this study. The data science approach proposed and evaluated in this study for mineral prospectivity prediction has huge significance, not only for delineating of known Cu deposits, but also for exploring other deposits related to the remote sensing satellite derived hydrothermal mineral system on a larger scale.

10. CONCLUSIONS AND RECOMMENDATIONS

10.1. Introduction

This research was conducted to enhance the effectiveness and accuracy of mineral resource explorations and estimations, particularly for low-income developing economies. Exploration for mineral resources is an expensive and time-consuming process, with a lower success rate and return on investment than other sectors. In light of declining discovery success rates and increasing demand for essential minerals, industry professionals are being encouraged to use new data and approaches to locate economically viable mineral resources at the lowest cost and within the shortest period. This thesis consists of a series of case studies highlighting the effectiveness of data science by integrating limited field data with satellite remote sensing data and utilizing advanced machine learning algorithms for mineral exploration and estimate. The geochemical samples from stream sediments were used as an input into several machine learning models for predictive modeling in order to develop the further exploration areas in the region. Several machine learning models including Support Vector Regression (SVR), Generalized Linear Model (GLM), Deep Neural Network (DNN), Decision Tree (DT), and Random Forest (RF) regression, were applied for prediction using the Sentinel-2 satellite multispectral dataset. Overall, the performance of SVR and RF outperforms the other machine learning models with the highest testing accuracy. In the other case, the predictive modelling of geochemical samples involved the utilization of inverse distance weighting and geostatistical kriging techniques. Multiple iterations of the Inverse Distance Weighting (IDW) model were implemented, incorporating different power functions and varying numbers of sample points. The most optimal model was determined by evaluating the root mean square prediction error statistics. The same statistical technique were employed in the selection of the most appropriate semivariogram models, namely circular, spherical, exponential, and Gaussian, for each geochemical parameter in the Kriging process. The development of the mineral prospectivity map for the area was ultimately achieved through the application of the geochemical accumulation index (GAI).. In order to generate more accurate predictions, it was suggested that IDW and kriging should be used together. In a separate case study, Landsat 8 satellite data was optimized and fused with a digital elevation model. An automated

GIS-based lineament extraction algorithm was applied to the data, and the lineament density was calculated. The results demonstrated that the orientation and distribution of the extracted lineaments correlated with geology and mineralized zones of the region. This method of automatic lineament extraction was found to be highly effective and can be employed in other regions where geological and mineralogical information is limited. The multispectral satellite data with a moderate resolution was used to map altered rocks. To identify hydrothermal alteration minerals, digital image processing techniques such as band ratios, spectral band combinations, and principal component analysis (PCA) were utilized. The combination of spectral bands and PCA produced efficient results, suggesting that remotely sensed satellite data is a useful tool for mineral alteration mapping. In the last case study, a data science approach was applied to create nine predictor maps for mineral prospectivity mapping using machine learning algorithms including Convolutional Neural Network (CNN), Random Forest (RF) and Support Vector Machine (SVM) in combination with remote sensing and limited field datasets. To assess the performance of the predictive models, a confusion matrix, metrics and ROC curve were utilized. The RF model demonstrated the highest accuracy, consistency and interpretability, and was the most efficient in capturing known mineral deposits in the region. In addition to other models, the RF model was used to generate a prospectivity map depicting regions with high to low potential for further investigation. This has been verified by the discovery of a new deposit within the predicted prospective zones, demonstrating the strength and efficacy of the prospectivity modeling approach given in this study for identifying exploration targets.

10.2. Recommendations

The following are some recommendations for consideration in future research:

Extended research can be done to compare the use of Landsat-8 and Sentinel-2 satellite datasets with higher spatial and spectral resolution satellite datasets and drone data using the methodology applied in this research.

When developing machine learning models for mineral exploration and estimation with satellite remote sensing data, it is crucial to account for the multi-source, multi-scale, high-dimensional, dynamic state, isomer, and non-linear properties of remote sensing data.

Other deep learning models, such as Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Multilayer Perceptron (MLPs), and Radial Basis Function Networks (RBFNs), should also be evaluated and compared for potential mineral exploration and estimation.

Other researchers have reported inconsistent findings regarding the accuracy of ML-based methods, the comparison between ML and non-ML methods, and comparisons between various different ML-based methods, despite the vast number of research projects that have employed ML-based methods. Therefore, it is strongly suggested that the scientific community adopt a uniform framework for comparing the performance of various ML methods and non-ML methods.

The proposed data science framework through several case studies in this research can be deployed to the inaccessible and/or new areas with limited ground knowledge for efficient, cost-effective and time-saving minerals exploration and their quantification.

The same method may be used to the analysis of drillhole samples, expanding the depth dimension of ore body characterization and quantification to provide an improved understanding of mineral domains.

Exploring geologists experienced challenges as a result of the constraints and unique requirements of various Machine Learning methods. Utilizing advanced analytics in mineral exploration is essential for the mining industry to accomplish its sustainable development goals. These economical approaches may assist geoscientists in mitigating the negative consequences of mineral prospecting activities on the ecosystem, environment, and climate.

REFERENCE LIST

Abbaszadeh, M., Hezarkhani, A., & Soltani-Mohammadi, S. (2013). An SVM-based machine learning method for the separation of alteration zones in Sungun porphyry copper deposit. *Geochemistry*, *73*(4), 545-554.

Abburu, S., & Golla, S. B. J. I. j. o. c. a. (2015). Satellite image classification methods and techniques: A review. *119*(8).

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.

Abdullah, A., Akhir, J. M., & Abdullah, I. J. E. J. o. G. E. (2010). Automatic mapping of lineaments using shaded relief images derived from digital elevation model (DEMs) in the Maran–Sungi Lembing area, Malaysia. *15*, 1-9.

Abedi, M., & Norouzi, G.-H. (2016). A general framework of TOPSIS method for integration of airborne geophysics, satellite imagery, geochemical and geological data. *International journal of applied earth observation geoinformation*, *46*, 31-44.

Abedi, M., Norouzi, G.-H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, *46*, 272-283.

Abhary, A., Hassani, H. J. I. J. o. A. E., Management, & Science. (2016). Mapping Hydrothermal Mineral Deposits Using PCA and BR Methods in Baft 1: 100000 Geological Sheet, Iran. *2*(9).

Abubakar, A. J. a., Hashim, M., & Pour, A. B. (2018). Using ASTER satellite data for mapping hydrothermal alteration as a tool in geothermal exploration with GPS field validation. *Advanced Science Letters*, *24*(6), 4489-4495.

Adem, K. (2022). Impact of activation functions and number of layers on detection of exudates using circular Hough transform and convolutional neural networks. *Expert Systems with Applications*, *203*, 117583.

Adiri, Z., El Harti, A., Jellouli, A., Lhissou, R., Maacha, L., Azmi, M., Zouhair, M., & Bachaoui, E. M. J. A. i. S. R. (2017). Comparison of Landsat-8, ASTER and Sentinel 1 satellite remote sensing data in automatic lineaments extraction: A case study of Sidi Flah-Bouskour inlier, Moroccan Anti Atlas. *60*(11), 2355-2367.

Adiri, Z., Lhissou, R., El Harti, A., Jellouli, A., & Chakouri, M. (2020a). Recent advances in the use of public domain satellite imagery for mineral exploration: A review of Landsat-8 and Sentinel-2 applications. *Ore Geology Reviews*, *117*, 103332.

Adiri, Z., Lhissou, R., El Harti, A., Jellouli, A., & Chakouri, M. (2020b). Recent advances in the use of public domain satellite imagery for mineral exploration: A review of Landsat-8 and Sentinel-2 applications. *Ore geology reviews*, 103332.

Afzal, P., Asl, R. A., Adib, A., & Yasrebi, A. B. (2015). Application of fractal modelling for Cu mineralisation reconnaissance by ASTER multispectral and stream sediment data in Khoshname area, NW Iran. *Journal of the Indian Society of Remote Sensing*, *43*(1), 121-132.

Agrawal, N., Govil, H., Chatterjee, S., Mishra, G., & Mukherjee, S. (2022). Evaluation of machine learning techniques with AVIRIS-NG dataset in the identification and mapping of minerals. *Advances in Space Research*.

Ahmad, I. J. S. (2016). Preliminary Investigations for the Upgradation of North Waziristan Manganese Ore. *35*(1), 22-25.

Ahmed, N., Firoze, A., & Rahman, R. M. (2020). Machine learning for predicting landslide risk of Rohingya refugee camp infrastructure. *Journal of Information Telecommunication*, *4*(2), 175-198.

Ahmed, S. J. I. J. o. C. A. (2014). Lineament extraction from southern Chitradurga Schist Belt using Landsat TM, ASTERGDEM and geomatics techniques. *93*(12).

Al-Anazi, A., & Gates, I. (2010). Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. *Computers & Geosciences*, *36*(12), 1494-1503.

Alvarez-Berríos, N. L., & Aide, T. M. J. E. R. L. (2015). Global demand for gold is another threat for tropical forests. *10*(1), 014006.

Arndt, N. T., Fontboté, L., Hedenquist, J. W., Kesler, S. E., Thompson, J. F., & Wood, D. G. (2017). Section 3. Mineral Exploration: Discovering and Defining Ore Bodies. *Geochemical Perspectives*, *6*(1), 52-85.

Atif, I., Cawood, F., & Mahboob, M. (2020). Modelling and analysis of the Brumadinho tailings disaster using advanced geospatial analytics. *Journal of the Southern African Institute of Mining and Metallurgy*, *120*(7), 405-414.

Atwizukye, T. (2022). *Using Satellite Based Remote Sensing in Copper Ore Exploration, A case study of Kasese District* Makerere University].

Ayoobi, I., & Tangestani, M. H. (2017). The effect of minimum noise fraction data input on success of artificial neural network in lithological mapping of a magmatic terrain with ASTER data; A case study from SE Iran. *Remote Sensing Applications: Society and Environment*, *7*, 21-26.

Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, *23*(1), 49-60.

Ball, T., & TK, B. (1976). PRELIMINARY MINERAL RECONNAISSANCE OF CENTRAL WALES.

Bancalà, G., Cingiloglu, O., & Guerra, I. (2017). Integrated Multi-physics Approach for Lineament Extraction in Complex Geological Setting-A Case Study from Gediz Graben, Turkey. 79th EAGE Conference and Exhibition 2017,

Bauer, T. E., Lynch, E. P., Sarlus, Z., Drejing-Carroll, D., Martinsson, O., Metzger, N., & Wanhainen, C. (2022). Structural Controls on Iron Oxide Copper-Gold Mineralization and Related Alteration in a Paleoproterozoic Supracrustal Belt: Insights from the Nautanen Deformation Zone and Surroundings, Northern Sweden. *Economic geology*, *117*(2), 327-359.

Bedini, E. (2009). Mapping lithology of the Sarfartoq carbonatite complex, southern West Greenland, using HyMap imaging spectrometer data. *Remote Sensing of Environment*, *113*(6), 1208-1219.

Beedles, W. L., & Simkowitz, M. A. (1978). A note on skewness and data errors. *the Journal of Finance*, *33*(1), 288-292.

Beguería, S. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*, *37*, 315-329.

Bemis, S. P., Micklethwaite, S., Turner, D., James, M. R., Akciz, S., Thiele, S. T., & Bangash, H. A. (2014). Ground-based and UAV-based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *Journal of Structural Geology*, *69*, 163-178.

Benaissi, L., Tarek, A., Tobi, A., Ibouh, H., Zaid, K., Elamari, K., & Hibti, M. (2022). Geological mapping and mining prospecting in the Aouli inlier (Eastern Meseta, Morocco) based on remote sensing and geographic information systems (GIS). *China Geology*, *5*(4), 614-625.

Berk, A., Bernstein, L., Anderson, G., Acharya, P., Robertson, D., Chetwynd, J., & Adler-Golden, S. (1998). MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Remote Sensing of Environment*, *65*(3), 367-375.

Bersi, M., Saibi, H., & Chabou, M. C. (2016). Aerogravity and remote sensing observations of an iron deposit in Gara Djebilet, southwestern Algeria. *Journal of African Earth Sciences*, *116*, 134-150.

Beygi, S., Talovina, I. V., Tadayon, M., & Pour, A. B. (2021). Alteration and structural features mapping in Kacho-Mesqal zone, Central Iran using ASTER remote sensing data for porphyry copper exploration. *International Journal of Image and Data Fusion*, *12*(2), 155-175.

Bhattacharya, S., Majumdar, T., Rajawat, A., Panigrahi, M., & Das, P. (2012). Utilization of Hyperion data over Dongargarh, India, for mapping altered/weathered and clay minerals along with field spectral measurements. *International Journal of Remote Sensing*, *33*(17), 5438-5450.

Biland, J., Çöltekin, A. J. C., & Science, G. I. (2017). An empirical assessment of the impact of the light direction on the relief inversion effect in shaded relief maps: NNW is better than NW. *44*(4), 358-372.

Blandine, K. T. A., Jules, T. K., Martial, F. E., Ludovic, A. M., Julios, E. A., Robinson, S. B., Ousmanou, S., & Maurice, K. (2023). Geological mapping and structural interpretation of the Dschang-Santchou-escarpment (West, Cameroon), using Landsat 8 OLI/TIRS sensors/SRTM and field observations. *Geological Journal*, *58*(3), 1111-1130.

Booysen, R., Zimmermann, R., Lorenz, S., Gloaguen, R., Nex, P. A., Andreani, L., & Möckel, R. (2019). Towards Multiscale and Multisource Remote Sensing Mineral Exploration Using RPAS: A Case Study in the Lofdal Carbonatite-Hosted REE Deposit, Namibia. *Remote Sensing*, *11*(21), 2500.

Bouabdellah, M., & Sangster, D. F. (2016). Geology, Geochemistry, and Current Genetic Models for Major Mississippi Valley-Type Pb–Zn Deposits of Morocco. In *Mineral Deposits of North Africa* (pp. 463-495). Springer.

Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, *158*, 1429-1441.

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

Brown, M. (1993). *Exploration of gold in central wales*. B. G. Survey. https://www.bgs.ac.uk/downloads/start.cfm?id=1968

Cai, Y., Li, J., Li, X., Li, D., & Zhang, L. (2018). Estimating soil resistance at unsampled locations based on limited CPT data. *Bulletin of Engineering Geology and the Environment*, 1-12.

Calvin, W. M., Littlefield, E. F., & Kratt, C. (2015). Remote sensing of geothermal-related minerals for resource exploration in Nevada. *Geothermics*, *53*, 517-526.

Cao, M., Yin, D., Zhong, Y., Lv, Y., & Lu, L. (2023). Detection of geochemical anomalies related to mineralization using the Random Forest model optimized by the Competitive Mechanism and Beetle Antennae Search. *Journal of Geochemical Exploration*, 107195.

Cardoso-Fernandes, J., Lima, A., & Teodoro, A. (2018). Potential of Sentinel-2 data in the detection of lithium (Li)-bearing pegmatites: a study case. Earth resources and environmental remote sensing/GIS applications IX,

Cardoso-Fernandes, J., Teodoro, A. C., & Lima, A. (2019). Remote sensing data in lithium (Li) exploration: A new approach for the detection of Li-bearing pegmatites. *International Journal of Applied Earth Observation and Geoinformation*, *76*, 10-25.

Carranza, E. J. M. (2008). *Geochemical anomaly and mineral prospectivity mapping in GIS*. Elsevier.
Carranza, E. J. M. (2009). Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features. *Ore geology reviews*, *35*(3-4), 383-400.
Carranza, E. J. M. (2017). Natural resources research publications on geochemical anomaly and mineral potential mapping, and introduction to the special issue of papers in these fields. *Natural Resources Research*, *26*(4), 379-410.
Carranza, E. J. M., & Laborte, A. G. (2015). Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. *Ore geology reviews*, *71*, 777-787.
Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining,
Céleste, T. S., Hermann, F. D., Gautier, K. P., Eric, D. S., Willy, L., & Elvis, B. W. W. (2021). Comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) Satellite Remote Sensing Data in Auto-Matic Lineaments Extraction: A Case Study of Nkolezom, Southern Part of Cameroon. *The International Journal of Engineering and Science*, *10*, 24-31.
Chander, G., & Markham, B. (2003). Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(11), 2674-2677.
Chatterjee, S., & Bandopadhyay, S. (2011a). Goodnews Bay Platinum resource estimation using least squares support vector regression with selection of input space dimension and hyperparameters. *Natural Resources Research*, *20*, 117-129.
Chatterjee, S., & Bandopadhyay, S. (2011b). Goodnews Bay Platinum Resource Estimation Using Least Squares Support Vector Regression with Selection of Input Space Dimension and Hyperparameters. *Natural resources research*, *20*(2), 117-129.
Chatterjee, S., Bandopadhyay, S., & Machuca, D. (2010a). Ore grade prediction using a genetic algorithm and clustering based ensemble neural network model. *Mathematical Geosciences*, *42*(3), 309-326.
Chatterjee, S., Bandopadhyay, S., & Machuca, D. (2010b). Ore grade prediction using a genetic algorithm and clustering based ensemble neural network model. *Mathematical Geosciences*, *42*, 309-326.
Chatterjee, S., Bandopadhyay, S., & Rai, P. (2008). Genetic algorithm-based neural network learning parameter selection for ore grade evaluation of limestone deposit. *Mining Technology*, *117*(4), 178-190.
Chatterjee, S., Bhattacherjee, A., Samanta, B., & Pal, S. K. (2010). Image-based quality monitoring system of limestone ore grades. *Computers in Industry*, *61*(5), 391-408.
Chattoraj, S., Sharma, R., Kumar, C., & Sengar, V. (2020). Identification and characterization of hydrothermally altered minerals using surface and space-based reflectance spectroscopy, in parts of south-eastern Rajasthan, India. *SN Applied Sciences*, *2*(4), 1-9.
Chavez, P. S. (1996). Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing*, *62*(9), 1025-1035.

CHEN, L., GUO, D., HU, Z., SHENG, Y., & ZHANG, H. J. P. I. G. (2004). A Study on Remote Sensing Monitoring Land Use Change and Reclamation Measures of Subsided Land in Xuzhou Coal Mining Area [J]. *2*.

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

Chen, W., Pourghasemi, H. R., Kornejady, A., & Zhang, N. (2017). Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma*, *305*, 314-327.

Chen, Y., & Wu, W. (2017). Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochemistry: Exploration, Environment, Analysis*, *17*(3), 231-238.

Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore geology reviews*, *32*(1-2), 314-324.

Cheng, Q. (2012). Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. *Journal of Geochemical Exploration*, *122*, 55-70.

Cheng, Q., Bonham-Carter, G., Wang, W., Zhang, S., Li, W., & Qinglin, X. (2011). A spatially weighted principal component analysis for multi-element geochemical data for mapping locations of felsic intrusions in the Gejiu mineral district of Yunnan, China. *Computers & Geosciences*, *37*(5), 662-669.

Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B., & Kim, K.-W. (2008). Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain. *Remote Sensing of Environment*, *112*(7), 3222-3233.

Ciampalini, A., Garfagnoli, F., Antonielli, B., Moretti, S., & Righini, G. (2013). Remote sensing techniques using Landsat ETM+ applied to the detection of iron ore deposits in Western Africa. *Arabian Journal of Geosciences*, *6*(11), 4529-4546.

Ciampalini, A., Garfagnoli, F., Del Ventisette, C., & Moretti, S. (2013). Potential use of remote sensing techniques for exploration of iron deposits in Western Sahara and Southwest of Algeria. *Natural Resources Research*, *22*(3), 179-190.

Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of remote sensing*, *3*(3-58), 2.2-4.

Clark, R. N., King, T. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research: Solid Earth*, *95*(B8), 12653-12680.

CNCrusher. (2018). *OBUASI WIKIPEDIA*. CNCrusher Retrieved 12 May 2019 from https://boerhaave-sliedrecht.nl/Jul-18/5855.html

Cohen, D., Silva-Santisteban, C., Rutherford, N., Garnett, D., & Waldron, H. (1999). Comparison of vegetation and stream sediment geochemical patterns in northeastern New South Wales. *Journal of Geochemical Exploration*, *66*(3), 469-489.

Coimbra, R., Rodriguez-Galiano, V., Olóriz, F., & Chica-Olmo, M. (2014). Regression trees for modeling geochemical data—An application to Late

Jurassic carbonates (Ammonitico Rosso). *Computers Geosciences*, *73*, 198-207.

Corgne, S., Magagi, R., Yergeau, M., & Sylla, D. J. R. S. o. E. (2010). An integrated approach to hydro-geological lineament mapping of a semi-arid region of West Africa using Radarsat-1 and GIS. *114*(9), 1863-1875.

Correia, M. I. T., & Waitzberg, D. L. (2003). The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. *Clinical nutrition*, *22*(3), 235-239.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

Cozzolino, D., & Moron, A. (2004). Exploring the use of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes. *Animal Feed Science Technology*, *111*(1-4), 161-173.

Cressie, N. A. (1993). Statistics for spatial data: Wiley series in probability and mathematical statistics. *Find this article online*.

Crippen, R. E. (1990). Selection of Landsat TM band and band-ratio combinations to maximize lithologic information in color composite displays.

Cyples, N. N., Ielpi, A., & Dirszowsky, R. W. (2020). Planform and stratigraphic signature of proximal braided streams: remote-sensing and ground-penetrating-radar analysis of the Kicking Horse River, Canadian Rocky Mountains. *Journal of Sedimentary Research*, *90*(1), 131-149.

da Cunha Frutuoso, R. M. (2015). Mapping hydrothermal gold mineralization using Landsat 8 data. A case of study in Chaves license, Portugal.

da Silva, G. F., Silva, A. M., Toledo, C. L. B., Junior, F. C., & Klein, E. L. (2022). Predicting mineralization and targeting exploration criteria based on machine-learning in the Serra de Jacobina quartz-pebble-metaconglomerate Au-(U) deposits, São Francisco Craton, Brazil. *Journal of South American Earth Sciences*, *116*, 103815.

Dabrowski, J. M., Ashton, P. J., Murray, K., Leaner, J. J., & Mason, R. P. (2008). Anthropogenic mercury emissions in South Africa: Coal combustion in power plants. *Atmospheric Environment*, *42*(27), 6620-6626.

Dagdelen, K., & Vega, A. N. (1997). Geostatistics applied to mine waste characterization at Leadville, Colorado, USA. *International Journal of Surface Mining, Reclamation and Environment*, *11*(4), 175-188.

Das Goswami, A., Mishra, M., & Patra, D. (2017). Investigation of general regression neural network architecture for grade estimation of an Indian iron ore deposit. *Arabian Journal of Geosciences*, *10*(4).

de Mulder, E. F., Cheng, Q., Agterberg, F., & Goncalves, M. (2016). New and game-changing developments in geochemical exploration. *Episodes*, *39*(1), 70-71.

Deutsch, C. V., & Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide. Hauptbd*. Oxford university press.

Di Tommaso, I., & Rubinstein, N. (2007). Hydrothermal alteration mapping using ASTER data in the Infiernillo porphyry deposit, Argentina. *Ore geology reviews*, *32*(1), 275-290.

Ding, Q., Huang, R., Wang, F., Chen, J., Wang, M., & Zhang, X. J. M. P. i. E. (2018). Multi-Parameter Dominant Grouping of Discontinuities in Rock Mass Using Improved ISODATA Algorithm. *2018*.

Ding, X., Hasanipanah, M., Rad, H. N., & Zhou, W. (2020). Predicting the blast-induced vibration velocity using a bagged support vector regression optimized with firefly algorithm. *Engineering with Computers*, 1-12.

Ding, Z., Mei, G., Cuomo, S., Xu, N., & Tian, H. (2017). Performance evaluation of gpu-accelerated spatial interpolation using radial basis functions for building explicit surfaces. *International Journal of Parallel Programming*, 1-29.

Dornan, T., O'Sullivan, G., O'Riain, N., Stueeken, E., & Goodhue, R. (2020). The application of machine learning methods to aggregate geochemistry predicts quarry source location: An example from Ireland. *Computers Geosciences*, 104495.

Drews-Armitage, S., Romberger, S., & Whitney, C. (1996). Clay alteration and gold deposition in the Genesis and Blue Star deposits, Eureka County, Nevada. *Economic geology*, *91*(8), 1383-1393.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. Advances in neural information processing systems,

Ducart, D. F., Silva, A. M., Toledo, C. L. B., & Assis, L. M. d. (2016). Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajás Mineral Province, Brazil. *Brazilian Journal of Geology*, *46*(3), 331-349.

Durand, J. (2012). The impact of gold mining on the Witwatersrand on the rivers and karst system of Gauteng and North West Province, South Africa. *Journal of African Earth Sciences*, *68*, 24-43.

Dutta, S., Bandopadhyay, S., Ganguli, R., & Misra, D. (2010). Machine learning algorithms and their application to ore reserve estimation of sparse and imprecise data. *Journal of Intelligent Learning Systems and Applications*, *2*(02), 86.

Ekneligoda, T. C., Henkel, H. J. C., & Geosciences. (2010). Interactive spatial analysis of lineaments. *36*(8), 1081-1090.

El-Makky, A. M. (2011). Statistical analyses of La, Ce, Nd, Y, Nb, Ti, P, and Zr in bedrocks and their significance in geochemical exploration at the Um Garayat gold mine Area, Eastern Desert, Egypt. *Natural Resources Research*, *20*(3), 157.

Es-Sabbar, B., Essalhi, M., Essalhi, A., & Mhamdi, H. S. (2020). Lithological and structural lineament mapping from Landsat 8 OLI images in Ras Kammouna arid area (Eastern Anti-Atlas, Morocco). *Economic and Environmental Geology*, *53*(4), 425-440.

Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental pollution*, *114*(3), 313-324.

Farahbakhsh, E., Chandra, R., Olierook, H. K., Scalzo, R., Clark, C., Reddy, S. M., & Müller, R. D. (2020). Computer vision-based framework for extracting tectonic lineaments from optical remote sensing data. *International Journal of Remote Sensing*, *41*(5), 1760-1787.

Farahbakhsh, E., Chandra, R., Olierook, H. K., Scalzo, R., Clark, C., Reddy, S. M., & Muller, R. D. J. a. p. a. (2018). Computer vision-based framework for extracting geological lineaments from optical remote sensing data.

Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science Control Engineering: An Open Access Journal*, *2*(1), 602-609.

Fink, G. (2018). Computer vision-based motion control and state estimation for unmanned aerial vehicles (UAVs).

Fletcher, W. (1997). Stream sediment geochemistry in today’s exploration world. Proceedings of exploration, Franco-Sepulveda, G., Del Rio-Cuervo, J. C., & Pachón-Hernández, M. A. (2019). State of the art about metaheuristics and artificial neural networks applied to open pit mining. *Resources Policy*, *60*, 125-133.

Freedman, J. (1972). *Geochemical prospecting for zinc, lead, copper, and silver, Lancaster Valley, southeastern Pennsylvania*. US Government Printing Office.

Frutuoso, R., Lima, A., & Teodoro, A. C. (2021). Application of remote sensing data in gold exploration: targeting hydrothermal alteration using Landsat 8 imagery in northern Portugal. *Arabian Journal of Geosciences*, *14*, 1-18.

Fu, Y., Cheng, Q., Jing, L., Ye, B., & Fu, H. (2023). Mineral Prospectivity Mapping of Porphyry Copper Deposits Based on Remote Sensing Imagery and Geochemical Data in the Duolong Ore District, Tibet. *Remote Sensing*, *15*(2), 439.

Gabr, S., Ghulam, A., & Kusky, T. (2010). Detecting areas of high-potential gold mineralization using ASTER data. *Ore geology reviews*, *38*(1), 59-69.

Gad, S., & Kusky, T. (2006). Lithological mapping in the Eastern Desert of Egypt, the Barramiya area, using Landsat thematic mapper (TM). *Journal of African Earth Sciences*, *44*(2), 196-202.

Gale. (2017). *Education, Learning and Research Resources Online - Gale*. Retrieved 15 Nov. 2017 from

Gao, R., Xue, C., Zhao, X., Chen, X., Li, Z., & Symons, D. (2019). Source and possible leaching process of ore metals in the Uragen sandstone-hosted Zn-Pb deposit, Xinjiang, China: Constraints from lead isotopes and rare earth elements geochemistry. *Ore geology reviews*, *106*, 56-78.

Ge, W., Cheng, Q., Jing, L., Armenakis, C., & Ding, H. (2018). Lithological discrimination using ASTER and Sentinel-2A in the Shibanjing ophiolite complex of Beishan orogenic in Inner Mongolia, China. *Advances in Space Research*, *62*(7), 1702-1716.

Geranian, H., Tabatabaei, S. H., Asadi, H. H., & Carranza, E. J. M. (2016). Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay gold deposit, NW Iran. *Natural Resources Research*, *25*, 145-159.

Gholami, R., Moradzadeh, A., & Yousefi, M. (2012). Assessing the performance of independent component analysis in remote sensing data processing. *Journal of the Indian Society of Remote Sensing*, *40*(4), 577-588.

Ghorbani, A., Honarmand, M., Shahriari, H., & Hassani, M. J. (2019). Regional scale prospecting for non-sulphide zinc deposits using ASTER data and different spectral processing methods. *International Journal of Remote Sensing*, *40*(23), 8647-8667.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, *11*(2), 196.

Girouard, G., Bannari, A., El Harti, A., & Desrochers, A. (2004). Validated spectral angle mapper algorithm for geological mapping: comparative study between QuickBird and Landsat-TM. XXth ISPRS congress, geo-imagery bridging continents, Istanbul, Turkey,

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294-300.

Global, S. (2022). *S&P Global*.

Gocht, W. R., Zantop, H., & Eggert, R. G. (2012). *International mineral economics: mineral exploration, mine valuation, mineral markets, international mineral policies*. Springer Science & Business Media.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, *228*(1-2), 113-129.

Govil, H., Gill, N., Rajendran, S., Santosh, M., & Kumar, S. J. O. G. R. (2018). Identification of new base metal mineralization in Kumaon Himalaya, India, using hyperspectral remote sensing and hydrothermal alteration. *92*, 271-283.

Grunsky, E. (1996). Spatial Data Integration for Mineral Exploration, Resource Assessment and Environmental Studies-A Guidebook. By International Atomic Energy Agency. *NONRENEWABLE RESOURCES*, *5*(1), 78-79.

Grunsky, E. C., Mueller, U. A., & Corrigan, D. (2014). A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: applications for predictive geological mapping. *Journal of Geochemical Exploration*, *141*, 15-41.

Hajaj, S., El Harti, A., Jellouli, A., Pour, A. B., Mnissar Himyari, S., Hamzaoui, A., & Hashim, M. (2023). Evaluating the Performance of Machine Learning and Deep Learning Techniques to HyMap Imagery for Lithological Mapping in a Semi-Arid Region: Case Study from Western Anti-Atlas, Morocco. *Minerals*, *13*(6), 766.

Han, T., & Nelson, J. (2015). Mapping hydrothermally altered rocks with Landsat 8 imagery: A case study in the KSM and Snowfield zones, northwestern British Columbia. *British Columbia Geological Survey Paper*.

Hattis, D., Ogneva-Himmelberger, Y., & Ratick, S. (2012). The spatial variability of heat-related mortality in Massachusetts. *Applied Geography*, *33*, 45-52.

Hauff, P. (2008). An overview of VIS-NIR-SWIR field spectroscopy as applied to precious metals exploration. *Spectral International Inc*, *80001*, 303-403.

Hemalatha, S., & Anouncia, S. M. (2017). Unsupervised segmentation of remote sensing images using FD based texture analysis model and ISODATA. *International Journal of Ambient Computing and Intelligence (IJACI)*, *8*(3), 58-75.

Holmes, R., & Lu, L. (2015). Introduction: overview of the global iron ore industry. In *Iron Ore* (pp. 1-42). Elsevier.

Horel, J. (1984). Complex principal component analysis: Theory and examples. *Journal of climate and Applied Meteorology*, *23*(12), 1660-1673.

Hörfarter, C., Reischer, J., & Schiegl, M. (2019). From Maps to Apps—Entering Another Technological Dimension Within the Geological Survey of Austria. In *Service-Oriented Mapping* (pp. 345-363). Springer.

Horrocks, T., Wedge, D., Holden, E.-J., Kovesi, P., Clarke, N., & Vann, J. (2015). Classification of gold-bearing particles using visual cues and cost-sensitive machine learning. *Mathematical Geosciences*, *47*(5), 521-545.

Hosseini, S., Asghari, O., & Emery, X. (2017). Direct block-support simulation of grades in multi-element deposits: application to recoverable mineral resource estimation at Sungun porphyry copper-molybdenum deposit. *Journal of the Southern African Institute of Mining and Metallurgy*, *117*(6), 577-585.

Hou, D., O'Connor, D., Nathanail, P., Tian, L., & Ma, Y. (2017). Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. *Environmental pollution*, *231*, 1188-1200.

Hunt, G. R. (1977). Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, *42*(3), 501-513.
Hussain, F., & Jeong, J. (2015). Exploiting deep neural networks for digital image compression. 2015 2nd world symposium on web applications and networking (WSWAN),
Imamverdiyev, Y., & Sukhostat, L. (2019). Lithological facies classification using deep convolutional neural network. *Journal of Petroleum Science and Engineering*, *174*, 216-228.
Imran, M., Ahmad, S., Sattar, A., & Tariq, A. (2022). Mapping sequences and mineral deposits in poorly exposed lithologies of inaccessible regions in Azad Jammu and Kashmir using SVM with ASTER satellite data. *Arabian Journal of Geosciences*, *15*(6), 538.
Jafrasteh, B., & Fathianpour, N. (2017). A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. *Neurocomputing*, *235*, 217-227.
Jafrasteh, B., Fathianpour, N., & Suárez, A. (2018). Comparison of machine learning methods for copper ore grade estimation. *Computational Geosciences*, *22*(5), 1371-1388.
Jalloh, A. B., Kyuro, S., Jalloh, Y., & Barrie, A. K. (2016). Integrating artificial neural networks and geostatistics for optimum 3D geological block modeling in mineral reserve estimation: A case study. *International Journal of Mining Science and Technology*, *26*(4), 581-585.
Jensen, J. R. (2015). *Introductory digital image processing: a remote sensing perspective*. Prentice Hall Press.
Johnson, J. R., Bell III, J. F., Bender, S., Blaney, D., Cloutis, E., Ehlmann, B., Fraeman, A., Gasnault, O., Kinch, K., & Le Mouélic, S. (2016). Constraints on iron sulfate and iron oxide mineralogy from ChemCam visible/near-infrared reflectance spectroscopy of Mt. Sharp basal units, Gale Crater, Mars. *American Mineralogist*, *101*(7), 1501-1514.
Johnston, K., Hoef, J., Krivoruchko, K., & Lucas, N. (2012). ArcGIS® 9: using ArcGIS® geostatistical analyst. ESRI. In.
Jolliffe, I. T. (2002). Principal component analysis and factor analysis. *Principal component analysis*, 150-166.
Jones, R. (2002). Algorithms for using a DEM for mapping catchment areas of stream sediment samples. *Computers & Geosciences*, *28*(9), 1051-1060.
Jung, D., & Choi, Y. (2021). Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. *Minerals*, *11*(2), 148.
Kaiser, P. K., Henning, J. G., Cotesta, L., & Dasys, A. (2002). Innovations in mine planning and design utilizing collaborative immersive virtual reality (CIRV). Proceedings of the 104th CIM Annual General Meeting,
Kaplan, U. E., Dagasan, Y., & Topal, E. (2021). Mineral grade estimation using gradient boosting regression trees. *International Journal of Mining, Reclamation and Environment*, *35*(10), 728-742.
Karaman, M., Özelkan, E., & Tasdelen, S. (2018). Influence of basin hydrogeology in the detectability of narrow rivers by Sentinel2-A satellite images: A case study in Karamenderes (Çanakkale). *Journal of Natural Hazards Environment*, *4*, 140-155.
Karan, S. K., Samadder, S. R., & Maiti, S. K. J. J. o. e. m. (2016). Assessment of the capability of remote sensing and GIS techniques for monitoring reclamation success in coal mine degraded lands. *182*, 272-283.

Karim, M., Maanan, M., Maanan, M., Rhinane, H., Rueff, H., & Baidder, L. (2019). Assessment of water body change and sedimentation rate in Moulay Bousselham wetland, Morocco, using geospatial technologies. *International journal of sediment research*, *34*(1), 65-72.
Kaufmann, H. (1988). Concepts, processing and results. *International Journal of Remote Sensing*, *9*(10-11), 1639-1658.
Kayet, N., Pathak, K., Chakrabarty, A., & Sahoo, S. (2018). Mapping the distribution of iron ore minerals and spatial correlation with environmental variables in hilltop mining areas. *Environmental Earth Sciences*, *77*, 1-14.
Kent, J., & Leitner, M. (2007). Efficacy of standard deviational ellipses in the application of criminal geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, *4*(3), 147-165.
Keykhay-Hosseinpoor, M., Kohsary, A.-H., Hossein-Morshedy, A., & Porwal, A. (2020). A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran. *Ore geology reviews*, *116*, 103234.
Khalajmasoumi, M., Sadeghi, B., Carranza, E. J. M., & Sadeghi, M. (2017). Geochemical anomaly recognition of rare earth elements using multi-fractal modeling correlated with geological features, Central Iran. *Journal of Geochemical Exploration*, *181*, 318-332.
Khaleghi, M., Ranjbar, H., Abedini, A., & Calagari, A. A. (2020). Synergetic use of the Sentinel-2, ASTER, and Landsat-8 data for hydrothermal alteration and iron oxide minerals mapping in a mine scale. *Acta Geodyn. Geromater*, *17*, 311-329.
Khamis, N., Sin, T. C., & Hock, G. C. (2018). Segmentation of Residential Customer Load Profile in Peninsular Malaysia using Jenks Natural Breaks. 2018 IEEE 7th International Conference on Power and Energy (PECon),
Khan, E. (2000). Flotation of copper minerals from north waziristan copper ore, on pilot-scale. *Quarterly Sci. Vision*, *6*(1), 10-20.
Khan, I. U., Khan, A., & Ullah, A. (2022). Causes and factors responsible for Operation Zarb-e-Azb: perspective of internally displaced persons of North Waziristan, Pakistan. *Liberal Arts and Social Sciences International Journal (LASSIJ)*, *6*(1), 181-200.
Khan, S. R., Jan, M. Q., Khan, T., & Khan, M. A. (2007). Petrology of the dykes from the Waziristan Ophiolite, NW Pakistan. *Journal of Asian Earth Sciences*, *29*(2-3), 369-377.
Kirkwood, C., Everett, P., Ferreira, A., & Lister, B. (2016). Stream sediment geochemistry as a tool for enhancing geological understanding: an overview of new data from south west England. *Journal of Geochemical Exploration*, *163*, 28-40.
Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering version 2.3. *Engineering*, *45*(4ve), 1051.
Kleinschmidt, I., Bagayoko, M., Clarke, G. P. Y., Craig, M., & Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, *29*(2), 355-361. https://doi.org/Doi 10.1093/Ije/29.2.355
Koga, S. C., Sugita, S., Kamata, S., Ishiguro, M., Hiroi, T., Tatsumi, E., & Sasaki, S. (2018). Spectral decomposition of asteroid Itokawa based on principal component analysis. *Icarus*, *299*, 386-395.
Köhler, M., Hanelli, D., Schaefer, S., Barth, A., Knobloch, A., Hielscher, P., Cardoso-Fernandes, J., Lima, A., & Teodoro, A. C. (2021). Lithium potential

mapping using artificial neural networks: A case study from central portugal. *Minerals*, *11*(10), 1046.

Koike, K., Nagano, S., Ohmi, M. J. C., & Geosciences. (1995). Lineament analysis of satellite images using a segment tracing algorithm (STA). *21*(9), 1091-1104.

Komnitsas, K., Guo, X., & Li, D. (2010). Mapping of soil nutrients in an abandoned Chinese coal mine and waste disposal site. *Minerals Engineering*, *23*(8), 627-635.

Kong, Y., Chen, G., Liu, B., Xie, M., Yu, Z., Li, C., Wu, Y., Gao, Y., Zha, S., & Zhang, H. (2022). 3D Mineral Prospectivity Mapping of Zaozigou Gold Deposit, West Qinling, China: Machine Learning-Based Mineral Prediction. *Minerals*, *12*(11), 1361.

Kotnise, G., & Chennabasappa, S. (2015). Application of Remote Sensing Techniques in Identification of Lithological Rock Units in Southern Extension of Kolar Schist Belt from Chigargunta, Chittoor District, Andhra Pradesh to Maharajagadai, Krishnagiri District, Tamil Nadu. *International Journal of Innovative Science, Engineering Technology*, *2*(11).

Kravchenko, A., & Bullock, D. G. (1999). A comparative study of interpolation methods for mapping soil properties. *Agronomy Journal*, *91*(3), 393-400.

Kröger, M., Sauer-Greff, W., Urbansky, R., Lorang, M., & Siegrist, M. (2016). Performance evaluation on contour extraction using Hough transform and RANSAC for multi-sensor data fusion applications in industrial food inspection. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2016,

Krupski, J., Graniszewski, W., & Iwanowski, M. (2021). Data transformation schemes for cnn-based network traffic analysis: A survey. *Electronics*, *10*(16), 2042.

Kruse, F. A. (2015). Comparative analysis of airborne visible/infrared imaging spectrometer (AVIRIS), and hyperspectral thermal emission spectrometer (HyTES) longwave infrared (LWIR) hyperspectral data for geologic mapping. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXI,

Lefever, D. W. (1926). Measuring geographic concentration by means of the standard deviational ellipse. *American Journal of Sociology*, *32*(1), 88-94.

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, *53*, 173-189.

Li, M., Guo, S., Chen, J., Chang, Y., Sun, L., Zhao, L., Li, X., & Yao, H. (2023). Stability Analysis of Unmixing-Based Spatiotemporal Fusion Model: A Case of Land Surface Temperature Product Downscaling. *Remote Sensing*, *15*(4), 901.

Li, N. (2010). *Textural and rule-based lithological classification of remote sensing data, and geological mapping in Southwestern Prieska sub-basin, Transvaal Supergroup, South Africa* lmu].

Li, T., Zuo, R., Xiong, Y., & Peng, Y. (2021). Random-drop data augmentation of deep convolutional neural network for mineral prospectivity mapping. *Natural Resources Research*, *30*, 27-38.

Li, X.-L., Li, L.-h., Zhang, B.-l., & Guo, Q.-j. (2013). Hybrid self-adaptive learning based particle swarm optimization and support vector regression model for grade estimation. *Neurocomputing*, *118*, 179-190.

Li, X.-l., Xie, Y.-l., Guo, Q.-j., & Li, L.-h. (2010). Adaptive ore grade estimation method for the mineral deposit evaluation. *Mathematical and Computer Modelling*, *52*(11-12), 1947-1956.

Li, X., Ao, Y., Guo, S., & Zhu, L. (2019). Combining Regression Kriging With Machine Learning Mapping for Spatial Variable Estimation. *IEEE Geoscience and Remote Sensing Letters*, *17*(1), 27-31.

Lim, E.-P., Foo, S., Khoo, C., Chen, H., Fox, E., Shalini, U., & Thanos, C. (2002). *Digital Libraries: People, Knowledge, and Technology: 5th International Conference on Asian Digital Libraries, ICADL 2002, Singapore, December 11-14, 2002, Proceedings* (Vol. 2555). Springer Science & Business Media.

Lin, X., Hu, Y., Meng, G., & Zhang, M. (2020). Geochemical patterns of Cu, Au, Pb and Zn in stream sediments from Tongling of East China: Compositional and geostatistical insights. *Journal of Geochemical Exploration*, *210*, 106457.

Litvinenko, V. (2020). Digital economy as a factor in the technological development of the mineral sector. *Natural Resources Research*, *29*(3), 1521-1541.

Liu, J., Xue, Y., Ren, K., Song, J., Windmill, C., & Merritt, P. (2019). High-performance time-series quantitative retrieval from satellite images on a GPU cluster. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(8), 2810-2821.

Liu, L., Li, Y., Zhou, J., Han, L., & Xu, X. (2018). Gold-copper deposits in Wushitala, Southern Tianshan, Northwest China: Application of ASTER data for mineral exploration. *Geological Journal*, *53*, 362-371.

Liu, X., Wu, J., & Xu, J. (2006). Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. *Environmental pollution*, *141*(2), 257-264.

Liu, Y., Cheng, Q., Carranza, E. J. M., & Zhou, K. (2018). Assessment of Geochemical Anomaly Uncertainty Through Geostatistical Simulation and Singularity Analysis. *Natural Resources Research*, 1-14.

Lobo, F. d. L., Souza-Filho, P. W. M., Novo, E. M. L. d. M., Carlos, F. M., & Barbosa, C. C. F. (2018). Mapping mining areas in the brazilian amazon using msi/sentinel-2 imagery (2017). *Remote Sensing*, *10*(8), 1178.

Lotfi, M., Ghanbari, H., Arefi, H., & Bahroudi, A. (2017). Mapping Alteration Zones using Gaussian Mixture Model and Spectral Angle Mapper. *Journal of Geomatics Science and Technology*, *6*(4), 217-229.

Loughlin, W. (1991). Principal component analysis for alteration mapping. *Photogrammetric engineering and remote sensing*, *57*(9), 1163-1169.

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*, *34*(9), 1044-1055.

Lu, Y., Yang, C., & He, R. (2022). Towards lithology mapping in semi-arid areas using time-series Landsat-8 data. *Ore geology reviews*, 105163.

Maduaka, A. C. (2014). *Contributions of Solid Mineral Sectors to Nigeria's Economic Development* Eastern Mediterranean University (EMU)-Doğu Akdeniz Üniversitesi (DAÜ)].

Mahandani, S., Puta, I., Hirawan, A., Abbas, R., & Titisari, A. (2018). Integrated Remote Sensing and Geological Mapping to Identify Landslide Prone Zone in Loano, Purworejo, Central Java. EAGE-HAGI 1st Asia Pacific Meeting on Near Surface Geoscience and Engineering,

Mahboob, Iqbal, & Atif. (2015). Modeling and Simulation of Glacier Avalanche: A Case Study of Gayari Sector Glaciers Hazards Assessment. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(11), 5824-5834.

Mahboob, M., Celik, T., & Genc, B. (2022). Review of machine learning-based Mineral Resource estimation. *Journal of the Southern African Institute of Mining and Metallurgy*, *122*(11), 655-664.
Mahboob, M., Genc, B., Celik, T., Ali, S., & Atif, I. (2019). Mapping hydrothermal minerals using remotely sensed reflectance spectroscopy data from Landsat. *Journal of the Southern African Institute of Mining and Metallurgy*, *119*(3), 279-289.
Mahboob, M. A., Celik, T., & Genc, B. (2020). Predictive modeling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments. *Arabian Journal of Geosciences*, *13*(20), 1-21.
Mahboob, M. A., & Genc, B. (2019). Evaluation of ISODATA Clustering Algorithm for Surface Gold Mining Using Satellite Data. 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE),
Mahboob, M. A., Iqbal, J., & Atif, I. (2015). Modeling and simulation of glacier avalanche: A case study of gayari sector glaciers hazards assessment. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(11), 5824-5834.
Mahboob;, Atif;, & Iqbal. (2015). Remote sensing and GIS applications for assessment of urban sprawl in Karachi, Pakistan. *Science, Technology and Development*, *34*(3), 179-188.
Malhotra, R. (2015). A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing*, *27*, 504-518.
Malkani, M. S., Mahmood, Z., Alyani, M. I., & Siraj, M. (2017). Mineral Resources of Khyber Pakhtunkhwa and FATA, Pakistan. *Geological Survey of Pakistan, Information Release*, *996*, 1-61.
Malkani, M. S. J. P. J. o. H. E. S., Special. (2013). Natural resources of Southern Khyber Pakhtunkhwa and FATA regions (Kohat sub-basin and part of northern Sulaiman Basin and Western Indus Suture), Pakistan-A review. Abstract Volume, Sustainable Utilization of Natural Resources of the Khyber Pakhtunkhwa and FATA, February 11, Peshawar. *2013*, 30-31.
Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, *43*(24), 8200-8214.
Manuel, R., Brito, M. d. G., Chichorro, M., & Rosa, C. (2017). Remote Sensing for Mineral Exploration in Central Portugal. *Minerals*, *7*(10), 184.
Marghany, M., & Hashim, M. J. I. J. o. P. S. (2010). Lineament mapping using multispectral remote sensing satellite data. *5*(10), 1501-1507.
Maria Navin, J., & Pankaja, R. (2016). Performance analysis of text classification algorithms using confusion matrix. *Int. J. Eng. Tech. Res. IJETR*, *6*, 75-78.
Marjoribanks, R. (2010). Geological Mapping in Exploration. In *Geological Methods in Mineral Exploration and Mining* (pp. 13-49). Springer.
Martinez, J.-M., Guyot, J.-L., Filizola, N., & Sondag, F. (2009). Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. *Catena*, *79*(3), 257-264.
Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., & Lim, T.-K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68-72.
Masoud, A., Koike, K. J. C., & Geosciences. (2017). Applicability of computer-aided comprehensive tool (LINDA: LINeament Detection and Analysis) and

shaded digital elevation model for characterizing and interpreting morphotectonic features from lineaments. *106*, 89-100.

Masoud, A., & Koike, K. J. J. o. A. E. S. (2006). Tectonic architecture through Landsat-7 ETM+/SRTM DEM-derived lineaments and relationship to the hydrogeologic setting in Siwa region, NW Egypt. *45*(4-5), 467-477.

Mather, P., & Tso, B. (2016). *Classification methods for remotely sensed data*. CRC press.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246-1266.

MatLab, P. (2020). 9.9. 0.1495850 (R2020b). *The MathWorks Inc.: Natick, MA, USA*.

Mehsud, S. (2012). Combating Militancy in Bajaur and North-Waziristan Agency in Federally Administered Tribal Areas (FATA) of Pakistan: A Comparative Analysis. Tigah. *A Journal o Peace and Development, FATA Research Centre, Islamabad*.

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software*, *18*(8-9), 761-799.

Meshkani, S. A., Mehrabi, B., Yaghubpur, A., & Sadeghi, M. J. O. G. R. (2013). Recognition of the regional lineaments of Iran: Using geospatial data and their implications for exploration of metallic ore deposits. *55*, 48-63.

Mia, B., & Fujimitsu, Y. (2012). Mapping hydrothermal altered mineral deposits using Landsat 7 ETM+ image in and around Kuju volcano, Kyushu, Japan. *Journal of earth system science*, *121*(4), 1049-1057.

Mia;, & Fujimitsu. (2012). Mapping hydrothermal altered mineral deposits using Landsat 7 ETM+ image in and around Kuju volcano, Kyushu, Japan. *Journal of earth system science*, *121*(4), 1049-1057.

Miao, F., Wu, Y., Xie, Y., & Li, Y. (2018). Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. *Landslides*, *15*(3), 475-488.

Mielke, C., Boesche, N. K., Rogass, C., Segl, K., & Kaufmann, H. (2014). Multi- and hyperspectral satellite sensors for mineral exploration, new applications to the Sentinel-2 and EnMAP mission. Proceedings of the 34th EARSeL Symposium, Poland, Warsaw,

Miller, J., & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *ecological modelling*, *157*(2-3), 227-247.

Mohamed Taha, A. 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*(1), 49.

Mondini, A. C. (2017). Measures of spatial autocorrelation changes in multitemporal SAR images for event landslides detection. *Remote Sensing*, *9*(6), 554.

Montgomery, M. (2016). *GSP 216 Online*. Humboldt State University Retrieved 18 March 2020 from http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson3-1/composites.html

Moral, F. J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *30*(4), 620-631.

Mueller, T., Pusuluri, N., Mathias, K., Cornelius, P., Barnhisel, R., & Shearer, S. (2004). Map quality for ordinary kriging and inverse distance weighted interpolation. *Soil Science Society of America Journal*, *68*(6), 2042-2047.

Munyai, V. H. (2017). *Assessing the impact of gold mining on the land use land cover change using GIS & Remote Sensing: case study in Yatela gold mine, Mali (1999-2015)*

Mwaniki, M. W., Moeller, M. S., Schellmann, G. J. I. A. o. t. P., Remote Sensing, & Sciences, S. I. (2015). A comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) in mapping geology and visualising lineaments: A case study of central region Kenya.

Namdeo, A., & Singh, D. (2021). Challenges in evolutionary algorithm to find optimal parameters of SVM: A review. *Materials Today: Proceedings*.

Nasab, M. H., & Agah, A. (2023). Mapping Hydrothermal Alteration Zones Associated with Copper Mineralization using ASTER Data: A Case Study from the Mirjaveh Area, Southeast Iran. *Transactions A: Basics*, *36*(04), 720.

Nforba, M. T., Api, L., Berinyuy, N., & Fils, S. C. N. (2019). Lineament Mapping Using RS and GIS Techniques at Mbateka, SE Cameroon: Implication for Mineralization. In *Advances in Remote Sensing and Geo Informatics Applications* (pp. 197-201). Springer.

Nieto, A., Bai, Y., & Brownson, J. (2014). Combined Life Cycle Assessment and Costing Analysis Optimization Model Using Multiple Criteria Decision Making in Earth-Resource Systems. *Natural Resources*, *5*(08), 351.

Nieto, A., & Toffait, Y. (2007). Tonnage-grade test of nonlinear estimators: indicator kriging, disjunctive kriging and uniform conditioning. *TRANSACTIONS-SOCIETY FOR MINING METALLURGY AND EXPLORATION INCORPORATED*, *320*, 45.

Nourali, H., & Osanloo, M. (2019). Mining capital cost estimation using Support Vector Regression (SVR). *Resources Policy*, *62*, 527-540.

Novak, I., & Soulakellis, N. (2000). Identifying geomorphic features using LANDSAT-5/TM data processing techniques on Lesvos, Greece. *Geomorphology*, *34*(1-2), 101-109.

O'leary, D., Friedman, J., & Pohn, H. J. G. S. o. A. B. (1976). Lineament, linear, lineation: some proposed new standards for old terms. *87*(10), 1463-1469.

Obodai, J., Adjei, K. A., Odai, S. N., & Lumor, M. (2019). Land use/land cover dynamics using landsat data in a gold mining basin-the Ankobra, Ghana. *Remote Sensing Applications: Society and Environment*, *13*, 247-256.

Okujeni, A., van der Linden, S., Tits, L., Somers, B., & Hostert, P. (2013). Support vector regression and synthetically mixed training data for quantifying urban land cover. *Remote Sensing of Environment*, *137*, 184-197.

Olea, R. A. (2006). A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment*, *20*(5), 307-318.

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. *Journal of Asian Earth Sciences*, *146*, 90-102.

Ottesen, R., Bogen, J., Bølviken, B., & Volden, T. (1989). Overbank sediment: a representative sample medium for regional geochemical mapping. *Journal of Geochemical Exploration*, *32*(1-3), 257-277.

Ousmanou, S., Fozing, E. M., Kwékam, M., Fodoue, Y., & Jeatsa, L. D. A. (2023). Application of remote sensing techniques in lithological and mineral exploration: discrimination of granitoids bearing iron and corundum deposits in southeastern Banyo, Adamawa region-Cameroon. *Earth Science Informatics*, 1-27.

Panahi, A., Cheng, Q., & Bonham-Carter, G. F. (2004). Modelling lake sediment geochemical distribution using principal component, indicator kriging and multifractal power-spectrum analysis: a case study from Gowganda, Ontario. *Geochemistry: Exploration, Environment, Analysis*, *4*(1), 59-70.

Parcutela, N., Dimalanta, C., Armada, L., Austria, R., Gabo-Ratio, J., & Yumul, G. (2022). Band processing of Landsat 8-OLI multi-spectral images as a tool for delineating alteration zones associated with porphyry prospects: A case from Suyoc, Benguet, Philippines. IOP Conference Series: Earth and Environmental Science,

Patel, A. K., Chatterjee, S., & Gorai, A. K. (2017a). Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. *Arabian Journal of Geosciences*, *10*, 1-16.

Patel, A. K., Chatterjee, S., & Gorai, A. K. (2017b). Development of online machine vision system using support vector regression (SVR) algorithm for grade prediction of iron ores. 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA),

Pendock, N., & McKay, A. (2019). Use of Thermal Infrared Remote Sensing for Targeting Mineral Deposits. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium,

Perez, C. A., Estévez, P. A., Vera, P. A., Castillo, L. E., Aravena, C. M., Schulz, D. A., & Medina, L. E. (2011). Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. *International Journal of Mineral Processing*, *101*(1-4), 28-36.

Piel, F. B., Patil, A. P., Howes, R. E., Nyangiri, O. A., Gething, P. W., Dewi, M., Temperley, W. H., Williams, T. N., Weatherall, D. J., & Hay, S. I. (2013). Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet*, *381*(9861), 142-151.

Piepho, H. P. (2019). A coefficient of determination (R2) for generalized linear mixed models. *Biometrical Journal*, *61*(4), 860-872.

Płaczkowska, E., Górnik, M., Mocior, E., Peek, B., Potoniec, P., Rzonca, B., & Siwek, J. (2015). Spatial distribution of channel heads in the Polish Flysch Carpathians. *Catena*, *127*, 240-249.

Pour, A. B., & Hashim, M. (2012). The application of ASTER remote sensing data to porphyry copper and epithermal gold deposits. *Ore geology reviews*, *44*, 1-9.

Pour, A. B., & Hashim, M. (2015). Regional hydrothermal alteration mapping using Landsat-8 data. Space Science and Communication (IconSpace), 2015 International Conference on,

Pour, A. B., Hashim, M., Park, Y., & Hong, J. K. (2017). Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. *Geocarto International*, 1-26.

Pour, A. B., Hashim, M., Park, Y., & Hong, J. K. (2018a). Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. *Geocarto International*, *33*(12), 1281-1306.

Pour, A. B., Hashim, M., Park, Y., & Hong, J. K. J. G. I. (2018b). Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. *33*(12), 1281-1306.

Pour, A. B., & Hashim, M. J. O. G. R. (2015). Structural mapping using PALSAR data in the Central Gold Belt, Peninsular Malaysia. *64*, 13-22.
Pournamdari, M., Hashim, M., & Pour, A. B. J. A. i. S. R. (2014). Spectral transformation of ASTER and Landsat TM bands for lithological mapping of Soghan ophiolite complex, south Iran. *54*(4), 694-709.
Pozdnoukhov, A., & Kanevski, M. (2007). Multi-scale support vector regression for hotspot detection and modeling.
Pushparaj, J., & Hegde, A. V. J. A. J. o. G. (2017). Comparison of various pan-sharpening methods using Quickbird-2 and Landsat-8 imagery. *10*(5), 119.
Qiu, J.-T., Zhang, C., & Hu, X. (2015). Integration of concentration-area fractal modeling and spectral angle mapper for ferric iron alteration mapping and uranium exploration in the Xiemisitan Area, NW China. *Remote Sensing*, *7*(10), 13878-13894.
Rajan Girija, R., & Mayappan, S. (2019). Mapping of mineral resources and lithological units: A review of remote sensing techniques. *International Journal of Image and Data Fusion*, *10*(2), 79-106.
Rajesh, H. (2004). Application of remote sensing and GIS in mineral resource mapping-An overview. *Journal of mineralogical and Petrological Sciences*, *99*(3), 83-103.
Ramakrishnan, D., & Bharti, R. (2015). Hyperspectral remote sensing and geological applications. *Curr Sci*, *108*(5), 879-891.
Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*.
Rathod, P. H., Müller, I., Van der Meer, F. D., & de Smeth, B. (2016). Analysis of visible and near infrared spectral reflectance for assessing metals in soil. *Environmental monitoring and assessment*, *188*(10), 558.
Redouane, M., Mhamdi, H. S., Haissen, F., Raji, M., & Sadki, O. (2022). Lineaments Extraction and Analysis Using Landsat 8 (OLI/TIRS) in the Northeast of Morocco. *Open Journal of Geology*, *12*(5), 333-357.
Reed, G. F., Lynn, F., & Meade, B. D. (2002). Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and diagnostic laboratory immunology*, *9*(6), 1235-1239.
Repacholi, M. (2012). *Clay mineralogy: spectroscopic and chemical determinative methods*. Springer Science & Business Media.
Richter, R. (1997). Correction of atmospheric and topographic effects for high spatial resolution satellite imagery. *International Journal of Remote Sensing*, *18*(5), 1099-1111.
Rigol-Sanchez, J., Chica-Olmo, M., & Abarca-Hernandez, F. (2003). Artificial neural networks as a tool for mineral potential mapping with GIS. *International Journal of Remote Sensing*, *24*(5), 1151-1156.
Robinson, R. *2018*. Retrieved 21 March from https://mlnotebook.github.io/post/CNN1/
Robinson, T., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and electronics in agriculture*, *50*(2), 97-108.
Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore geology reviews*, *71*, 804-818.
Rosenbaum, L., Dörr, A., Bauer, M. R., Boeckler, F. M., & Zell, A. (2013). Inferring multi-target QSAR models with taxonomy-based multi-task learning. *Journal of cheminformatics*, *5*(1), 1-20.

Rossel, R. V., Behrens, T., Ben-Dor, E., Brown, D., Demattê, J., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., & Adamchuk, V. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, *155*, 198-230.

Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.

Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, *185*, 57-70.

Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore geology reviews*, *14*(3), 157-183.

Sabins, F. F. (2007). *Remote sensing: principles and applications*. Waveland Press.

Sadowski, Ł., Nikoo, M., & Nikoo, M. (2015). Principal component analysis combined with a self organization feature map to determine the pull-off adhesion between concrete layers. *Construction and Building Materials*, *78*, 386-396.

Saepuloh, A., Haeruddin, H., Heriawan, M. N., Kubo, T., Koike, K., & Malik, D. J. G. (2018). Application of lineament density extracted from dual orbit of synthetic aperture radar (SAR) images to detecting fluids paths in the Wayang Windu geothermal field (West Java, Indonesia). *72*, 145-155.

Safari, M., Maghsoudi, A., & Pour, A. B. (2017). Application of Landsat-8 and ASTER satellite remote sensing data for porphyry copper exploration: a case study from Shahr-e-Babak, Kerman, south of Iran. *Geocarto International*, 1-16.

Salawu, N. B., Omosanya, K. O. L., Eluwole, A. B., Saleh, A., & Adebiyi, L. S. (2023). Structurally-controlled Gold Mineralization in the Southern Zuru Schist Belt NW Nigeria: Application of Remote Sensing and Geophysical Methods. *Journal of Applied Geophysics*, *211*, 104969.

Salisbury, J. W., D'Aria, D. M., & Jarosewich, E. (1991). Midinfrared (2.5–13.5 μm) reflectance spectra of powdered stony meteorites. *Icarus*, *92*(2), 280-297.

Salminen, R., & Gregorauskien, V. (2000). Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. *Applied Geochemistry*, *15*(5), 647-653.

Salui, C. L. J. J. o. t. G. S. o. I. (2018). Methodological Validation for Automated Lineament Extraction by LINE Method in PCI Geomatica and MATLAB based Hough Transformation. *92*(3), 321-328.

Samanta, B. (2010). Radial basis function network for ore grade estimation. *Natural Resources Research*, *19*(2), 91-102.

Samanta, B., & Bandopadhyay, S. (2009). Construction of a radial basis function network using an evolutionary algorithm for grade estimation in a placer gold deposit. *Computers & Geosciences*, *35*(8), 1592-1602.

Samanta, B., Bandopadhyay, S., & Ganguli, R. (2002). Data segmentation and genetic algorithms for sparse data division in Nome placer gold grade estimation using neural network and geostatistics. *Exploration and Mining Geology*, *11*(1-4), 69-76.

Samanta, B., Bandopadhyay, S., & Ganguli, R. (2006). Comparative evaluation of neural network learning algorithms for ore grade estimation. *Mathematical Geology*, *38*, 175-197.

Samanta, B., Bandopadhyay, S., & Ganguli, R. (2006). Comparative evaluation of neural network learning algorithms for ore grade estimation. *Mathematical Geology*, *38*(2), 175-197.
Samanta, B., Banopadhyay, S., Ganguli, R., & Dutta, S. (2005). A comparative study of the performance of single neural network vs. Adaboost algorithm based combination of multiple neural networks for mineral resource estimation. *Journal of the Southern African Institute of Mining and Metallurgy*, *105*(4), 237-246.
Samanta, B., Ganguli, R., & Bandopadhyay, S. (2005). Comparing the predictive performance of neural networks with ordinary kriging in a bauxite deposit. *Mining Technology*, *114*(3), 129-139.
Santos, D., Cardoso-Fernandes, J., Lima, A., Müller, A., Brönner, M., & Teodoro, A. C. (2022). Spectral analysis to improve inputs to random forest and other boosted ensemble tree-based algorithms for detecting NYF pegmatites in Tysfjord, Norway. *Remote Sensing*, *14*(15), 3532.
Santos, D., Teodoro, A., Lima, A., & Cardoso-Fernandes, J. (2019). Remote sensing techniques to detect areas with potential for lithium exploration in Minas Gerais, Brazil. Earth resources and environmental remote sensing/GIS applications X,
Savu-Krohn, C., Rantitsch, G., Auer, P., Melcher, F., & Graupner, T. (2011). Geochemical fingerprinting of coltan ores by machine learning on uneven datasets. *Natural Resources Research*, *20*(3), 177-191.
Schroeder, T. A., Cohen, W. B., Song, C., Canty, M. J., & Yang, Z. (2006). Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment*, *103*(1), 16-26.
Schueler, V., Kuemmerle, T., & Schröder, H. (2011). Impacts of surface gold mining on land use systems in Western Ghana. *Ambio*, *40*(5), 528-539.
Schwertmann, U. (1993). Relations between iron oxides, soil color, and soil formation. *Soil color*(soilcolor), 51-69.
Seidl, D. E., Paulus, G., Jankowski, P., & Regenfelder, M. (2015). Spatial obfuscation methods for privacy protection of household-level data. *Applied Geography*, *63*, 253-263.
Sekandari, M., Masoumi, I., Beiranvand Pour, A., M Muslim, A., Rahmani, O., Hashim, M., Zoheir, B., Pradhan, B., Misra, A., & Aminpour, S. M. (2020). Application of Landsat-8, Sentinel-2, ASTER and WorldView-3 spectral imagery for exploration of carbonate-hosted Pb-Zn deposits in the Central Iranian Terrane (CIT). *Remote Sensing*, *12*(8), 1239.
Sener, E., Davraz, A., & Ozcelik, M. J. H. J. (2005). An integration of GIS and remote sensing in groundwater investigations: a case study in Burdur, Turkey. *13*(5-6), 826-834.
Shahbeik, S., Afzal, P., Moarefvand, P., & Qumarsy, M. (2014). Comparison between ordinary kriging (OK) and inverse distance weighted (IDW) based on estimation error. Case study: Dardevey iron ore deposit, NE Iran. *Arabian Journal of Geosciences*, *7*(9), 3693-3704.
Sharifi, A., Malian, A., & Soltani, A. J. J. o. t. I. S. o. R. S. (2018). Efficiency Evaluating of Automatic Lineament Extraction by Means of Remote Sensing (Case Study: Venarch, Iran). *46*(9), 1507-1518.
Sheng, L., Zhang, T., Niu, G., Wang, K., Tang, H., Duan, Y., & Li, H. (2015). Classification of iron ores by laser-induced breakdown spectroscopy (LIBS) combined with random forest (RF). *Journal of Analytical Atomic Spectrometry*, *30*(2), 453-458.

Shimoda, H., & Kimura, T. (2017). 1.09 Japanese Space Program. *Comprehensive Remote Sensing*, 246.

Shirazi, A., Hezarkhani, A., Shirazy, A., & Shahrood, I. (2018). Remote Sensing Studies for Mapping of Iron Oxide Regions, South of Kerman, IRAN. *International Journal of Science and Engineering Applications*, *7*(4), 45-51.

Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, *268*, 112750.

Short, N. M., & Lowman Jr, P. D. (1973). Earth observations from space: outlook for the geological sciences.

Šilhavý, J., Minár, J., Mentlík, P., Sládek, J. J. C., & Geosciences. (2016). A new artefacts resistant method for automatic lineament extraction using Multi-Hillshade Hierarchic Clustering (MHHC). *92*, 9-20.

Smirnoff, A., Boisvert, E., & Paradis, S. J. (2008). Support vector machine for 3D modelling from sparse geological information of various origins. *Computers & Geosciences*, *34*(2), 127-143.

Soe, M., Kyaw, T. A., & Takashima, I. (2005). Application of remote sensing techniques on iron oxide detection from ASTER and Landsat images of Tanintharyi coastal area, Myanmar.

Solano-Acosta, W., Mastalerz, M., & Schimmelmann, A. J. I. J. o. C. G. (2007). Cleats and their relation to geologic lineaments and coalbed methane potential in Pennsylvanian coals in Indiana. *72*(3-4), 187-208.

Soloviev, S. G., Kryazhev, S. G., Dvurechenskaya, S. S., Vasyukov, V. E., Shumilin, D. A., & Voskresensky, K. I. (2019). The superlarge Malmyzh porphyry Cu-Au deposit, Sikhote-Alin, eastern Russia: Igneous geochemistry, hydrothermal alteration, mineralization, and fluid inclusion characteristics. *Ore geology reviews*, *113*, 103112.

Song, C., & Woodcock, C. E. (2003). Monitoring forest succession with multitemporal Landsat images: Factors of uncertainty. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(11), 2557-2567.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sensing of Environment*, *75*(2), 230-244.

Sothe, C., De Almeida, C., Schimalski, M., La Rosa, L., Castro, J., Feitosa, R., Dalponte, M., Lima, C., Liesenberg, V., & Miyoshi, G. (2020). Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience & Remote Sensing*, *57*(3), 369-394.

Soydan, H., Koz, A., & Düzgün, H. Ş. (2021). Secondary iron mineral detection via hyperspectral unmixing analysis with sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *101*, 102343.

Spychała-Kij, M. (2020). Civil-Military Cooperation in Post Conflict Development: A Case of North Waziristan. *NUST Journal of International Peace and Stability*, *3*(1), 102-107.

Ssempiira, J., Nambuusi, B., Kissa, J., Agaba, B., Makumbi, F., Kasasa, S., & Vounatsou, P. (2017). Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria prevalence in children less than 5 years in Uganda. *PloS one*, *12*(4), e0174948.

Sun, T., Chen, F., Zhong, L., Liu, W., & Wang, Y. (2019). GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore geology reviews*, *109*, 26-49.

Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., & Hu, Z. (2020). Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: a case study from southern Jiangxi Province, China. *Minerals*, *10*(2), 102.

Sunuprapto, H., Danoedoro, P., & Ritohardoyo, S. (2016). Evaluation of pan-sharpening method: applied to artisanal gold mining monitoring in Gunung Pani Forest area. *Procedia Environmental Sciences*, *33*, 230-238.

Sutton, M., Weiersbye, I., Galpin, J., & Heller, D. (2006). A GIS-based history of gold mine residue deposits and risk assessment of post-mining land uses on the Witwatersrand Basin, South Africa. *Mine Closure*, 667-678.

Sutton, M. W. (2013). *Use of remote sensing and GIS in a risk assessment of gold and uranium mine residue deposits and identification of vulnerable land use*

Swenson, J. J., Carter, C. E., Domec, J.-C., & Delgado, C. I. J. P. o. (2011). Gold mining in the Peruvian Amazon: global prices, deforestation, and mercury imports. *6*(4), e18875.

Tahmasebi, P., & Hezarkhani, A. (2010a). Application of adaptive neuro-fuzzy inference system for grade estimation; case study, Sarcheshmeh porphyry copper deposit, Kerman, Iran. *Journal of Basic and Applied Sciences*, *4*(3), 408-420.

Tahmasebi, P., & Hezarkhani, A. (2010b). Application of adaptive neuro-fuzzy inference system for grade estimation; case study, Sarcheshmeh porphyry copper deposit, Kerman, Iran. *Australian Journal of Basic and Applied Sciences*, *4*(3), 408-420.

Tahmasebi, P., & Hezarkhani, A. (2012). A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers & Geosciences*, *42*, 18-27.

Tamani, F., Hadji, R., Hamad, A., & Hamed, Y. (2019). Integrating remotely sensed and GIS data for the detailed geological mapping in semi-arid regions: case of Youks les Bains Area, Tebessa Province, NE Algeria. *Geotechnical and Geological Engineering*, *37*(4), 2903-2913.

Tan, M., Song, X., Yang, X., & Wu, Q. (2015). Support-vector-regression machine technology for total organic carbon content prediction from wireline logs in organic shale: A comparative study. *Journal of Natural Gas Science Engineering*, *26*, 792-802.

Tangestani, M., & Moore, F. (2001). Comparison of three principal component analysis techniques to porphyry copper alteration mapping: a case study, Meiduk area, Kerman, Iran. *Canadian Journal of Remote Sensing*, *27*(2), 176-182.

Tao, J., Zhang, N., Chang, J., Chen, L., Zhang, H., & Chi, Y. (2022). Unlabeled sample selection for mineral prospectivity mapping by semi-supervised support vector machine. *Natural Resources Research*, *31*(5), 2247-2269.

Tedesco, S. A. (2012). *Surface geochemistry in petroleum exploration*. Springer Science & Business Media.

Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., & Bui, D. T. (2019). A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using logitboost machine learning classifier and multi-source geospatial data. *Theoretical Applied Climatology*, *137*(1-2), 637-653.

Tessier, J., Duchesne, C., & Bartolacci, G. (2007). A machine vision approach to on-line estimation of run-of-mine ore composition on conveyor belts. *Minerals Engineering*, *20*(12), 1129-1144.

Thomas, R. (2019). The Impact of Illegal Artisanal Gold Mining on the Peruvian Amazon: Benefits of Taking a Direct Mercury Analyzer into the Rain Forest to Monitor Mercury Contamination.

Todorovski, L., & Džeroski, S. (2006). Integrating knowledge-driven and data-driven approaches to modeling. *ecological modelling*, *194*(1-3), 3-13.

Toghill, P. (2011). *The geology of Britain: an introduction*. Crowood.

Tompolidi, A.-M., Sykioti, O., Koutroumbas, K., & Parcharidis, I. (2020). Spectral unmixing for mapping a hydrothermal field in a volcanic environment applied on ASTER, landsat-8/OLI, and sentinel-2 MSI satellite multispectral data: the Nisyros (Greece) case study. *Remote Sensing*, *12*(24), 4180.

Triantafilis, J., Odeh, I., & McBratney, A. (2001). Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. *Soil Science Society of America Journal*, *65*(3), 869-878.

Tuba, E., Ribic, I., Capor-Hrosik, R., & Tuba, M. (2017). Support vector machine optimized by elephant herding algorithm for erythemato-squamous diseases detection. *Procedia computer science*, *122*, 916-923.

Turlapaty, A. C., Anantharaj, V. G., & Younan, N. H. (2010). A pattern recognition based approach to consistency analysis of geophysical datasets. *Computers & Geosciences*, *36*(4), 464-476.

Udvardi, B., Kovács, I. J., Fancsik, T., Kónya, P., Bátori, M., Stercel, F., Falus, G., & Szalai, Z. (2017). Effects of Particle Size on the Attenuated Total Reflection Spectrum of Minerals. *Applied spectroscopy*, *71*(6), 1157-1168.

Van Der Meer, F. (2004). Analysis of spectral absorption features in hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, *5*(1), 55-68.

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

Van der Werff, H., & Van der Meer, F. (2016). Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing. *Remote Sensing*, *8*(11), 883.

van Leeuwen, W. J. (2009). Visible, near-IR, and shortwave IR spectral characteristics of terrestrial surfaces. In (pp. 33-50): SAGE Publications Ltd. London.

Vanhellemont, Q. (2019). Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sensing of Environment*, *225*, 175-192.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, *10*(5), 988-999.

Varouchakis, E., & Hristopulos, D. (2013). Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environmental monitoring and assessment*, *185*(1), 1-19.

Vassilas, N., Perantonis, S., Charou, E., Tsenoglou, T., Stefouli, M., & Varoufakis, S. (2002). Delineation of lineaments from satellite data based on efficient neural network and pattern recognition techniques. 2nd Hellenic Conf. on AI, SETN-2002, Verma, A. K., Garg, P. K., & Prasad, K. H. J. A. J. o. G. (2017). Sugarcane crop identification from LISS IV data using ISODATA, MLC, and indices based decision tree approach. *10*(1), 16.

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*, 46-56.

Vermote, E. F., Tanré, D., Deuze, J. L., Herman, M., & Morcette, J.-J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(3), 675-686.

Voltz, M., & Webster, R. (1990). A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *Journal of Soil Science*, *41*(3), 473-490.

Vorster, C. (2012a). *Simplified Geology and Gold Deposits, South Africa Lesotho and Swaziland*. COUNCIL FOR GEOSCIENCE. http://www.geoscience.org.za/images/Maps/gold.gif

Vorster, C. (2012b). *Simplified Geology and Iron Deposits, South Africa Lesotho and Swaziland*. COUNCIL FOR GEOSCIENCE. http://www.geoscience.org.za/images/Maps/iron.gif

Wang, B., Shi, W. Z., & Miao, Z. L. (2015). Confidence Analysis of Standard Deviational Ellipse and Its Extension into Higher Dimensional Euclidean Space. *PloS one*, *10*(3). https://doi.org/ARTN e0118537

10.1371/journal.pone.0118537

Wang, Q., Li, F., Jiang, X., Wu, S., & Xu, M. (2020). On-stream mineral identification of tailing slurries of tungsten via NIR and XRF data fusion measurement techniques. *Analytical Methods*.

Wang, Q., Li, Q., Liu, H., Wang, Y., & Zhu, J. (2014). An improved ISODATA algorithm for hyperspectral image classification. 2014 7th International Congress on Image and Signal Processing,

Wang, X., Yuan, P., Mao, Z., & You, M. (2016). Molten steel temperature prediction model based on bootstrap feature subsets ensemble regression trees. *Knowledge-Based Systems*, *101*, 48-59.

Wang, Z., Zuo, R., & Dong, Y. (2019). Mapping geochemical anomalies through integrating random forest and metric learning methods. *Natural Resources Research*, *28*(4), 1285-1298.

Wei, Z., Xia, W., Canjiong, L., Renzhong, S., & Yuzhuo, W. (2016). Determination of Flow Accumulation Threshold Based on Multiple Regression Model in Raster River Networks Extraction. *Transactions of the Chinese Society for Agricultural Machinery*, *10*, 018.

Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, *54*(1), 41-59.

Wessels, R. L., Vaughan, R. G., Patrick, M. R., & Coombs, M. L. (2013). High-resolution satellite and airborne thermal infrared imaging of precursory unrest and 2009 eruption at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, *259*, 248-269.

Wijaya, D. R., Afianti, F., Arifianto, A., Rahmawati, D., & Kodogiannis, V. S. (2022). Ensemble machine learning approach for electronic nose signal processing. *Sensing and Bio-Sensing Research*, *36*, 100495.

Wong, D. W. (2016). Interpolation: Inverse-Distance Weighting. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, 1-7.

Woodhead, J., & Landry, M. (2021). Harnessing the power of artificial intelligence and machine learning in mineral exploration—opportunities and cautionary notes. *SEG Discovery*(127), 19-31.

Wu, C., Wu, J., Luo, Y., Zhang, H., Teng, Y., & DeGloria, S. D. (2011). Spatial interpolation of severely skewed data with several peak values by the approach integrating kriging and triangular irregular network interpolation. *Environmental Earth Sciences*, *63*(5), 1093-1103.

Wu, X., & Zhou, Y. (1993a). Reserve estimation using neural network techniques. *Computers & Geosciences*, *19*(4), 567-575.

Wu, X., & Zhou, Y. (1993b). Reserve estimation using neural network techniques. *Computers & Geosciences*, *19*(4), 567-575.

Xiong, Y., & Zuo, R. (2020). Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. *Computers & Geosciences*, *140*, 104484.

Xu, S., Zhao, Q., Yin, K., Zhang, F., Liu, D., & Yang, G. (2019a). Combining random forest and support vector machines for object-based rural-land-cover classification using high spatial resolution imagery. *Journal of Applied Remote Sensing*, *13*(1), 014521-014521.

Xu, S., Zhao, Q., Yin, K., Zhang, F., Liu, D., & Yang, G. J. J. o. A. R. S. (2019b). Combining random forest and support vector machines for object-based rural-land-cover classification using high spatial resolution imagery. *13*(1), 014521.

Yan, J., & Wang, L. (2016). Suitability evaluation for products generation from multisource remote sensing data. *Remote Sensing*, *8*(12), 995.

Yang, C.-S., Kao, S.-P., Lee, F.-B., & Hung, P.-S. (2004). Twelve different interpolation methods: A case study of Surfer 8.0. Proceedings of the XXth ISPRS Congress,

Yang, H.-h., Wang, Q., Li, Y.-b., Lin, B., Song, Y., Wang, Y.-y., He, W., Li, H.-w., Li, S., & Li, J.-l. (2022). Geology and mineralization of the Tiegelongnan supergiant porphyry-epithermal Cu (Au, Ag) deposit (10 Mt) in western Tibet, China: A review. *China Geology*, *5*(1), 136-159.

Yang, J. J. J. G. I. S. (2008). Detecting landscape changes pre-and post surface coal mining in Indiana, USA. *14*(1), 36-43.

Yang, N., Zhang, Z., Yang, J., & Hong, Z. (2022). Mineral prospectivity prediction by integration of convolutional autoencoder network and random forest. *Natural Resources Research*, *31*(3), 1103-1119.

Yang, N., Zhang, Z., Yang, J., Hong, Z., & Shi, J. (2021). A convolutional neural network of GoogLeNet applied in mineral prospectivity prediction based on multi-source geoinformation. *Natural Resources Research*, *30*(6), 3905-3923.

Yeomans, C. M., Middleton, M., Shail, R. K., Grebby, S., Lusty, P. A. J. C., & Geosciences. (2019). Integrated Object-Based Image Analysis for semi-automated geological lineament detection in southwest England. *123*, 137-148.

Yin, J., & Li, N. (2022). Ensemble learning models with a Bayesian optimization algorithm for mineral prospectivity mapping. *Ore geology reviews*, *145*, 104916.

Yousefi, M. (2017). Analysis of zoning pattern of geochemical indicators for targeting of porphyry-Cu mineralization: a pixel-based mapping approach. *Natural Resources Research*, *26*(4), 429-441.

Yousefi, M., Kamkar-Rouhani, A., & Carranza, E. J. M. (2012). Geochemical mineralization probability index (GMPI): a new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping. *Journal of Geochemical Exploration*, *115*, 24-35.

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, *13*(5), 839-856.

Yu, L., Porwal, A., Holden, E.-J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, *45*, 229-239.

Yu, L., Xu, Y., Xue, Y., Li, X., Cheng, Y., Liu, X., Porwal, A., Holden, E.-J., Yang, J., & Gong, P. J. O. G. R. (2018). Monitoring surface mining belts using multiple remote sensing datasets: A global perspective. *101*, 675-687.

Zaini, N., Van Der Meer, F., Van Ruitenbeek, F., De Smeth, B., Amri, F., & Lievens, C. (2016). An alternative quality control technique for mineral chemistry analysis of Portland cement-grade limestone using shortwave infrared spectroscopy. *Remote Sensing*, *8*(11), 950.

Zangerl, C., Loew, S., & Eberhardt, E. J. E. g. h. (2006). Structure, geometry and formation of brittle discontinuities in anisotropic crystalline rocks of the Central Gotthard Massif, Switzerland. *99*(2), 271-290.

Zanter, K. (2016). Landsat 8 (L8) data users handbook. *Landsat Science Official Website. Available online: https://landsat. usgs. gov/landsat-8-l8-data-users-handbook (accessed on 20 January 2018)*.

Zeng, X., Liu, Z., He, C., Ma, Q., & Wu, J. J. J. o. A. R. S. (2017). Detecting surface coal mining areas from remote sensing imagery: an approach based on object-oriented decision trees. *11*(1), 015025.

Zhang, X., Pazner, M., & Duke, N. (2007). Lithologic and mineral information extraction for gold exploration using ASTER data in the south Chocolate Mountains (California). *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*(4), 271-282.

Zhang, X., Song, S., Li, J., & Wu, C. (2013). Robust LS-SVM regression for ore grade estimation in a seafloor hydrothermal sulphide deposit. *Acta Oceanologica Sinica*, *32*(8), 16-25.

Zhang, Y., & Mishra, R. K. (2012). A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion. Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International,

Zhang, Y., Song, S., You, K., Zhang, X., & Wu, C. (2017a). Relevance vector machines using weighted expected squared distance for ore grade estimation with incomplete data. *International Journal of Machine Learning and Cybernetics*, *8*, 1655-1666.

Zhang, Y., Song, S., You, K., Zhang, X., & Wu, C. (2017b). Relevance vector machines using weighted expected squared distance for ore grade estimation with incomplete data. *International Journal of Machine Learning and Cybernetics*, *8*(5), 1655-1666.

Zhang, Z., Yang, J., Wang, Y., Dou, D., & Xia, W. (2014). Ash content prediction of coarse coal by image analysis and GA-SVM. *Powder Technology*, *268*, 429-435.

Zhao, Z.-F., Zhou, J.-X., Lu, Y.-X., Chen, Q., Cao, X.-M., He, X.-H., Fu, X.-H., Zeng, S.-H., & Feng, W.-j. (2021). Mapping alteration minerals in the Pulang porphyry copper ore district, SW China, using ASTER and WorldView-3 data: Implications for exploration targeting. *Ore geology reviews*, *134*, 104171.

Zou, Y., Chen, Y., & Deng, H. (2021). Gradient boosting decision tree for lithology identification with well logs: a case study of zhaoxian gold deposit, shandong peninsula, China. *Natural Resources Research*, *30*(5), 3197-3217.

Zuluaga, M. C., Norini, G., Lima, A., Albanese, S., David, C. P., & De Vivo, B. (2017). Stream sediment geochemical mapping of the Mount Pinatubo-Dizon Mine area, the Philippines: implications for mineral exploration and environmental risk. *Journal of Geochemical Exploration*, *175*, 18-35.

Zuo, R. (2020). Geodata science-based mineral prospectivity mapping: A review. *Natural Resources Research*, *29*, 3415-3424.

Zuo, R., Carranza, E. J. M., & Wang, J. (2016). Spatial analysis and visualization of exploration geochemical data. *Earth-Science Reviews*, *158*, 9-18.

Zuo, R., Kreuzer, O. P., Wang, J., Xiong, Y., Zhang, Z., & Wang, Z. (2021). Uncertainties in GIS-based mineral prospectivity mapping: Key types, potential impacts and possible solutions. *Natural Resources Research*, *30*, 3059-3079.

Zuo, R., & Xiong, Y. (2018). Big data analytics of identifying geochemical anomalies supported by machine learning methods. *Natural Resources Research*, *27*(1), 5-13.

APPENDIX A. LIST OF RESEARCH PUBLICATIONS USED IN THIS SYSTEMATIC LITERATURE REVIEW AFTER QUALITY CONTROL AND ASSESSMENT

Study ID	Title	Reference
S-01	Reserve Estimation Using Neural Network Techniques.	(Wu & Zhou, 1993b)
S-02	Data Segmentation and Genetic Algorithms for Sparse Data Division in Nome Placer Gold Grade Estimation Using Neural Network and Geostatistics.	(Samanta et al., 2002)
S-03	A Comparative Study of the Performance of Single Neural Network vs. Adaboost Algorithm Based Combination of Multiple Neural Networks for Mineral Resource Estimation.	(B Samanta et al., 2005)
S-04	Comparing the Predictive Performance of Neural Networks with Ordinary Kriging in a Bauxite Deposit.	(Biswajit Samanta et al., 2005)
S-05	Comparative Evaluation of Neural Network Learning Algorithms for Ore Grade Estimation.	(B Samanta et al., 2006)
S-06	A Machine Vision Approach to On-Line Estimation of Run-Of-Mine Ore Composition on Conveyor Belts.	(Tessier et al., 2007)
S-07	Genetic Algorithm-Based Neural Network Learning Parameter Selection for Ore Grade Evaluation of a Limestone Deposit.	(Chatterjee et al., 2008)
S-08	Ore Grade Prediction Using a Genetic Algorithm and Clustering Based Ensemble Neural Network Model.	(Chatterjee, Bandopadhyay, et al., 2010a)
S-09	Construction of a Radial Basis Function Network Using an Evolutionary Algorithm for Grade Estimation in a Placer Gold Deposit.	(Samanta & Bandopadhyay, 2009)

S-10	Application of Adaptive Neuro-Fuzzy Inference System for Grade Estimation; Case Study, Sarcheshmeh Porphyry Copper Deposit, Kerman, Iran.	(Tahmasebi & Hezarkhani, 2010a)
S-11	Image-based Quality Monitoring System of Limestone Ore Grades.	(Chatterjee, Bhattacherjee, et al., 2010)
S-12	Radial Basis Function Network for Ore Grade Estimation.	(Samanta, 2010)
S-13	Machine Learning Algorithms and Their Application to Ore Reserve Estimation of Sparse and Imprecise Data.	(Dutta et al., 2010)
S-14	Adaptive Ore Grade Estimation Method for the Mineral Deposit Evaluation.	(Li et al., 2010)
S-15	Ore Grade Estimation by Feature Selection and Voting Using Boundary Detection in Digital Image Analysis.	(Perez et al., 2011)
S-16	Geochemical Fingerprinting of Coltan Ores by Machine Learning on Uneven Datasets.	(Savu-Krohn et al., 2011)
S-17	Goodnews Bay Platinum Resource Estimation Using Least Squares Support Vector Regression with Selection of Input Space Dimension and Hyperparameters.	(Chatterjee & Bandopadhyay, 2011b)
S-18	A Hybrid Neural Networks-Fuzzy Logic-Genetic Algorithm for Grade Estimation.	(Tahmasebi & Hezarkhani, 2012)
S-19	Hybrid Self-Adaptive Learning Based Particle Swarm Optimization and Support Vector Regression Model for Grade Estimation.	(Li et al., 2013)
S-20	An SVM-based Machine Learning Method for The Separation of Alteration Zones in Sungun Porphyry Copper Deposit.	(Abbaszadeh et al., 2013)
S-21	Robust LS-SVM Regression for Ore Grade Estimation in A Seafloor Hydrothermal Sulphide Deposit.	(Zhang et al., 2013)
S-22	Ash Content Prediction of Coarse Coal by Image Analysis and GA-SVM.	(Zhang et al., 2014)

S-23	Classification of Iron Ores by Laser-Induced Breakdown Spectroscopy (LIBS) Combined with Random Forest (RF).	(Sheng et al., 2015)
S-24	Classification of Gold-Bearing Particles Using Visual Cues and Cost-Sensitive Machine Learning.	(Horrocks et al., 2015)
S-25	Integrating Artificial Neural Networks and Geostatistics for Optimum 3D Geological Block Modeling in Mineral Reserve Estimation: A Case Study.	(Jalloh et al., 2016)
S-26	A Hybrid Simultaneous Perturbation Artificial Bee Colony and Back-Propagation Algorithm for Training a Local Linear Radial Basis Neural Network on Ore Grade Estimation.	(Jafrasteh & Fathianpour, 2017)
S-27	Investigation of General Regression Neural Network Architecture for Grade Estimation of An Indian Iron Ore Deposit.	(Das Goswami et al., 2017)
S-28	Development of Online Machine Vision System using Support Vector Regression (SVR) Algorithm for Grade Prediction of Iron Ores.	(Patel et al., 2017b)
S-29	Relevance Vector Machines Using Weighted Expected Squared Distance for Ore Grade Estimation with Incomplete Data	(Zhang et al., 2017b)
S-30	Comparison of Machine Learning Methods for Copper Ore Grade Estimation.	(Jafrasteh et al., 2018)
S-31	Combining Regression Kriging with Machine Learning Mapping for Spatial Variable Estimation	(Li et al., 2019)

APPENDIX B. PUBLICATION DOMAINS OF THE SELECTED RESEARCH STUDIES USED FOR SLR

Study ID	Title of Journal/Conference	Publisher
S-01	Computers & Geosciences	Elsevier
S-02	Exploration and Mining Geology	Canadian Institute of Mining, Metallurgy and Petroleum
S-03	The Journal of The Southern African Institute of Mining and Metallurgy	The South African Institute of Mining and Metallurgy
S-04	Mining Technology Transactions of the Institutions of Mining and Metallurgy: Section A	Taylor & Francis
S-05	Mathematical Geology	Springer
S-06	Minerals Engineering	Elsevier
S-07	Mining Technology Transactions of the Institutions of Mining and Metallurgy: Section A	Taylor & Francis
S-08	Mathematical Geosciences	Springer
S-09	Computers & Geosciences	Elsevier
S-10	Australian Journal of Basic and Applied Sciences	Australian Journal of Basic and Applied Sciences
S-11	Computers in Industry	Elsevier
S-12	Natural Resources Research	Springer
S-13	Journal of Intelligent Learning Systems and Applications	Scientific Research Publishing Inc
S-14	Mathematical and Computer Modelling	Elsevier
S-15	International Journal of Mineral Processing	Elsevier
S-16	Natural Resources Research	Springer
S-17	Natural Resources Research	Springer

S-18	Computers & Geosciences	Elsevier
S-19	Neurocomputing	Elsevier
S-20	Geochemistry	Elsevier
S-21	Acta Oceanologica Sinica	Springer
S-22	Powder Technology	Elsevier
S-23	Journal of Analytical Atomic Spectrometry	Royal Society of Chemistry
S-24	Mathematical Geosciences	Springer
S-25	International Journal of Mining Science and Technology	Elsevier
S-26	Neurocomputing	Elsevier
S-27	Arabian Journal of Geosciences	Springer
S-28	2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)	IEEE Xplore
S-29	International Journal of Machine Learning and Cybernetics	Springer
S-30	Computational Geosciences	Springer
S-31	IEEE Geoscience and Remote Sensing Letters	IEEE Xplore

APPENDIX C. THE TYPE OF ML-MODELS APPLIED IN THE RESEARCH STUDIES FOR MRE

Study ID	ML model applied for MRE
S-01	Support Vector Machine (SVM)
S-02	Coactive Neuro-Fuzzy Inference System based on Artificial Neural Networks Adaptive Neuro-Fuzzy Inference System (ANFIS)
S-03	Adaptive Neuro-Fuzzy Inference System (ANFIS) Artificial Neural Networks (ANN) Kriging
S-04	Support Vector Regression (SVR)
S-05	Support Vector Machine (SVM) Neural Network Kriging
S-06	Local Linear Radial Basis Function (LLRBF) neural network LLRBF network with skewed Gaussian activation function Simultaneous Perturbation Artificial Bee Colony (SPABC) algorithm Back Propagation (BP) Standard artificial bee colony Covariance Matrix Adaptation Evolution Strategy (CMAES) Particle Swarm Optimization (PSO) Support Vector Machine (SVM)
S-07	Support Vector Machine (SVM) Random Forest (RF)
S-08	Multi-layer perceptron neural network
S-09	Support Vector Machine (SVM)
S-10	Support Vector Machine (SVM)

	A naïve Bayes classifier A majority decision table
S-11	Least square support vector machine regression Inverse distance weight Ordinary kriging Back propagation neural network
S-12	Support Vector Machine (SVM)
S-13	Multi-layer feed forward neural network Simple kriging Ordinary kriging Kriging with linear drift function Kriging with a quadratic drift function Lognormal kriging
S-14	Support Vector Machines (SVM) Linear Programming Boosting
S-15	General regression neural network Multilayer perceptron neural network Ordinary kriging
S-16	Radial Basis Function (RBF)
S-17	Radial basis function network Ordinary kriging
S-18	Support Vector Machine (SVM) Regression
S-19	Single neural network Multiple neural network
S-20	Support Vector Machine (SVM)
S-21	Layered feedforward artificial neural network
S-22	Support Vector Regression (SVR)

S-23	Artificial Neural Networks (ANN) Geostatistics
S-24	Artificial Neural Networks (ANN) Ordinary kriging
S-25	Relevance vector machine Expected squared distance Weighted expected squared distance Inverse distance weighted Kriging
S-26	Neural networks Support Vector Regression (SVR) Ordinary kriging
S-27	Neural networks
S-28	Wavelet neural network
S-29	Multilayer feedforward neural network
S-30	Neural networks Random Forests (RF) Gaussian processes
S-31	Ordinary kriging Regression kriging Machine learning mapping Hybrid method