ARDHI UNIVERSITY

UTILIZING SURFACE GEOCHEMISTRY USING GEOSTATISTICS AND GIS ANALYSIS IN GEOCHEMICAL (GOLD) TARGET GENERATION

A Case Study of Mara Greenstone Belt

THOBIAS, LYDIA S.

BSc Geoinformatics

Dissertation

Ardhi University, Dar es Salaam

July, 2023

UTILIZING SURFACE GEOCHEMISTRY USING GEOSTATISTICS AND GIS ANALYSIS IN GEOCHEMICAL (GOLD) TARGET GENERATION

A Case Study of Mara Greenstone Belt

THOBIAS, LYDIA S.

A Dissertation Submitted to the Department of Geospatial Sciences and Technology in Partially Fulfilment of the Requirements for the Award of Bachelor of Science in Geoinformatics (BSc. GI) of Ardhi University

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance by the Ardhi University dissertation titled **"Utilizing Surface Geochemistry using Geostatistics and GIS Analysis in Geochemical (Gold) Target Generation, a case study of Mara Greenstone Belt"** in partial fulfillment of the requirements for the award of degree of Bachelor of Science in Geoinformatics at Ardhi University.

………………………………… ……………………………..

Dr. Anastazia Msusa **Mr. Iriaeli Mlay**

(Supervisor) (Supervisor)

Date ……………………… Date ………………………

DECLARATION AND COPYRIGHT

I, THOBIAS, LYDIA S. hereby declare that, the contents of this dissertation are the results of my own findings through my study and investigation, and to the best of my knowledge they have not been presented anywhere else as a dissertation for diploma, degree or any similar academic award in any institution of higher learning.

………………………………..

THOBIAS, LYDIA S.

22734/T.2019

(Candidate)

ACKNOWLEDGEMENT

I am deeply grateful to Almighty God for bestowing upon me knowledge, good health, and the ability to successfully complete my research. I acknowledge and hold dear the words from **Zechariah 4:6b**:

"Not by might, nor by power, but by my spirit, saith the LORD of hosts."

I would like to express my profound appreciation for the unwavering support and guidance provided by my supervisors, Mr. Iriaeli Mlay and Dr. Anastazia Msusa. Their sincerity, inspiration, guidance, and extensive knowledge have profoundly influenced me throughout the entirety of my dissertation.

I extend my heartfelt gratitude to the Shanta Mining Company Limited for their support and for providing the valuable data that complemented my study. I would also like to extend sincere thanks to Mr. Peter Myegeta and Dr. Julian Ijumilana for their close support, direction, vision, and assistance during my dissertation.

Furthermore, I would like to convey my utmost appreciation to Mr. Michael Mavura, Mr. Gadiel Mchau, and the staff of the Geospatial Sciences and Technology Department (DGST) for their constructive comments, encouragement, and unwavering support. I am truly grateful for their contributions.

I cannot overlook the invaluable support and companionship provided by my Geoinformatics classmates and colleagues. The camaraderie and assistance offered will forever be cherished. Additionally, I would like to express special thanks to my family for their unwavering moral support, patience, and consistent encouragement throughout this study.

I am fully aware that the completion of this research would not have been possible without the participation and assistance of numerous individuals who may not have been mentioned here. I sincerely appreciate and acknowledge their invaluable contributions.

I offer my heartfelt thanks to everyone who has played a role in this journey.

DEDICATION

This dissertation is the culmination of tireless dedication and sacrifices. It is with great pride and heartfelt sincerity that I dedicate this work to the individuals who have served as my inspiration. From my parents, Rachel Mathias and Benedict Anthony, to my dissertation supervisors, and even to my younger siblings, my classmates, and circle of friends who offered their assistance during the challenges faced while conducting this research.

I would also like to pay tribute to my guardian, the late Rev. Bertha Joseph Maghembe (may your soul rest in peace), who nurtured and educated me from childhood, shaping me into the person I am today. I am immensely grateful for everything you done for me. Your love and support are immeasurable, and I can never adequately express how much I treasure and appreciate you. Your absence has left a void that can never be filled, and their presence will always be cherished.

Above all, my utmost gratitude goes to our Almighty God, who has bestowed His blessings upon us in our daily lives. I am especially thankful for the strength, courage, patience, wisdom, time, and guidance provided throughout the realization of this work.

In conclusion, I extend my deepest gratitude and appreciation to all those who have played a significant role in this journey.

ABSTRACT

Successful gold exploration projects rely heavily on obtaining clear information regarding the relationship between gold, trace elements, and factors influencing mineralization. Soil geochemistry has proven to be a valuable tool in identifying potential exploration targets during the initial stages of exploration. The objective of this study was to establish targets for gold distribution, understand the association between gold and its pathfinder elements (such as Ag, Cd, Mo, Sb, Te, Tl, W, Bi, and Se), and identify the lithologies contributing to the overlying residual soils. Cluster analysis revealed that the elements associated with gold exhibited an average level of similarity, indicating that both associated elements could serve as promising pathfinders. Principal component analysis (PCA), factor analysis (FA), and Pearson's correlation matrix of transformed data for Ag, Cd, Mo, Sb, Te, Tl, W, Bi, and Se suggested that the source of gold mineralization might be related to hornblende gneisses interlayered with quartzite, tonalite, and tonalitic orthogneiss. Through the contour map and gridded map of gold and its pathfinder elements, it was observed that the anomalies and generated targets were localized in the northern and eastern parts of the area. These targets exhibited an ESE-WNW trend, nearly parallel to the shear zones, which appear to be controlling factors for the emplacement of gold mineralization. Additionally, SRTM DEM was utilized for extraction and mapping of lineaments which control mineralization, the extracted lineaments indicate that the presence of gold may be associated with the presence of lineaments, with gold showing a direct proportionality to the abundance of lineaments. This study has provided valuable insights into the association between gold, trace elements, and mineralization controlling factors. The identified lithologies, anomalies, and lineaments serve as potential targets for further exploration and contribute to the understanding of gold distribution in the study area

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

CA	Cluster Analysis
DEM	Digital Elevation Model
FA	Factor Analysis
MLFE	Manual Linear Feature Extraction
PCA	Principal Component Analysis
PGM	Platinum Group Metals
SRTM	Shuttle Radar Topographic Mission

CHAPTER ONE

INTRODUCTION

1.0 BACKGROUND

Mineral exploration is a complex and costly process that involves the identification of potential mineral deposits and the assessment of their economic viability. The general principle works by extracting pieces of geological information from several places, and extrapolating this over the larger area to develop a geological picture. Exploration works in stages of increasing sophistication, with cheap, cruder methods implemented at the start, and if the resultant information is economically interesting, this warrants the next, more advanced (and expensive) techniques. However, it is very rare to find sufficiently enriched ore bodies, and so most exploration campaigns stop after the first/couple of stages, (Google search). Geochemical exploration is one of the most effective methods for identifying new mineral deposits. It involves the collection and analysis of geochemical data from rocks, soils, stream sediments, and other materials to identify anomalous geochemical signatures that may indicate the presence of mineral deposits. However, the interpretation of geochemical data can be challenging, and the identification of potential mineral deposits requires the integration of multiple datasets.

Tanzania offers the exploration industry a favorable investment climate and operating environment with significant potential for new discoveries as many commodities still are highly under explored. GST has been involved in the mineral exploration projects. These projects have ranged from regional- to prospect-scale and have led to the discovery of a number of significant deposits such as Kahama Gold Mine and Geita Gold Mine which are now world class mines.

Much of the present exploration activity in Tanzania, is concentrated in gold, base metals, platinum group metals (PGM), uranium, gemstones, diamonds and industrial minerals. Tanzania has excellent geological databases, good infrastructure, attractive mineral policy and readily available exploration services. These factors make investing in Tanzania attractive and cost effective. Mineral exploration in Tanzania is governed by the mining act 2010, also the exploration tenements are provided to licensed areas to conduct exploration including geological sampling

Sampling is limited to specific designed location and it could be in regular or irregular grids depend on preliminary information on metal presence. Samples are managed so that there is no contamination between different samples and then are sent to laboratory for analysis to determine the chemical composition of metal of choice

In recent years, the use of geostatistics and GIS analysis has become increasingly popular in geochemical target generation. while GIS (Geographic Information Systems) is a software program used to organize, view, and analyze geographic data, geostatistics is a statistical method used to evaluate spatial data by providing interpolation or extrapolation capabilities. Based on the spatial distribution of geochemical anomalies and other geological and topographic factors, geostatistics and GIS analysis could be used to identify locations with significant mineral potential (Cheng et al., 2021).

Numerous researches have looked into the use of geostatistics and GIS analysis in the creation of geochemical targets. For example, (Nweke et al., 2021) used geostatistical techniques to identify areas with high potential for gold mineralization in the Ilesha schist belt in southwestern Nigeria. The authors used geochemical data, geological maps, and remote sensing data to generate a mineral prospectivity map that identified areas with high potential for gold mineralization. However, the type of metal being investigated may necessitate a different method of data preparation and analysis when the two spatial problem-solving methods are combined.

In this study, aim to investigate the effectiveness of 2D geostatistics and GIS analysis in geochemical data processing and target generation to identify potential mineral deposits in a selected study area. A range of geostatistical techniques and GIS spatial analysis (Geoprocessing) will be applied to generate a mineral prospectivity map that identifies areas with high potential for mineralization given the mineralization criteria. The results of this study can be used to guide mineral exploration programs and contribute to the sustainable development of the mining industry. The research on utilizing 2D geostatistics and GIS analysis in geochemical target generation contributes to the advancement of scientific knowledge in the fields of geology, geostatistics, and GIS. It can also serve as a basis for future research in the field of mineral exploration and mining.

1.1 RESEARCH PROBLEM STATEMENT

Exploration companies have collected extensive databases of geochemical data, providing valuable information on the concentrations of chemical elements in rocks. However, visually analyzing and interpreting this data can be challenging due to its complex nature and inherent imprecision. As a result, there is a growing need to employ technological approaches for effectively dealing with such vast amounts of geochemical data.

One of the main challenges is the difficulty in evaluating different geochemical datasets without interpolating the point data, primarily due to limited sampling. Furthermore, removing the geochemical background from interpretation studies poses an additional challenge. To address these issues, researchers are exploring the integration of geostatistical and Geographic Information System (GIS) methodologies. Geostatistics provides statistical tools for analyzing spatial patterns and relationships in the data, while GIS offers a platform for visualizing and analyzing the data within a spatial context. By combining these approaches, geologists can gain better insights and make informed interpretations of the geochemical data.

This integration allows for the interpolation of data points, filling in spatial gaps and providing a more comprehensive view of the geochemical distribution. GIS tools enable the visualization of interpolated data on maps, aiding in the identification of trends, anomalies, and areas of interest. Additionally, the combined approach helps in distinguishing meaningful geochemical patterns from background variations, supporting geologists in properly processing and interpreting the data. Ultimately, this integration of geostatistical and GIS methodologies enhances the understanding of geochemical characteristics, enabling more effective exploration and analysis of geochemical datasets.

1.2 OBJECTIVES OF THE RESEARCH

1.2.1 Main Objective

- To develop a methodology that combines geostatistical and GIS methodologies to effectively process and interpret geographically dispersed geochemical data to delineate areas potential for gold mineralization using surface geochemistry.

1.2.2 Specific Objectives

- To develop techniques for interpolating geochemical data to estimate element concentrations at unsampled locations, addressing the problem of restricted sampling and spatial gaps.
- To design methods for accurately remove the geochemical background from interpretation studies, allowing for a more precise analysis of anomalous concentrations.
- To integrate geostatistical tools with GIS platforms to visualize and analyze the interpolated data in a spatial context.
- Identify and map lineaments and anomalous trends based on the gridded maps and overlay them with geochemical data to identify areas of interest for further investigation.
- Determine the zones with high concentrations of gold and associated elements through multivariate analysis, highlighting potential target areas for further exploration.
- Generate contour maps and gridded maps using the ArcGIS software platform to visualize the distribution of geochemical elements and identify spatial patterns or anomalies.

1.3 RESEARCH QUESTION

- How could GIS and Geostatistics be combined as a robust tool in mineral prospectivity?
- Could GIS be used to thematically visualize multivariate geochemical data?

1.4 SCOPE AND LIMITATION

1.4.1 Scope

The study also intends to investigate the efficacy and efficiency of various methodologies and methods for integrating geostatistical and GIS approaches to provide a better understanding of the underlying spatial patterns of geochemical data and their relationship to geological features.

1.4.2 Limitation

This research is restricted to investigating the use of GIS and geostatistics to assess the geochemical data at hand and identify target locations with statistical high values. Since there are many variables that could influence the results, such as the quality of the geochemical data used, which may be constrained by the availability of sampling data, or the analytical techniques used to analyze the samples. It is therefore not the goal of this study to discuss much about the accuracy of the data and the science behind GIS and geostatistics towards the obtained results.

1.5 SIGNIFICANCE OF THE STUDY

The research has significant practical and scientific importance in the field of mineral exploration and mining. The main significances of this research are

- Improved efficiency and cost-effectiveness: By utilizing 2D geostatistics and GIS analysis, mineral exploration companies can more efficiently and cost-effectively identify and prioritize areas of mineralization potential. This can lead to more targeted exploration programs and a higher likelihood of discovering economically viable mineral deposits.

- Enhanced accuracy and precision: Geochemical data is complex and challenging to analyze due to non-normal distribution, multivariate geographic autocorrelation, and imprecision. By combining geostatistical and GIS approaches, the accuracy and precision of geochemical data analysis can be enhanced, leading to more accurate identification of mineral targets.

- Identification of previously undiscovered mineral deposits: The use of 2D geostatistics and GIS analysis can help to identify mineralization potential in areas that were previously overlooked or underexplored. This can lead to the discovery of previously unknown mineral deposits and the expansion of mining operations.

- Sustainability: Mineral exploration and mining activities can have significant impacts on the environment. By utilizing 2D geostatistics and GIS analysis to identify areas of mineralization potential more accurately and efficiently, exploration companies can minimize their environmental impact while still discovering economically viable mineral deposits.

- Advancement of scientific knowledge: The research on utilizing 2D geostatistics and GIS analysis in geochemical target generation contributes to the advancement of scientific knowledge in the fields of geology, geostatistics, and GIS. It can also serve as a basis for future research in the field of mineral exploration and mining.

1.6 BENEFICIARIES OF THE RESEARCH

Beneficiaries are those who benefit from the research topic, the following are beneficiaries of the research output

- Mineral exploration and mining companies: The research can benefit mineral exploration and mining companies by providing more efficient and accurate methods to identify and prioritize areas of mineralization potential. This can lead to more targeted exploration programs, more successful discoveries of economically viable mineral deposits, and reduced exploration costs.

- Investors and shareholders: The research can benefit investors and shareholders by increasing the probability of successful mineral discoveries and ultimately leading to increased profits for mining companies.

- Governments and regulatory bodies: The research can benefit governments and regulatory bodies by improving the accuracy and precision of mineral exploration and mining activities, leading to better management of natural resources, and ensuring compliance with environmental and social regulations.

- Local communities: The research can benefit local communities by reducing the environmental impact of mineral exploration and mining activities, providing employment opportunities, and contributing to local economic development.

- Society and the environment: The research can benefit society and the environment by improving the sustainability of mineral exploration and mining activities, reducing environmental impacts, and ensuring responsible natural resource management.

1.7 EXPECTED OUTPUTS

- Geoprocessing analytical results, proposed target areas that meet the geological criteria for mineral prospectivity in the study area.
- Geodatabase of the GIS files resulting from converting raw data from MS Access and MS Excel formats

1.8 STUDY AREA

1.8.1 Geographical Location

The Mara Greenstone Belt is located in Tanzania, East Africa. It is part of the larger Lake Victoria Goldfields, which encompasses several greenstone belts known for their gold mineralization. The Mara Greenstone Belt is situated in the northwestern region of Tanzania, near the country's border with Kenya. It covers an area that includes parts of the Mara and Musoma districts within the Mara Region of Tanzania. Mara Greenstone Belt is generally located in the following approximate coordinates: Latitude: 1.6°S to 2.3°S Longitude: 34.2°E to 35.0°E

In terms of activities conducted in the Mara Greenstone Belt, the area is known for its significant gold mineralization. Therefore, gold exploration and mining activities are prominent in this region. Exploration companies and mining operations are involved in prospecting, drilling, geological mapping, geochemical sampling, and the extraction of gold-bearing ore. These activities aim to identify and develop economically viable gold deposits within the Mara Greenstone Belt. Additionally, associated activities such as environmental monitoring, infrastructure development, and community engagement may also take place in the area.

1.8.2 Geology of the study area

The Mara Greenstone Belt is a geological formation that is part of the larger Lake Victoria Goldfields in East Africa. It is characterized by a complex assemblage of rocks that represent a significant period of Earth's history. The Mara Greenstone Belt consists of a variety of rock types, including greenstone rocks, metavolcanics, metasediments, and intrusive igneous rocks. Greenstone rocks, such as basalt and andesite, are volcanic in origin and commonly exhibit a greenish hue due to the presence of minerals like chlorite and amphibole. The rocks have undergone regional metamorphism, which has altered their original characteristics. Metamorphism has resulted in the formation of various metamorphic rocks, such as schists and gneisses, and has led to the development of foliation and banding within the rocks. The Belt is situated within a region of extensive tectonic activity. It is associated with the Tanzanian Craton, a stable continental crust, and is located along the eastern edge of the craton. The formation of the belt is related to multiple episodes of volcanic activity, sediment deposition, and deformation during the Archean and Proterozoic eras.

The Mara Greenstone Belt is renowned for its gold mineralization. The geological processes and structures within the belt have created favorable conditions for the concentration of gold-bearing fluids. Hydrothermal systems associated with volcanic activity and deformation have played a crucial role in the deposition of gold within the rocks. The gold mineralization in the Mara Greenstone Belt occurs in quartz veins, shear zones, and disseminated form within the host rocks. In case of structural features, the belt exhibits a complex network of structural features, including faults, shear zones, and fold structures. These structures have been influenced by regional tectonic events, and they play a significant role in controlling the distribution of gold mineralization and the overall geology of the belt.

The geology of the Mara Greenstone Belt reflects a dynamic geological history involving volcanic activity, sedimentation, metamorphism, deformation, and gold mineralization. The understanding of these geological characteristics is crucial for exploration companies and geologists involved in mineral exploration and mining activities in the region.

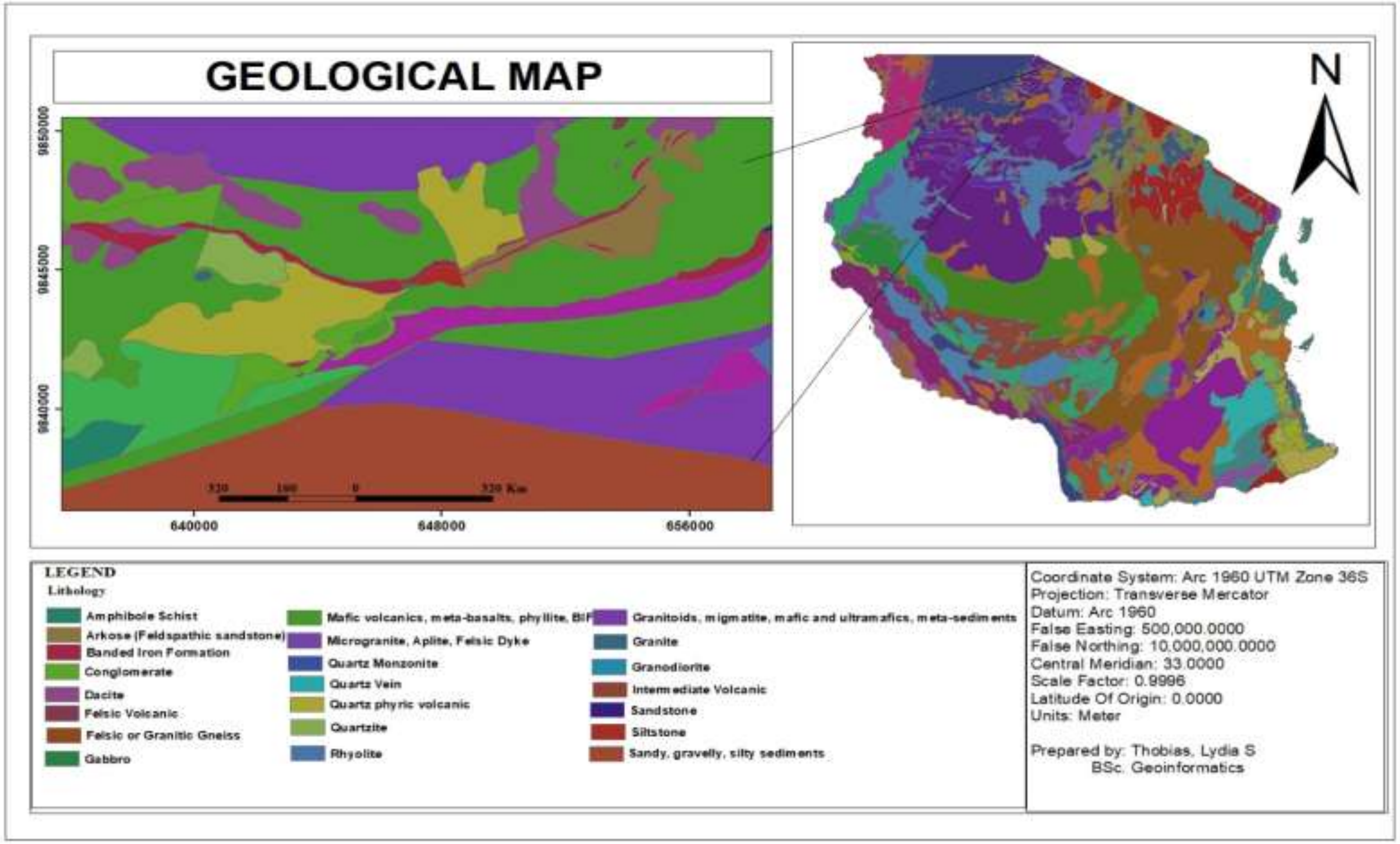

Figure 1: 1 Geological Map of the study area

CHAPTER TWO

LITERATURE REVIEW

2.0 OVERVIEW

This chapter reviews different literature relating to mineral exploration, gold mineralization and sampling, a review of factors influencing methods used to integrate geochemical data to identify potential mineral deposits.

2.1 MINERAL EXPLORATION

Mineral exploration is a complex and costly process that involves the identification of potential mineral deposits and the assessment of their economic viability. The general principle works by extracting pieces of geological information from several places, and extrapolating this over the larger area to develop a geological picture. Exploration works in stages of increasing sophistication, with cheap, cruder methods implemented at the start, and if the resultant information is economically interesting, this warrants the next, more advanced (and expensive) techniques. However, it is very rare to find sufficiently enriched ore bodies, and so most exploration campaigns stop after the first/couple of stages, (Google search)

Mineral reserves and resources, annual production versus consumption, demand versus supply, and index of per capita spending on commodities are measures that distinguish the rank of a country as developed, developing, and underdeveloped. Existing nonrenewable mineral reserve bases are limited and tend to reduce progressively with mine production. On the other hand, demand for minerals/metals gradually increases at w10% with annual growth of the population and uplifting standards of living. However, part of shortfall between demand and supply is partially substituted by recycling and reuse. (Evans, 2006; Haldar, 2013).

The process of mineral discovery and development to target production has a long gestation period of 5-10 years. Mineral sectors expect high-risk tolerance at all levels, extensive time, and rich sources for sustained cash flow. A small business unit in this field may often end its brief tenure with a total loss because of a failure to make a financial return. Many discoveries are not viable at current market prices, indicating the investment as a total loss. However, one discovery out of 100 or even 1000 attempts may pay back all the effort. The policymaker must plan timely allocation of funds for exploration and technology research of various mineral types, based on long-term demand and supply trends. An investment friendly environment, transparency, and political commitment of federal/state governments are essential for mineral development in any country (Evans, 2006; Haldar, 2013).

2.1.1. Mineral Exploration in Tanzania

Tanzania offers the exploration industry a favorable investment climate and operating environment with significant potential for new discoveries as many commodities still are highly under explored. GST has been involved in the mineral exploration projects. These projects have ranged from regional- to prospect-scale and have led to the discovery of a number of significant deposits such as Kahama Gold Mine and Geita Gold Mine which are now world class mines, (GST, 2023). Much of the present exploration activity in Tanzania, is concentrated in gold, base metals, platinum group metals (PGM), uranium, gemstones, diamonds and industrial minerals. Tanzania has excellent geological databases, good infrastructure, attractive mineral policy and readily available exploration services. These factors make investing in Tanzania attractive and cost effective, (GST,2023) One of the roles of GST is to acquire geoscientific data and information from new areas and mineral prospects to encourage further evaluation by the private sector. All discoveries and prospects are developed by the private sector through licensing by the Ministry of Energy and Minerals. Under the Mining Policy, 1997, the government has no role in the mining business, but is only promoter and facilitator, (GST, 2023) GST also provides confidential and customized expert services to exploration, mining companies and small-scale miners. These include all aspects and scales of mineral exploration and prospect evaluation, from planning and implementing regional exploration programs (geological, geochemical and geophysical), to detailed mineralogical studies and deposit modelling (GST, 2023)

Mineral exploration in Tanzania is governed by the mining act 2010 which states “An Act to re-enact with substantial amendments the provisions that regulate the law relating to prospecting for minerals, mining, processing and dealing in minerals, to granting, renewal and termination of mineral rights, payment of royalties, fees and other charges and any other relevant matters.

In particular section 41(4)(h) which holds that all applications for licenses must be accompanied, inter alia, by a proposed plan with respect to the Employment and training of citizens of Tanzania and succession plan for expatriate employees, if any as may be required by the Employment and Labor Relations Act; section 42(1)(d) which holds that the Minister is to take into account the size and nature of the proposed mining operations, the applicant's proposals for the employment and training of citizens of Tanzania and succession plan on expatriate employees and plan for procurement of goods and services available in the United Republic are adequate before granting a license; section 44(d)(iii) states that a mining license must contain, inter alia, information on the employment and training of citizens of Tanzania and succession plan, as required by the Employment and Labor Relations Act; section 47(b) states that holders of a special mining license are to employ and train citizens of Tanzania and implement succession plan on expatriate employees in accordance with his proposals as appended to the special mining license; section 49(2)(f) states that information on employment must be appended to a mining license application; section 72(2) regulates the entitlement to indemnities for employees in case of closer of a mine; section 111(2)(d) regulates insurance to cover indemnities for employees; licensees are also to keep records of the number of persons employed (Schedule one)", (ILO, 2023).

2.2 GEOLOGICAL BACKGROUND

In this part the Regional Geology and Geology of Lake Victoria Gold Field will be discussed;

2.2.1 Regional Geology

The Tanzania Craton extends from central Tanzania to western Kenya and south east Uganda, and is divided into two main geological units; The deformed high grade metamorphic terrain of central Tanzania locally known as Dodoman belt and the well-preserved low-grade granite-greenstone terrain of northern Tanzania (Manya and Maboko, 2008). The Dodoman belt is made up of granites, granodiorites, granitic gneiss and migmatites where the low-grade granite greenstone terrain of northern Tanzania consist of mafic to felsic volcanic rocks as well as meta-sedimentary rocks including shales, chert and Banded Iron Formation (BIF) (Manya and Maboko 2008, Mtoro et al.,2009). The low-grade granite-greenstone belt consist of well-preserved greenstone belts which include the Sukumaland, Shinyanga-Malita, Musoma-Mara, Iramba-Sekenke, Nzega and Kilimafedha greenstone belts (Manya and Maboko 2008), (Fig. 2-1). The greenstone belts of Tanzania Craton like most other greenstone belt in the world have suffered deformation and greenschist facies metamorphism (Sanislav et al., 2014) that result to the green color.

The granite greenstone belts stratigraphically comprise of greenstone which are traditionally assigned to the Nyanzian and Kavirondian Supergroup (Barth, 1990). The

Nyanzian Supergroup is predominantly made of volcano-sedimentary assemblages comprising pillow basalts, andesites, dacites and rhyolites, BIF, sandstone, shales and siltstones. The Kavirondian is largely sedimentary dominated Supergroup with most of its rock derived from the Nyanzian Supergroup and consist of conglomerate, sandstone, siltstone, and rare horizons of volcanic rocks (Kwelwa et al., 2013). Manya, (2006) noted that the Lake Victoria goldfield greenstone belts were formed at different time intervals (age bracketed being 2.8-2.65 Ga) and different tectonic setting from back arc to continental arcs. The Sukumaland greenstone belt comprises partially connected greenstone fragments (Borg and Schleton, 1997), the largest fragment in the northern part of the belt has been reclassified as the Geita Greenstone Belt (Sanislav, 2014).

2.2.2. Geology of the Lake Victoria Gold Field (LVGF)

The Lake Victoria Gold Field (LVGF) is situated on the stable Archaean Tanzania Craton, the central nucleus of the preserved Archaean continental crust in East Africa. The Craton extends from central Tanzania towards southwestern Kenya and is bordered by the Neoproterozoic Mozambique Belt rocks towards the east. Further towards southeast Uganda, the Craton is bounded by the 2.0 Ga Proterozoic metamorphic rocks, which locally constitute the Ruwenzori Belt (Clifford, 1970; Borg and Shacklecton, 1997). The western side the Craton is bordered by the newly defined Mesoproterozoic (1.37 Ga) Bukoban rocks, these rocks are comprised of gently folded to flat lying tabular and commonly reddish sedimentary units i.e., shales, siltstones, sandstones, arkoses, ortho-quartzites and dolomitic limestones with cherts, stromatoliths and occasional ooliths. These sediments are typically anorogenic and of continental origin and are partly associated with flood basalts and andesites (Pinna et. al., 2004). The southern part of the Archaean Tanzania Craton is bordered by the Paleoproterozoic high grade metamorphic Ubendian Belt, which joins the Mesoproterozoic (ca. 1.37-1.30 Ga) Karagwe-Ankolian metasediments to the far west. The latter is comprised of a variety of high-grade metamorphic rocks of both sedimentary and igneous origin, e.g., biotite-garnet-kyanite schists, quartzose-garnetiferous gneisses, amphibolite and hornblende gneisses, phyllites, crystalline limestones, graphitic schists, and granulites. Structural features of the Ubendian rocks were caused by a strong deformation that gave rise to both cataclastic and crystalloblastic deformation, which also resulted in the presence of mylonites, e.g., in the Livingstone Mountains and also phylonites found everywhere in the belt (Boniface, 2009). The eastern part of the Archaean Tanzania Craton is bordered by the Paleoproterozoic

Usagaran and Neoproterozoic Mozambique Belts. The belts are extensively migmatized and comprise high-grade metamorphic rocks ranging from schists to gneisses. Granulites and biotite gneisses of pelitic origin occupy a large portion of the total thickness of the Usagaran Belt, which is also considered to be the counterpart of the Ubendian Belt (Pinna et al., 2004). The Archaean Tanzania Craton has been subdivided into two main terranes: the deformed high-grade metamorphic terrane of central Tanzania and the low-grade granite–greenstone terrane of northern Tanzania, southeastern Uganda, and south-western Kenya (Clifford, 1970). Basing on the recent tectonic architecture developed from the syntheses of geochronological studies, which combined new SHRIMP U-Pb data and the interpretation of aero-geophysical data, Kabete et al. (2012) subdivided the country into seven Superterranes. Among the Superterranes, the East Lake Victoria Superterrane, the Lake Victoria-Lake Eyasi Superterrane, and the Lake Nyanza Superterrane are located in the northern part of the Craton. With regard to gold mineralization in the region, the East Lake Victoria Superterrane and the Lake Nyanza Superterrane comprise all major gold deposits discovered so far in the Archaean Tanzania Craton. However, by following the traditional tectonic domains of Barth (1990), the Archaean Tanzania Craton is composed of three lithostratgraphic/tectonic domains, namely the Kavirondian Supergroup, Nyanzian Supergroup and Dodoman Belt.

2.3 GOLD EXPLORATION AND PROSPECTING

Prospecting and exploration are the two processes for searching for mineral deposits. Gold exploration and prospecting can be performed by traditional methods, geological methods and remote sensing techniques as discussed below:

2.3.1 Traditional Prospecting Methods

There are different methods to find gold, in this subsection most common traditional methods will be discussed:

- **Gold Panning**

Gold panning is the traditional prospecting technique of separating particles of greater specific gravity (especially gold) from soil or gravels by washing in a pan with water. Gold pans are available in numerous sizes and shapes. However, the standard American gold pan at the top is perceived to be around 15 inches to 18 inches in diameter, while in depth it is two to two and half inches with the sides of the gold pan sloping 30 to 45 degrees

(Saliou and Huang, 2020). Additionally, the gold pans are built of plastic or metal. Figure 2-1 shows metal gold pans.

Hence, at the bottom of the gold pans heavy minerals are concentrated while at the top is where the lighter materials are removed to ensure that the basic operation of gold panning is simple. Gold panning has been widely utilized as the primary prospecting method in the early times. Nonetheless, it is evident that gold panning is exceedingly limited since it only recovers coarse gold while fine grained gold is normally washed away with the pebbles. This is due to most experienced panners can only process small amounts of gravel. In the modern world, gold panning has been used for cleaning concentrate and prospecting. The benefits of gold panning such as portability, immediate availability, and low price make it a vital instrument for prospectors. Gold panning is still being utilized for gathering geologic information.

Figure 2- 1: Gold Pans (Source: Saliou and Huang, 2020)

- **Dry Washing**

Dry washing is a gold prospecting method that uses a dry washer. Basically, a dry washer is perceived to be a waterless and short sluice. The dry washer uses pulsations of air via a porous medium to isolate gold from sand. In a dry washer, screened gravel goes down through inclined riffle boxes with cross riffles. The box’s bottom comprises of canvas or some other material/fabric (Melchiorre et al., 2017). Underneath the riffle boxes there are bellows that blow in strong and short puffs through the fabric/canvas. This process provides a classifying action and combined shaking to the material. Hence, the gold sinks downwards to the canvas while being held by the riffles. On the other hand, the waste goes over the riffles and out of the dry washer. Dry washer is made up of a frame that a heavy and well-braced screen is enclosed with burlap overlain with fly screen and/or window while being enclosed with fine linen. A power dry washer is estimated washer is estimated to process around 21 cubic feet of screened material on an hour basis (Makshakov and Kravtsova, 2018). On the other hand, hand-powered dry washers which are operated by two individuals are estimated to process approximately one or more cubic yards per eight hours. Nonetheless, this depends on the handled material size. The dry washer prospecting technique affects the gold or needs to be entirely disintegrated and dry. If the gold ore is considerably damp then prior to treatment it needs to be dried. Sun can be used to dry the gold ore for small-scale work. The dry gold ore in operations is put into the dry washer's vibrating screen where the gold fines drop through the box riffles while the oversized gold falls off the dry washer's edge. The screen and bellows of the dry washer are powered by small engines or operated by hand. The dry washer's bellows are operated at approximately 250 pulsations per minute with around 3 inches stroke (Rieck, R., 2017). However, there can be variation of these figures depending on the gold fineness and processed material's coarseness.

- **Metal detector**

A metal detector is an electronic device designed to find metal buried in the ground. The detector has a long shaft with a search coil that you hover just above the surface of the ground, and when it goes over a piece of metal you will get an audible sound (Melchiorre et al., 2017). While metal detectors are extremely effective for finding gold, they also have a very steep learning curve. To be successful, it is important to invest in a quality metal detector that is designed specifically for finding gold. There is literally TONS of metal in the ground; old rusty nails, cans, boot tacks, lead bullets, brass casings, etc. And a prospector has to be able to distinguish the sound of a gold nugget from the thousands of pieces of "trash" targets out there. Even a highly skilled metal detector operator will spend a lot of time digging trash. Unlike other methods, a metal detector will miss a lot of small gold because it requires a large enough piece of gold to get an audible tone, but if you are truly interested in finding large nuggets, a metal detector might be your best option.

2.3.2 Geological Prospecting Methods

In this subsection the geological prospecting methods were discussed which are:

- **Geological Mapping**

Geological map is a special purpose map made to show geological features, rock units/strata are shown by color or symbols to indicate where they are exposed at the surface. Geological map can be prepared from base map by georeferencing and digitization methods. The identification of rocks, minerals and an understanding of the environment in which they are formed. These surveys aim to find what rock types occur at or close to the surface and how these rock types are related to each other i.e., their boundaries, ages, and structure. A geological survey can be undertaken using a number of methods depending on the size of a region and the amount of information that is required (Moon et al., 2006)

i. Remote Sensing - some geological mapping can be done using satellite remote sensing methods. While most of these methods rely on geophysical rather than pure geological data, the use of this method can give broad scale views of surface geological structures such as folding, faulting, igneous intrusions etc.
ii. Air photo interpretation - this can give a broad overview of the geological relationships of
iii. an area with no detailed knowledge of the mineral composition or fabric of the rocks.
iv. Outcrop surveys - this is normally achieved by geologists driving along roads and walking traverses along creeks and rivers mapping the outcropping rock types. This can give a regional view of the rock types and their mineral content and fabric, but often no clear understanding of the relationships between rock layers (unless outcrop is exceptional).
v. Geology interpretation surveys-these more detailed outcrop surveys where geological boundaries are established and interpreted in a small area.

A geological map is invariably the first step in any mineral exploration program, and it remains an important control document for all subsequent stages of exploration and mining, including drilling, geochemistry, geophysics, geostatistics and mine planning. In an operating mine, geological mapping records the limits to visible ore in mine openings, and provides the essential data and ideas to enable projection of assay information beyond the sample points. Making a geological map is thus a fundamental skill for any exploration or Mine Geologist (Moon et al., 2006).

- **Geobotanical Prospecting**

Geobotanical also known as biogeochemical prospecting. Biogeochemical prospecting is established on the basis that trees and other plants might uptake some elements from primary substrate while accumulating then in their foliage and roots (Dunn, 2007). This is supported by Brooks who indicate that plants that collect increased metal concentrations from the substrate might be utilized as indicators plants for biogeochemical prospecting (Brooks, 1983). Numerous studies illustrate that biogeochemical exploration is a successful and effective mineral prospecting technique (Jung et al., 2011; Reid and Hill, 2010). Gold prospectors through using biogeochemical prospecting are capable of obtaining information regarding mineral deposits which happen at substantial depths. Nonetheless, there are issues in interpreting the obtained information which might make this prospecting technique to be impractical under several field conditions. This is evident in that some plants or bacteria due to their genetic makeup might concentrate elements selectively in their leaves, stems or roots in higher concentrations as compared in the underlying rocks and soil. It is vital for prospectors where possible to analyze rock and soil samples prior to come to the conclusion that the geo-botanical anomaly suggests the presence of gold ores in the region (Saliou and Huang, 2020). For example, in the Katanga region of southern Zaire, a small blue-flowered mint, Acrocephalus robertii, is restricted entirely to copper-bearing rock outcrops. Another example from Australian researchers confirms that deep-rooted this Eucalyptus trees (Figure 2-2) take gold from ore deposits underground and transport them into their leaves.

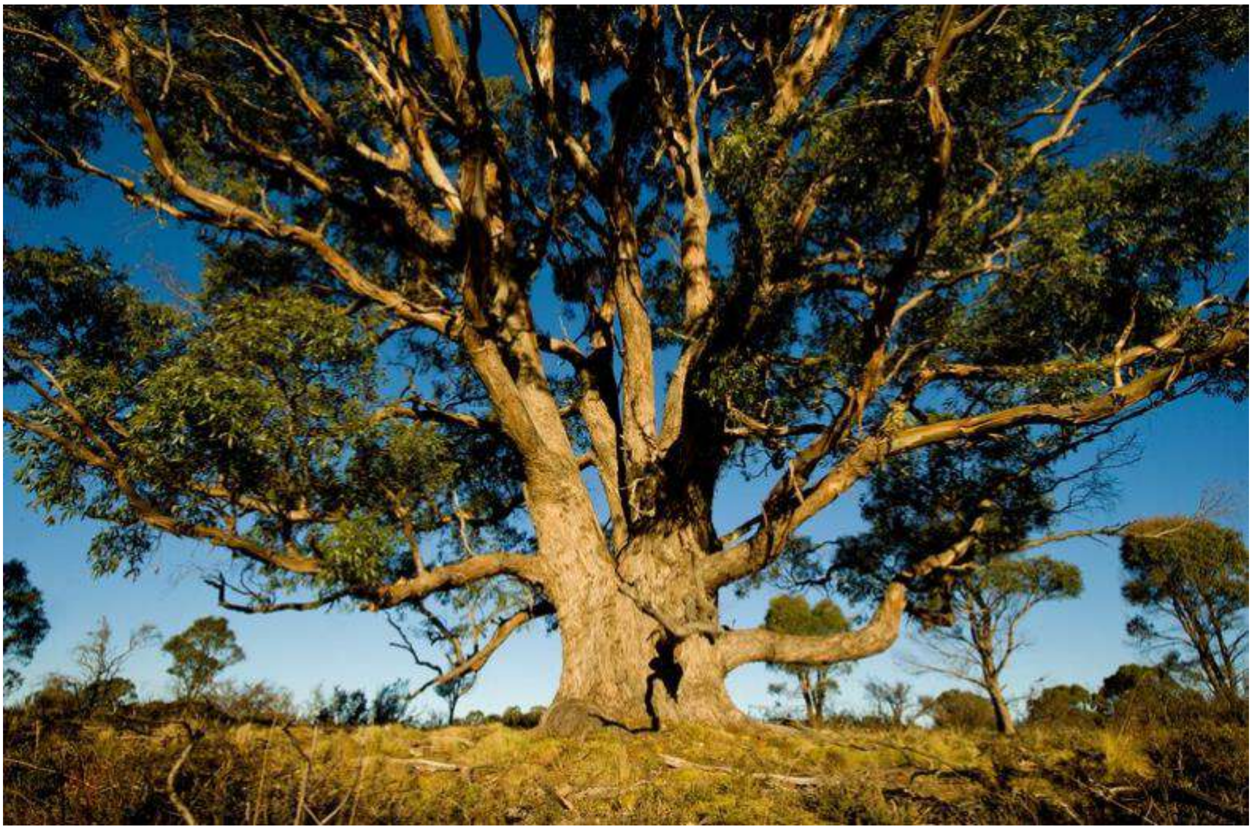

Figure 2- 2: Photograph to show Eucalyptus trees that used as indication of gold deposits

(Source: Photograph by Bill Hatcher, National Geographic).

- **Geochemical Prospecting**

Geochemical methods involve the measurement of the chemistry of the rock, soil, stream sediments or plants to determine abnormal chemical patterns which may point to areas of mineralization (Sittig, M., 1980). In this method, the most commonly chemical property calculated is the content of a main trace element. The zones in rocks and soil of anomalous or comparatively high concentrations of specific elements might lead prospectors to the elements in soils and rocks which comprise a geochemical anomaly (Chen, D., et al., 2019). When a mineral deposit forms, the concentration of the ore "metals" and a number of other elements in the surrounding rocks is usually higher than normal. These patterns are known as primary chemical halos. When a mineral deposit is exposed to surface processes, such as weathering and erosion, these elements become further distributed in the soil, groundwater, stream sediments or plants and this pattern is called a secondary chemical halo. Secondary halos aid in the search for deposits as they normally cover a greater area and therefore the chance of a chemical survey selecting a sample from these areas is greater than from a primary halo area. Different elements have different "mobility" in the environment based on their readiness to dissolve in water, their density, their ability to form compounds with other elements and the acidity (pH) of the environment. Subsequently, the secondary halo may not contain the "metal" for which a geochemical survey is searching but other "marker" elements (US Geological survey, 1980).

- **Geophysical Prospecting**

This method mixes the sciences of geology and physics in an effort to help prospectors in exploring both energy fuel and mineral deposits. There are numerous examples of geophysical prospecting such as geo-magnetic surveys for looking for gold deposits, magnetic surveys for looking for iron deposits as well as the utilization of scintillation counters for sensing radioactive uranium deposits (Yang et al., 2019). It should be noted that there are five main geophysical prospecting methods which include geo-electric, seismic, radiometric, gravimetric, and magnetic which are normally utilized in mineral exploration. Some of these techniques' applications need costly and complex tools as well as sophisticated processing techniques while others have been argued to be inexpensive and relatively simple.

2.3.3. Gold Mineralization in Tanzania

Tanzania is the 4th largest gold producer in Africa after South Africa, Ghana, and Mali and accounts for 1.3% of the total global gold production. Gold is mainly found in the lake Victoria region(which hosts Bulyanhulu, Geita, Buzwagi, and North Mara Mines), the Lupa Gold Field (which hosts the Shanta Gold Mine in Songwe), Mpada Mineral Field and the East Tanzania Gold Region(Mkurumu and Magam-bazi areas in Handeni). Tanzania's gold reserves are estimated at about 45 million ounces and Tanzania's gold production in Tanzania stands at around 50t per year. Gold exploration is mostly centered mostly on the greenstone belts around Lake Victoria, where several large deposits have been discovered and are being developed. Tanzania Gold Production Tanzania's total gold production reached 55.6 tonnes in the financial year 2020/21, versus 53.7 tonnes in the financial year 2019/20, and 42 tonnes in the financial year 2018/19. Gold exploitation is done by both large and medium-scale miners as well as small-scale miners. (TanzaniaInvest, 2023)

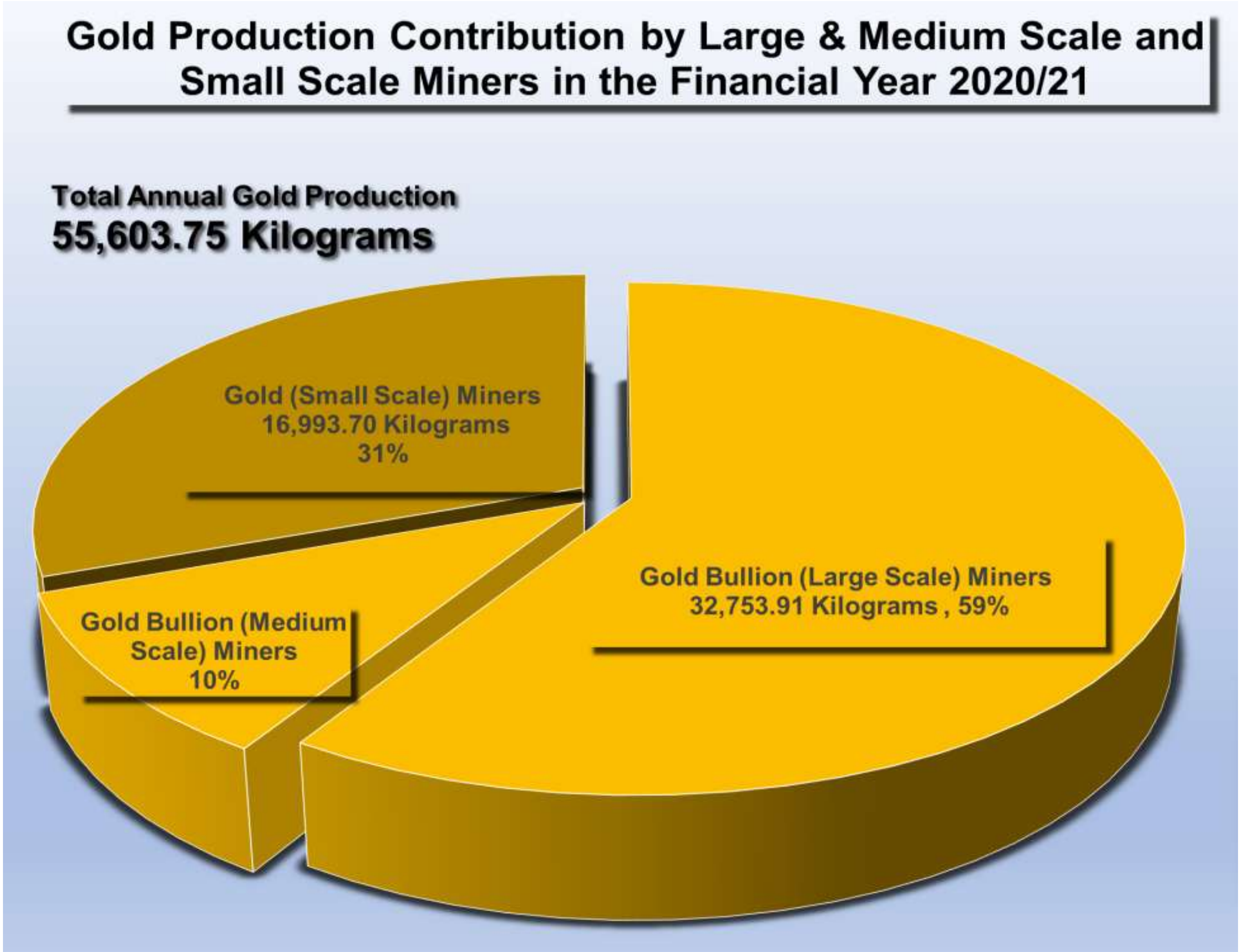

Figure 2-3: Gold Production Contribution

2.4. DATA COLLECTION AND SAMPLING

Collecting geochemical data involves the systematic sampling and analysis of geological materials to determine the composition and distribution of chemical elements. The data collection process for geochemical studies typically follows established protocols and guidelines to ensure accurate and reliable results. The key steps involved in data collection for geochemical studies:

1. Sampling Design: A well-designed sampling strategy is crucial to obtain representative and statistically valid geochemical data. The design should consider the specific objectives of the study, the nature of the geological materials being sampled, and the spatial distribution of the samples. Common sampling strategies include systematic sampling along transects or grids, random sampling, or stratified sampling based on geological units.
2. Sample Collection: Fieldwork involves collecting geological materials such as rocks, sediments, soils, water, or vegetation. Samples are usually collected using

appropriate sampling techniques, tools, and containers to prevent contamination and ensure sample integrity. The collection process should follow standard protocols and consider factors like sample size, sample location accuracy, and appropriate sample preservation methods.

3. Sample Preparation: Once collected, samples need to be prepared for laboratory analysis. This involves processes such as sample drying, crushing, grinding, and homogenization to obtain representative subsamples for analysis. The specific preparation methods depend on the nature of the samples and the analytical techniques to be employed.
4. Laboratory Analysis: Geochemical analysis is performed in specialized laboratories using various analytical techniques. Common methods include atomic absorption spectroscopy (AAS), inductively coupled plasma mass spectrometry (ICP-MS), X-ray fluorescence (XRF), or spectroscopy. These techniques determine the concentration and distribution of elements in the samples.
5. Quality Control: To ensure data accuracy and reliability, quality control measures are implemented throughout the data collection process. This includes the use of certified reference materials (CRMs) to validate analytical results, the inclusion of field and laboratory blanks, duplicate samples, and ongoing calibration and instrument checks.
6. Documentation: Accurate and comprehensive documentation of the data collection process is essential. This includes recording sample locations using GPS coordinates, describing the sampling environment, documenting sampling methods, and maintaining a clear chain of custody for the samples. Metadata, field notebooks, and detailed sample logs should be kept to ensure data traceability.

2.5 DATA STRUCTURE

In Geographic Information Systems (GIS), data structures refer to the specific ways in which geospatial data is organized and stored to enable efficient retrieval and manipulation. GIS data structures are designed to handle the unique characteristics of geographic data, which includes information about the Earth's surface, features, and their spatial relationships.

2.5.1. Structure of geochemical data

The data structure for geochemical data in GIS typically involves organizing the data in a tabular format, where each row represents a specific sample or measurement point, and columns represent different attributes or variables associated with the samples. This tabular structure allows for efficient storage, querying, and analysis of geochemical data.The following is an example of a typical structure for organizing geochemical data:To store and manage geochemical data, various file formats and databases can be used. Common file formats include CSV (Comma-Separated Values), Excel spreadsheets, or specialized formats such as Geochemical Data Exchange (GDX) files. Additionally, geochemical data can be stored and managed in relational databases, such as PostgreSQL with the PostGIS extension, or spatial databases like Esri's File Geodatabase or SQLite with SpatiaLite.

2.6. CORRELATION

Correlation refers to a bivariate analysis that measures the strength of association between two variables and the direction of the relationship. The correlation method is the most common method to use for numerical variables; it assigns a value between -1 and 1, where 0 is no correlation, 1 is total positive correlation, and -1 is total negative correlation (Boslaugh and Paul, 2008). This is interpreted as follows: a correlation value of 0.7 between two variables would indicate that a significant and positive relationship exists between the two. A positive correlation signifies that if variable A goes up, then B will also go up, whereas if the value of the correlation is negative, then if A increases, B decreases. Usually, in statistics, there are four types of correlations: Pearson correlation, Kendall rank correlation, Spearman correlation, and the Point-Biserial correlation

2.6.1. The Pearson Correlation

The Pearson correlation coefficient, also known as Pearson's r, is a statistical measure that quantifies the strength and direction of the linear relationship between two quantitative variables. It assesses how closely the data points cluster around a straight line, indicating the degree of linear association between the variables.

The coefficient ranges from -1 to 1, where a value of 1 represents a perfect positive linear relationship, 0 indicates no linear relationship, and -1 represents a perfect negative linear relationship. The sign of the coefficient (+ or -) indicates the direction of the relationship, while the magnitude (closer to 0 or 1) represents the strength of the relationship.

A positive correlation coefficient implies that as one variable increases, the other variable tends to increase as well. In contrast, a negative correlation coefficient suggests that as one variable increases, the other variable tends to decrease. A coefficient of 0 indicates no linear relationship between the variables.

2.6.2. Kendall rank correlation

The Kendall rank correlation coefficient, denoted as τ (tau), is a non-parametric measure of the strength and direction of the rank-based association between two variables. It is an alternative to the parametric Pearson correlation coefficient when the assumptions of the Pearson's correlation test are not met, such as when the data is not normally distributed or when there are outliers.

The Kendall correlation is particularly useful when working with small sample sizes and datasets that contain tied ranks, which can occur when multiple observations have the same value. It evaluates the similarity or dissimilarity in the rankings of the paired observations.

The formula to calculate Kendall's rank correlation coefficient is as follows:

$\tau = (nc - nd) / [0.5 * n * (n - 1)]$

Where:

- τ represents Kendall's rank correlation coefficient.
- nc is the number of concordant pairs, which are pairs of observations that have the same order or rank in both variables.
- nd is the number of discordant pairs, which are pairs of observations that have different orders or ranks in the two variables.
- n is the number of observations.

The resulting value of τ ranges from -1 to 1. A value of 1 indicates a perfect positive rank correlation, -1 represents a perfect negative rank correlation, and 0 suggests no rank correlation between the variables.

2.6.3 Spearman's correlation

Spearman's correlation coefficient, denoted as ρ (rho), is a non-parametric measure of the strength and direction of the monotonic relationship between two variables. It is derived by calculating the Pearson correlation coefficient on the ranked values of the data.

Spearman's correlation coefficient ranges between -1 and 1, similar to Pearson's correlation. A value of 1 indicates a perfect positive monotonic relationship, -1 represents a perfect negative monotonic relationship, and 0 suggests no monotonic relationship between the variables.

The formula to calculate Spearman's correlation coefficient is as follows:

$\rho = 1 - (6 * \Sigma(di^2)) / (n * (n^2 - 1))$

Where:

- ρ represents Spearman's rank correlation coefficient.
- di is the difference in ranks between corresponding pairs of observations in the two variables.
- n is the number of values in each dataset.

The numerator of the formula calculates the sum of squared differences in ranks, while the denominator adjusts for the number of observations and the ties in the data.

The resulting ρ value provides information about the strength and direction of the monotonic relationship between the variables. The closer ρ is to 1 or -1, the stronger the monotonic relationship. A positive ρ indicates that as one variable increases, the other tends to increase, while a negative ρ suggests that as one variable increases, the other tends to decrease.

Spearman's correlation is useful when analyzing data that do not follow a linear relationship or when dealing with ordinal or ranked data. It allows for the assessment of monotonic associations between variables without assuming linearity.

2.6.4 Point-Biserial Correlation

Point-biserial correlation is a statistical measure used to assess the strength and direction of the relationship between two variables, where one variable is binary (categorical with two levels) and the other variable is continuous. It is specifically designed to evaluate the association between a dichotomous variable and a quantitative variable.

The point-biserial correlation coefficient, denoted as rpb, ranges between -1 and 1. Similar to other correlation coefficients, a value of 1 indicates a perfect positive relationship, -1 represents a perfect negative relationship, and 0 suggests no relationship between the variables.

The point-biserial correlation coefficient is calculated using the following formula:

rpb = (M1 - M0) / (SD * sqrt(p * (1 - p)))

Where:

- rpb represents the point-biserial correlation coefficient.
- M1 is the mean of the continuous variable for the group with the binary variable at level 1.
- M0 is the mean of the continuous variable for the group with the binary variable at level 0.

- SD is the standard deviation of the continuous variable.
- p is the proportion of cases in the binary variable that have the value of 1.

The point-biserial correlation coefficient measures the degree of association between the binary variable and the continuous variable. A positive rpb indicates that higher values of the continuous variable tend to be associated with the presence of the binary variable at level 1. Conversely, a negative rpb suggests that higher values of the continuous variable tend to be associated with the absence of the binary variable at level 1.

2.7 MULTIVARIATE STATISTICS

Multivariate statistics refers to the branch of statistics that deals with the analysis of data sets involving multiple variables simultaneously(Johnson, 2007) . It focuses on understanding the relationships and patterns among multiple variables rather than analyzing them individually. Multivariate statistical methods allow for a more comprehensive and holistic analysis of data, taking into account the complex interdependencies that may exist between variables (Tabachnick, 2013).In multivariate statistics, the dataset consists of observations or cases, where each case is described by multiple variables. These variables can be either continuous (e.g., height, weight) or categorical (e.g., gender, occupation), and they can be interrelated in various ways (Rencher, 2003). Some commonly used techniques in multivariate statistics include; Multivariate regression, Principal Component Analysis (PCA), Factor Analysis (FA), Cluster Analysis (CA), Discriminant Analysis, Canonical Correlation Analysis, Multivariate Analysis of Variance (MANOVA)

2.7.1 Cluster Analysis

Cluster analysis is a statistical technique used to identify natural groupings or clusters within a dataset based on the similarity or dissimilarity of the elemental compositions or isotopic ratios of samples. Cluster analysis aims to group samples that exhibit similar geochemical characteristics and separate them from samples that have distinct compositions (Zuo et. Al, 2018).

Cluster analysis algorithms calculate the distances or similarities between pairs of samples and use these measures to create clusters. The choice of distance measure and clustering algorithm depends on the nature of the data and the objectives of the analysis. Commonly used distance measures in geochemical cluster analysis include Euclidean distance, Manhattan distance, and Mahalanobis distance (Liu et. al, 2020).

Once the clusters are formed, various visualization techniques can be employed to display the results. These may include dendrograms, which represent the hierarchical structure of the clusters, or scatter plots that show the samples colored or labeled according to their cluster assignments. Cluster analysis in geochemistry can provide insights into the compositional variability of different geological units, identify geochemical anomalies, delineate regions with similar elemental signatures, or assist in sample classification and discrimination. It can aid in the interpretation of geological processes, source identification, environmental assessments, and mineral exploration (Duffield, 2007).

2.7.2 Principal Component Analysis (PCA)

The Principal Components Analysis (PCA) is a technique used to enhance and separate certain spectral signatures from the background (Moradi, 2015; Bentahar et al., 2020). PCA is a multivariate statistical technique used to reduce the data redundancy by transforming the original data onto new orthogonal principal component axes producing an uncorrelated image, which has much higher contrast than the original bands.

Principal Component Analysis (PCA) can also be applied to geochemical data to enhance and separate certain spectral signatures or elemental patterns from the background. In this context, geochemical data refers to measurements of elemental concentrations or isotopic compositions in samples collected from geological materials such as rocks, soils, or sediments.

When applying PCA to geochemical data, the goal is to identify the dominant patterns or sources of variation in the dataset and to reduce the dimensionality of the data by transforming it into a new set of uncorrelated variables called principal components. These principal components represent linear combinations of the original variables that capture the most significant variability in the data.

The first principal component (PC1) represents the direction of maximum variability in the dataset and often corresponds to a common factor or dominant source of variation in the geochemical composition. It highlights features that are consistent across different elemental concentrations or isotopic ratios. PC1 can be interpreted as a measure of overall geochemical composition or as a proxy for a specific geological process or source. The second principal component (PC2) is orthogonal to PC1 and captures the second most significant source of variation that is independent of PC1. It highlights the spectral differences or elemental contrasts between different samples or regions. PC2 can reveal variations related to secondary processes, geological boundaries, or specific elemental associations. Similarly, the third principal component (PC3) captures additional variability that is orthogonal to PC1 and PC2. It may represent variations associated with specific geochemical processes, environmental factors, or interactions between different elements (Morad, 2015).

The other principal components beyond PC3 typically have progressively less variability and may capture more noise or minor sources of variation in the data.The magnitude and sign of the loadings (eigenvalues) of each principal component define the contribution of each original variable (element or isotope) to the overall variability captured by that component. By examining these loadings, it is possible to determine which elements or isotopes are most influential in driving the observed variations. CA can be applied in a standard manner, where all variables are included, or in a selective manner, where only a subset of variables relevant to a specific research question or geochemical process is considered.

2.7.3. Factor Analysis

factor analysis is a statistical technique used to identify underlying factors or latent variables that explain the correlation patterns among a set of observed geochemical variables. It aims to reduce the dimensionality of the data by grouping variables that are highly correlated and represent similar underlying processes or sources (Carranza, 2008). Factor analysis assumes that observed variables are influenced by a smaller number of latent factors. The technique estimates factor loadings, which indicate the strength and direction of the relationship between each variable and each factor. These factor loadings provide insights into which variables are most strongly associated with each factor(Carranza et. al, 2011).

Factor analysis help identify geochemical processes or sources that contribute to the observed variations in the dataset. For example, in geochemical studies, factors such as weathering, anthropogenic pollution, hydrothermal activity, or specific lithological sources can be extracted through factor analysis. These factors can provide valuable information for understanding geological processes, assessing environmental impacts, and exploring mineral resources. Factor analysis can also assist in data reduction and simplification, as it allows for the representation of a large number of variables with a smaller number of factors. This can aid in visualization, interpretation, and subsequent analysis of the dataset.

2.8. LINEAMENTS EXTRACTION

Lineaments are linear or curvilinear discontinuities in direct connection with the faults and composite fractures which are associated with geomorphological features and/or a various tectonic structure such as faults, fractures, fold axes and lithological contacts (Raghavan et al., 1995). In geology and hydrology applications, linear features on a satellite image regularly reflect the geological lineaments such as faults or fractures and hydrological structures such as river or shoreline (Lillesand et al., 2004). Among the lineament detection algorithms available in remote sensing, thosebased on filtering techniques using directional filters (Sobel) show good results (Masoumi et al., 2017). There are two methods for extracting lineaments using remote sensing techniques which are (i) manual and (ii) automated lineament extraction. Both methods depend on the data resolution and digital image processing (Raghavan et al, 1995).

2.8.1. Manual Lineament Extraction

Manual lineament extraction is a technique used to identify and extract lineaments from remote sensing data through visual interpretation. In this method, the knowledge and experience of the user play a crucial role in identifying and marking the lineaments present in the imagery. Lineaments are linear features or patterns observed in satellite images that can appear as straight lines or edges due to tonal variations. The manual extraction technique relies on the visual interpretation of these tonal variations to identify potential lineaments. Manual technique is composed of spatial convolution filtering also known as multi-band analysis. It involves applying different image enhancement techniques such as filtering operations, principal component analysis (PCA), band colour combination and spectral rationing (Sarp, 2005).

- **Filtering Operations**

Filtering operations are image processing techniques used to enhance or modify specific spatial frequencies in an image. In the context of lineament extraction from satellite images, filtering operations can be applied to sharpen the boundaries between different units or features present in the imagery. Satellite images exhibit spatial frequency, which refers to the rate of change in brightness values per unit distance within the image. Low frequency corresponds to areas with minimal changes in brightness values over a given area, while high frequency indicates significant and abrupt changes in brightness values over a short distance (Jensen, 1996).

Filtering operations are employed to accentuate or de-emphasize these spatial frequencies in an image. The operations involve the use of a moving window, also known as a kernel, with a specific size such as 3x3, 5x5, or 7x7 pixels. The window is systematically moved over the input image, and for each pixel within the window, a new numerical value is calculated based on the values of the surrounding pixels. This calculated value is then assigned to the central pixel of the window in the resultant image.

- **Band Color Combination**

Band color combination is a technique used to enhance an image by assigning specific colors to digital number (DN) values, thereby increasing the contrast of those values within the surrounding pixels (Jensen, 1996). This technique aims to improve the visual interpretation of the image by leveraging the human eye's ability to perceive a wider range of colors compared to shades of gray. In band color combination, the DN values of an image are mapped to specific colors. This can be done by assigning different color channels (such as red, green, and blue) to different bands of the image. By combining these color channels, either through additive or subtractive color models, a color image is created. The conversion of an entire grayscale image to a color image allows for a more visually appealing representation, as different features or DN values can be represented by distinct colors. Alternatively, specific portions of an image that correspond to certain DN values can be selectively colored, emphasizing the areas of interest (Raghavan et al., 1995).

- **Spectral rationing**

This is the technique that involves dividing band by band that helps to remove shadowing. Rationed images are useful for discriminating spectral variations in an image with brightness variations (Sarp, 2005). The enhanced discrimination is due to the fact that rationed images clearly display the variations in slopes of spectral reflectance curves between the two bands involved regarding reflectance values observed in the bands (Lillesand & Kiefer, 2000). By rationing the data from two different spectral bands the variations in the slopes of the spectral reflectance are enhanced. The variation in scene illumination as the result of topographic effect are reduced. Ratio images combined in RGB offer greater contrast between units in the image through individual TM band false colour images. By help of the band rationing most of the scene illumination effects are removed from the image and linear features are more easily identified from the rationed image (Jensen, 1996).

2.8.2. Automatic Lineament Extraction

Automatic lineament extraction involves the use of algorithms and software to extract lineaments from remote sensing data without the need for manual interpretation. Several algorithms are commonly employed for this purpose, including the Hough transform, Haar transform, and segment tracing algorithm.

The Hough transform is a widely used technique for detecting specific shapes, such as lines, circles, and ellipses, within an image. It requires the shape to be defined in a parametric form. The Hough transform is advantageous in that it is relatively robust to gaps in lines and noise, making it suitable for lineament extraction (Wang et al., 1990).

The Haar transform, another algorithm used for lineament extraction, focuses on the extraction of linear and anomalous patterns in an image. It captures differential energy concentrated in local regions by decomposing the image into low and high frequency components. The Haar transform can also be utilized for image enhancement (Kocal, 2004).

The Segment Tracing Algorithm (STA) is a method developed by Koike et al. (1995) for automatically detecting lines of pixels as vector elements. It examines the local variance of the gray level in a digital image to trace segments that represent lineaments.

These automatic extraction algorithms provide efficient and objective approaches to identifying and extracting lineaments from remote sensing data, reducing the reliance on manual interpretation and leveraging the computational capabilities of software tools.

2.9 VALIDATION

Cross-validation is a statistical method used to estimate the performance (or accuracy) of machine learning models (Kutner et al., 2005). It is used to protect against overfitting in a predictive model, particularly in a case where the amount of data may be limited. In cross-validation, you make a fixed number of folds (or partitions) of the data, run the analysis on each fold, and then average the overall error estimate. cross-validation methods are divided into non-exhaustive and exhaustive whereby non-exhaustive do not compute all ways of splitting the original data such as holdout and k-fold cross validation. Exhaustive is a cross validation method which test on all possible ways to divide the original sample into a training and a validation set, such as leave one out cross validation (LOOCV).

CHAPTER THREE

METHODOLOGY

3.0. OVERVIEW

This chapter describes overall workflow ranging from data collection to methods used. It includes the data collection, data pre-processing and data analysis methods that were used in obtaining the results. The methodological workflow is as shown in figure 3-1

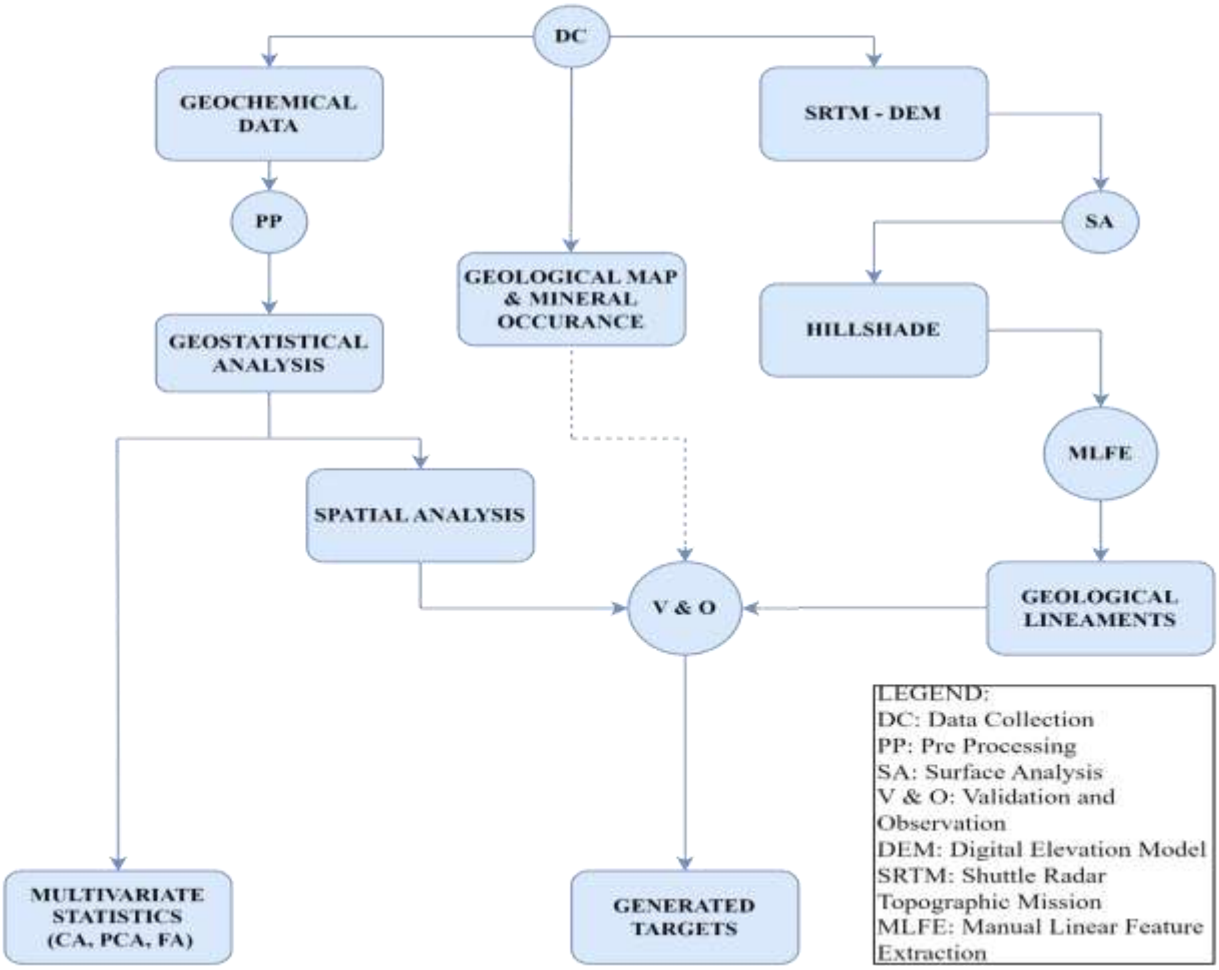

Figure 3-1: Workflow Methodology

3.1. DATA AND THEIR SOURCES

The data used in this study were collected from various sources and in different format, and are summarized as:

Table 3-1: Datasets and their sources

DATASETS	SPATIAL RESOLUTION	DATA FORMAT	DATA TYPE	SOURCE
SRTM-DEM	30m	TIFF	RASTER	USGS
Geochemical Data		Shapefile	VECTOR	Shanta Mining Company
Geological Data		Shapefile	RASTER	Shanta Mining Company
Mineral Occurrences		Shapefile	VECTOR	Shanta Mining Company
Administrative boundaries		Shapefile	VECTOR	GADM

3.2. SOFTWARE USED

- ArcGIS: a geographic information system (GIS) software that allows for the visualization, analysis, and management of geographic data.
- Geostatistical Analyst: an extension for ArcGIS that provides tools for spatial data analysis, including interpolation and geostatistics.
- R: a programming language and environment for statistical computing and graphics, which has a variety of packages for geostatistics and spatial analysis.

3.3. DATA COLLECTION

The study involves collection of required data from web archives and from Government Agencies. The dataset includes Geochemical data, Geological data. Collecting geochemical data involves the systematic sampling and analysis of geological materials to determine the composition and distribution of chemical elements. The data collection process for geochemical studies typically follows established protocols and guidelines to ensure accurate and reliable results The geochemical data were in excel format in which it consists of unique columns with unit of measurements. The geochemical data used for this study, elements were measured in ppm.

3.4. PRE-PROCESSING TECHNIQUES

Data preprocessing is an essential step in handling and preparing geochemical data for analysis. It involves various techniques and methods to clean, transform, and normalize the data, ensuring its quality and compatibility for further analysis.

3.4.1. Missing Data Handling

Missing data can occur in geochemical datasets due to various reasons such as data collection errors or incomplete sampling. A strategy used for handling missing data in the study is in the deletion of incomplete records where by the rows which contain missing information of some chemical elements were deleted.

3.4.2. Data Transformation

Geochemical datasets often exhibit skewed distributions or heteroscedasticity, which can violate the assumptions of statistical analyses. The study uses Logarithmic Data transformation technique to normalize the data and improve statistical validity.

3.5. PROCESSING TECHNIQUES

3.5.1. Descriptive statistics

In the descriptive statistics analysis mean, median, standard deviation, quartile, minimum, maximum, scatter plots and box plots of the raw data were generated using R software. The purpose of these values was to gain a comprehensive understanding of the data's central tendency, variability, spread, and extreme values. These insights are valuable in generating targets, as they provide information on the acceptable range and permissible deviations. It allows for informed decision-making and helps ensure that the generated targets align with the characteristics and patterns observed in the geochemical data. The data are displayed as a collection of points using Cartesian coordinates. The value of the variable is determined by the position of each point with respect to its horizontal and vertical axis.

Table 3-2: Descriptive Statistics of the raw data

Variables	MEAN	STD	Q1	MEDIAN	Q3	MIN	MAX
Au	0.028	0.120	0.002	0.003	0.01	0.0005	4.688
Ag	0.016	0.033	0	0.012	0.022	0	1.763
Al	398.225	1713.001	0	0	0	0	10350.000
As	6.680	65.385	0	0.618	1.987	0	3720.000
Ba	38.053	166.779	0	0	0	0	1250.000
Bi	0.014	0.058	0	0.001	0.004	0	2.000
Ca	26.582	129.773	0	0	0	0	1800.000
Ce	4.534	20.050	0	0	0	0	176.000
Cr	5.393	24.422	0	0	0	0	237.000
Cs	0.162	0.733	0	0	0	0	7.310
Cu	4.536	7.248	0	3.4	5.47	0	67.500
Fe	0.191	0.842	0	0	0	0	8.260
Hf	0.256	1.122	0	0	0	0	10.100
K	88.133	381.490	0	0	0	0	2860.000
La	2.057	9.200	0	0	0	0	96.700
Li	1.481	6.575	0	0	0	0	84.400
Mg	39.282	184.622	0	0	0	0	1820.000
Mn	33.205	150.269	0	0	0	0	2010.000
Mo	0.195	0.473	0	0.106	0.193	0	25.205
Na	67.657	307.493	0	0	0	0	5620.000
Nb	1.371	7.161	0	0	0	0	112.500
Ni	2.930	13.434	0	0	0	0	119.000
P	22.964	112.152	0	0	0	0	1980.000
Pb	1.205	5.550	0	0	0	0	73.000
Rb	3.761	16.279	0	0	0	0	104.000
S	0.850	3.981	0	0	0	0	70.000
Sb	0.106	0.479	0	0.01	0.02	0	9.930
Sc	0.595	2.652	0	0	0	0	25.200
Sr	10.476	48.256	0	0	0	0	763.000
Ta	0.080	0.406	0	0	0	0	6.000
Te	0.004	0.008	0	0.001	0.004	0	0.090
Th	0.557	2.431	0	0	0	0	20.500
Ti	18.089	78.670	0	0	0	0	761.000
V	4.237	18.846	0	0	0	0	182.000
W	0.089	0.291	0	0.017	0.034	0	3.600
Zn	5.253	16.828	0	0.805	2.167	0	156.000
Zr	9.887	43.771	0	0	0	0	416.000
Be	0.097	0.427	0	0	0	0	3.830
Cd	0.015	0.027	0	0.007	0.015	0	0.400
Co	0.959	4.321	0	0	0	0	43.700
Ga	1.017	4.386	0	0	0	0	28.300
Ge	0.010	0.043	0	0	0	0	0.360
In	0.003	0.013	0	0	0	0	0.139
Se	0.148	0.200	0	0.117	0.198	0	2.000
Sn	0.077	0.342	0	0	0	0	3.100
Tl	0.030	0.101	0	0.004	0.01	0	0.740
U	0.131	0.574	0	0	0	0	6.500
Y	1.010	4.461	0	0	0	0	42.800

3.6.1.1. Boxplot

The primary purpose of using box plots in data analysis is to identify and understand the type of relationship between variables. This graphical representation offers valuable insights by examining various aspects of the data distribution.

One of the key observations made from box plots is the comparison of medians. By comparing the medians of different variables from the dataset, make it possible to assess differences in their central tendency. A higher median in one variable was compared to another variable to suggests a potential relationship or variation between the variables being studied.

Another important aspect revealed by box plots is the spread of the data. The width of the box in a box plot provides information about the variability or dispersion of the data. Variables with wider boxes indicate more dispersed data, implying greater variability. On the other hand, narrower boxes suggest less variability. Analyzing the spread of different elements in geochemical data through box plots help identify variations in their distribution, such as knowing which variable can be indicative of specific relationships between the other variables.

Box plots are also effective in detecting outliers, which are extreme values that deviate significantly from the majority of the data. These outliers can indicate unusual behavior or potential errors in measurements. Identifying outliers is particularly crucial when analyzing geochemical data with multiple elements, as it allows for the identification of abnormal values or potential anomalies that could affect the interpretation of relationships between variables.

Furthermore, box plots was used provide insights into the shape of the distribution. When the median is positioned closer to one end of the box, it suggests asymmetry in the distribution. Such skewed distributions can indicate a non-linear relationship between variables or the presence of different patterns within the data. Analyzing the shape of distributions using box plots can further contribute to understanding the nature of relationships between variables and uncovering potential complexities in the data.

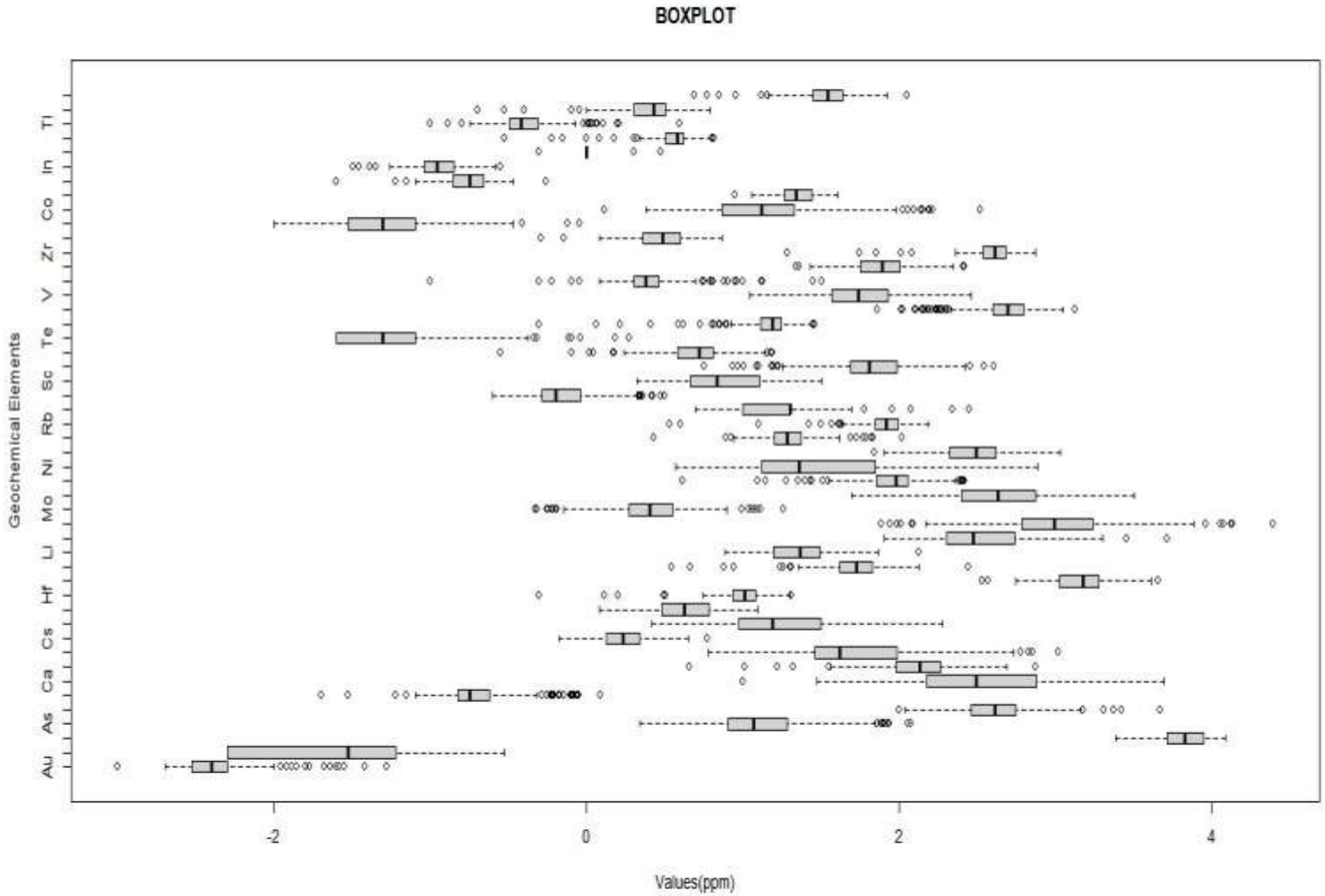

Figure 3-1: Boxplot of Geochemical Elements

3.5.2. Multivariate Statistics

Multivariate statistics such as cluster analysis, principal components and factor analysis are used to analyze and present the results. The data were transformed by normal logarithmic transformation. The R statistical software was used to select the number of components and the type of matrix (e.g. correlation, covariance) graphs to be plotted (e.g. score plot, scree plot, loadings plot) of the transformed data.

Cluster Analysis

The multivariate cluster analysis conducted on the 48 elements (variables) from 7001 samples aimed to measure the similarity or dissimilarity of the trace elements and gold concentration. The analysis utilized a process that involved the calculation of distances between initial clusters, which in this case, represent the individual variables. The distances were calculated based on a measure of dissimilarity, such as Euclidean distance or another appropriate distance metric. In this case, the distances between clusters were determined using the Pearson correlation coefficient. The Pearson correlation coefficient measures the linear relationship between two variables, ranging from -1 to 1. A value of -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship.

The next step in the process was the fusion of the two most similar clusters, meaning the two clusters with the smallest distance between them. The fusion process involved merging the two most similar clusters, which had the highest correlation coefficient, indicating a stronger linear relationship between the variables they represented. After the fusion, the distances between the newly formed cluster and the remaining clusters were recalculated based on the Pearson correlation coefficient.

This iterative process of merging similar clusters and recalculating distances based on the Pearson correlation coefficient continued until all variables eventually fell into one cluster. The stopping criterion for the iteration could be reaching a desired number of clusters or when all variables were merged into a single cluster.

By using the Pearson correlation coefficient as the inter-cluster linkage distance, the multivariate cluster analysis aimed to identify groups of variables with similar linear relationships or patterns. This analysis helps in understanding the associations between the trace elements and gold concentration, allowing for the detection of variables that tend to co-vary or exhibit similar behavior.

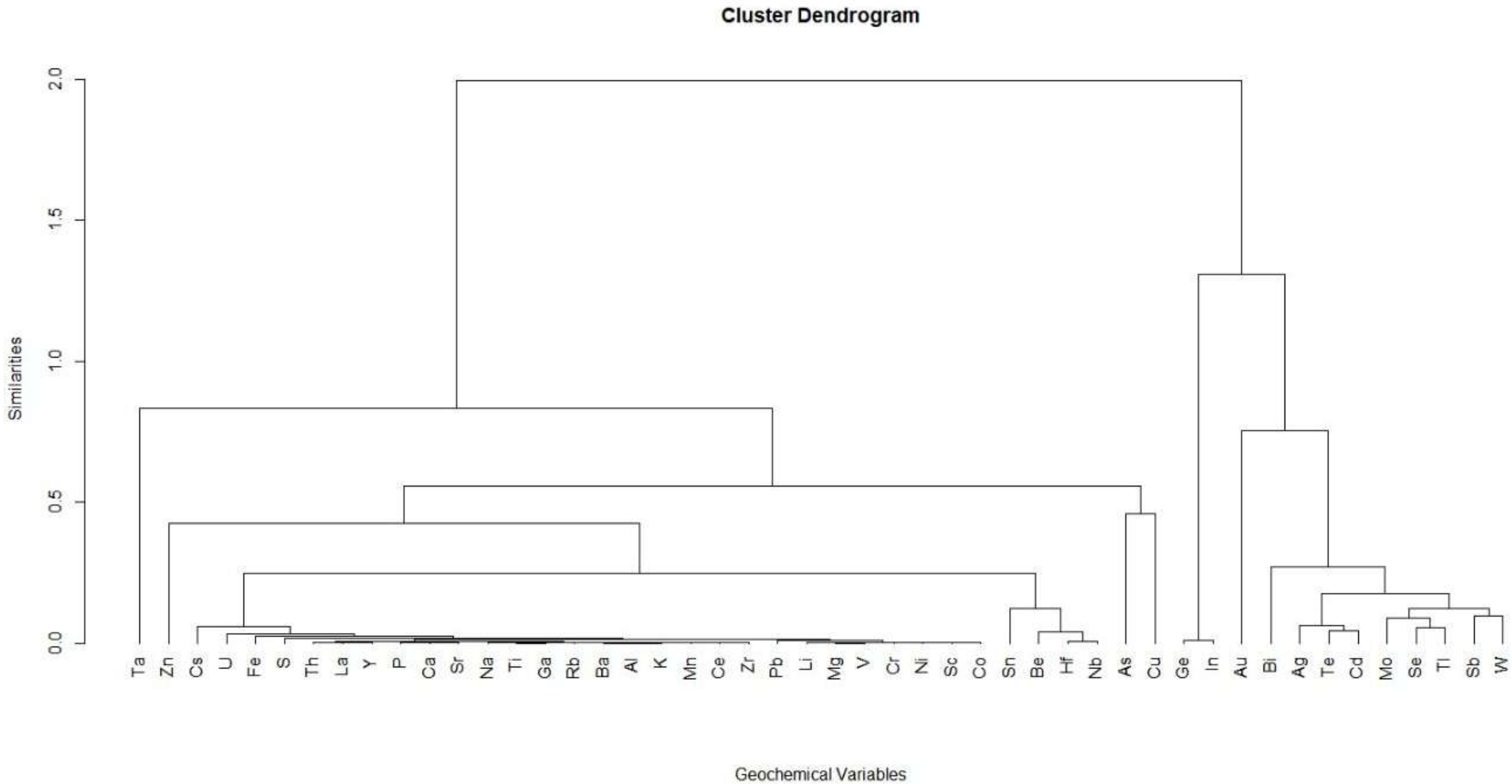

Figure 3-2: Cluster Dendrogram of Geochemical variables

Principal Component Analysis

During Principal Component Analysis (PCA), eigenvalues play a crucial role in data reduction. The eigenvalues of the correlation matrix represent the variances extracted by the principal components. Each eigenvalue corresponds to a principal component and signifies the amount of variance explained by that component.

The correlation matrix was computed so as to assess the linear relationships between the variables. The correlation matrix represents the pairwise correlations between all pairs of variables, resulting in a square matrix with dimensions equal to the number of variables (48 variables in this case).

Let's assume a dataset with n variables (48 in this case) and m samples (7001 in this case).

- Calculating the mean of each variable, denoted by μ_j, where j ranges from 1 to n.
- Calculating the standard deviation of each variable, denoted by σ_j, where j ranges from 1 to n.
- Computing the correlation coefficient (Pearson correlation) between variables i and j using the formula:
    - $corr(i, j) = \sum((x_i - \mu_i) * (x_j - \mu_j)) / (\sigma_i * \sigma_j)$
    - where x_i and x_j are the values of variables i and j, respectively.
    - Repeating the calculation for all pairs of variables to form an n x n correlation matrix

In performing the eigen decomposition (also known as eigenvalue decomposition) on the correlation matrix involves calculating the eigenvalues and eigenvectors of the correlation matrix. The eigenvalues represent the amount of variance explained by each principal component. They provide information about the importance of each component in capturing the variability in the data. The eigenvectors correspond to the principal components. Each eigenvector represents a linear combination of the original variables and describes a direction in the multidimensional space of the data.

Once the correlation matrix has been computed, eigen decomposition can be performed to obtain eigen values and eigen vectors.

Let A be the correlation matrix of dimension n x n. Computing the eigenvalues (λ_i) and eigenvectors (v_i) of A using the eigenvalue decomposition method. The eigenvalue decomposition of a square matrix A is represented by:

$A = V * D * V^{(-1)}$

where:

A is the square matrix of interest.

V is a matrix whose columns are the eigenvectors of A.

D is a diagonal matrix containing the eigenvalues of A.

The diagonal elements of D represent the eigenvalues of A, denoted as $\lambda_1, \lambda_2, ..., \lambda_n$. The corresponding columns of V represent the eigenvectors, denoted as $v_1, v_2, ..., v_n$.To calculate the eigenvalues and eigenvectors, the following formula is used:

$A * v = \lambda * v$

where:

A is the matrix for which you want to calculate the eigenvalues and eigenvectors.

v is the eigenvector.

λ is the eigenvalue associated with that eigenvector.

To find the eigenvalues, solving the characteristic equation:

$|A - \lambda I| = 0$

where:

I is the identity matrix of the same size as A.

Solving the characteristic equation will give you the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$.

Once the eigenvalues are computed, the corresponding eigenvectors can be computed by substituting each eigenvalue back into the equation $A * v = \lambda * v$ and solving for v. Each eigenvector v_i will correspond to the eigenvalue λ_i.

The eigenvalues (λ_i) represent the variances explained by each principal component.

The eigenvectors (v_i) correspond to the principal components and describe the directions in the multidimensional space of the data

3.6. LINEAMENTS EXTRACTION

The lineaments are the linear feature that could be mapped from the surfaces, and was morphological expression of geologic structure. Mapping geologic lineaments is important for mineral exploration because integrating geochemical data with lineament extraction allows to generate target areas for further exploration. By combining the spatial distribution of geochemical anomalies with the proximity to lineaments, help to prioritize areas with a higher potential for mineralization. This helps in focusing limited resources on areas that are more likely to host economically viable deposits, increasing the efficiency of exploration efforts. In this study lineaments were extracted by performing (i) hillshade analysis and (ii)manual linear feature extraction.

3.6.1. Hillshade Analysis

On the SRTM DEM data, to show the data of more representative images, it was needed to perform process of hillshade used Spatial Analyst tools on ArcGIS software. The hillshade process was performed on DEM data with 4 value variations of sun azimuth for 100°, 200°, and 315°; whereas for the input value of sun altitude were 60°, 50° and 45° respectively.

Hillshade=255.0 ×((cos $(Zenith\_rad)$×cos $(Slope\_rad)$)+(sin $(Zenith\_rad)$×sin $(Slope\_rad)$×cos $(Azimuth\_rad - Aspect\_rad)$))

3.6.2 Manual Linear Feature Extraction

After hill shade processing of the images, the lineaments were extracted using manual technique instead of automatic extraction using software applications such as PCI Geomatica. Manual extraction was preferred over automatic extraction because it was able to discriminate non-geological features such as roads, railways and buildings from the geological lineaments. The images as result of the hill shade process were digitized to get the linear features that was seen from all angles of azimuth and altitude.

3.7 TARGET GENERATION

To generate a target, the contour maps, gridded maps were plotted using the ArcGIS software platform where the geochemical data were used to create contour maps and gridded maps representing the distribution of gold and its associated elements across the study area. The kriging interpolation technique was used to generate continuous surfaces of element concentrations and then the contour intervals and grid resolution were adjusted to effectively visualize the patterns and variations in element concentrations.

Lineaments and anomalous trends were marked on gridded maps for several geochemical elements particularly gold and its associated elements that were identified through multivariate analysis. The maps containing the line segments were overlaid and the zones with a high concentration of gold and its associated elements were marked to generate targets for further investigation.

CHAPTER FOUR

RESULT, ANALYSIS AND DISCUSSION

4.0. OVERVIEW

A total of 48 elements were analyzed from 7001 ang 3500 regolith samples for data group 1 and 2 respectively, these were Au, Ag, Al, As, Ba, Bi, Ca, Ce, Cr, Cs, Cu, Fe, Hf, K, La, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Rb, S, Sb, Sc, Sr, Ta, Te, Th, Ti, V, W, Zn, Zr, Be, Cd, Co, Ga, Ge, In, Se, Sn, Tl, U, Y. To obtain clear data due diligence, a replicate sample and standard sample were used. The replicated samples indicated that their results were 96 ± 2% precisely except for As and Sb, which showed an average precision of 93 ± 2.4%. All the results of the control material were noted to be within the acceptable range.

4.1. DESCRIPTIVE STATISTICS

Descriptive statistics of the raw data obtained from 48 analyzed elements have shown a high kurtosis and skewness values deviated far from a normal distribution due to outliers. The parameters computed from the raw data were the mean, median, standard deviation, quartile, minimum and maximum values of each element. The raw results show that Al, As, Ba, Ca, Ce, Cr, Cu, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Rb, S, Sc, Sr, Th, Ti, V, Zn, Zr, Co, Ga, and Y have the highest maximum enrichment concentrations; while gold (Au) as a targeted indicator element showed minimum and maximum concentrations of (0.0005 and 4.688)ppm.

A box plot for the transformed data shows a narrow interquartile range (IQR) of boxes for Au, Th, Sr, La, Mg, Ce, Ti and Ta that is positively skewed. This implies that a greater number of samples have comparable concentrations. Pb, Zn, Mn, Ca, K, Y and Hf are positively skewed with wide IQR boxes, which suggests a great variability in the concentrations; while Cu, Zn, Ni, Co, Fe, As, V, P, Ba, Sn and Nb have wide ranges of concentration, showing a normal or near-normal distribution. A narrow range and a normal distribution are shown by Cr and W.

Table 4-1: Variables showing high values (ppm) of the raw data

Variables	values(ppm)	Variables	values(ppm)
Al	10350.000	Ni	119.000
As	3720.000	P	1980.000
Ba	1250.000	Pb	73.000
Ca	1800.000	Rb	104.000
Ce	176.000	S	70.000
Cr	237.000	Sc	25.200
Cu	67.500	Sr	763.000
K	2860.000	Th	20.500
La	96.700	Ti	761.000
Li	88.400	V	182.000
Mg	1820.000	Zn	156.000
Mn	2010.000	Zr	416.000
Mo	25.205	Co	43.700
Na	5620.000	Ga	28.300
Nb	112.500	Y	42.800

4.2. ELEMENTAL ASSOCIATION

To understand the relationship between elements and how these elements are associated with each other, elements were clustered into clustering phases. Based on the cluster analysis, it is observed that Au has three subclusters and the associated elements were suggested as pathfinders based on their level of similarity.

In all clustering phases, the cluster containing Au was considered for evaluation which are Bi, Ag, Te, Cd, Mo, Se, Tl, Sb, and W with a similarity level of 53.93% and distance level of 0.92 as shown in the dendrogram (figure 3-2). According to Korshunova et al. (2017), elements associated with gold mineralization tend to form a large secondary halo from the overlying soils. A study conducted on soils above the host rock of hypozonal deposit noted that Au was associated with Se, Ag, Te, Sb, Pb, Mo and Bi. As is used as a common pathfinder in gold exploration. However, in hypozonal deposit systems Au does not associate with As (most of the samples rich in Au have 0 ppm As), As is not considered to be a Au pathfinder. GST (1930), Groves et al. (2003), Groves (2010) and Kabete et al.

(2012a) reported that gold mineralization in a study area is typical of a hypozonal gold system. Similar hypozonal deposits in the Barberton greenstone belt (South Africa) of Paleoarchean–Neoarchean age show that Au deposits are associated with Ca, Se, Cu, Ni, Co and S (Kolb et al. 2015). The use of Se, Cu and Ni as pathfinders for gold mineralization in hypozonal gold systems will give a promising outcome.

4.3. PRINCIPAL COMPONENT ANALYSIS

During the principal component analysis (PCA), eigenvalues were generated for the 48 geochemical elements as an important tool in data reduction. The correlation matrix values (eigenvalues) are the variances extracted by the principal components The principal components as a linear combination of the original variables that account for the variance in the data signify that the maximum number of components extracted equals the number of variables. From the analyzed elements, nine main PCA components pulled out a cumulative percentage of 96% of the total variance by considering the eigenvalue of ≥0.5 (Table 4-2). The components are used in bedrock mapping, which in turn helps to infer the host rock of the gold mineralization.

PC1 has a large positive loading of all elements (except for Cu, Ge and In) including V, Cu, Ni, Fe, Mn, Cr and Co, which are related to the underlying hornblende gneiss interlayered with quartz (Fig. 1). PC1 constitutes 50.244% of the total variance. PC2 accounts for 39.985%, showing a large positive loading except for Au, Ag, Bi, Mo, Sb, Te, W, Cd, Ge, In, Se, and Tl which shows that negative loading is related to epitomized quartz rock and amphibole richmafic rocks or amphibolite/hornblende gneiss, and that negative loading is linked to the underlying felsic rocks such granites, granitic gneiss and quartz vein porphyry. PC3 constitutes 3.006%, with a negative loading from Au, Ag, As, Cu, Mo, Sb, Te, W, Zn, Cd, Ge, In, Se, and Tl and minor positive loading from K, S, U, Ag, Pb and Mn that is related to tonalite and tonalitic orthogneiss. It also has large negative loadings from Ni, As, Sb, Bi, W, Nb, Ta, Li and Rb (Table 4-3).

Table 4-2: Eigen analysis of the correlation Matrix

Principal Components	Eigenvalues	Proportion	Percentage	Cummulative
PC1	34.97401	0.7286251	72.86251	0.7286251
PC2	8.33291	0.1736026	17.36026	0.9022277
PC3	1.33115	0.02773223	2.773223	0.92995993
PC4	0.91839	0.0191331	1.91331	0.94909303
PC5	0.73385	0.01528855	1.528855	0.96438158
PC6	0.54224	0.01129665	1.129665	0.97567823
PC7	0.20734	0.00431967	0.431967	0.9799979
PC8	0.19526	0.004067976	0.4067976	0.984065876
PC9	0.14625	0.003046896	0.3046896	0.987112772
PC10	0.10339	0.002153881	0.2153881	0.989266653
PC11	0.08444	0.001759125	0.1759125	0.991025778
PC12	0.06954	0.001448809	0.1448809	0.992474587
PC13	0.05581	0.001162653	0.1162653	0.99363724
PC14	0.04717	0.000982796	0.09827962	0.994620036
PC15	0.04396	0.000915762	0.0915762	0.995535798
PC16	0.0359	0.000747995	0.0747995	0.996283793
PC17	0.03067	0.000638857	0.0638857	0.99692265
PC18	0.02502	0.000521262	0.05212617	0.997443912
PC19	0.0234	0.000487564	0.04875641	0.997931476
PC20	0.0218	0.000454119	0.04541194	0.998385595
PC21	0.01782	0.000371271	0.03712708	0.998756866
PC22	0.00162	0.00033733	0.03373301	0.999094196
PC23	0.00174	0.000199233	0.01992327	0.999293429
PC24	0.00551	0.00011485	0.01148496	0.999408279
PC25	0.00428	8.91803E-05	0.008918033	0.999497459
PC26	0.00379	7.89725E-05	0.007897249	0.999576431
PC27	0.00373	7.7748E-05	0.007774798	0.999654179
PC28	0.00282	5.87323E-05	0.005873225	0.999712912
PC29	0.00216	4.51924E-05	0.004519244	0.999758104
PC30	0.00202	4.21873E-05	0.004218731	0.999800291
PC31	0.00174	3.6374E-05	0.003637398	0.999836665
PC32	0.00142	2.9772E-05	0.002977195	0.999866437
PC33	0.00134	2.78951E-05	0.00278951	0.999894332
PC34	0.00109	2.26347E-05	0.002263466	0.999916967
PC35	0.00086	1.7928E-05	0.001792797	0.999934895
PC36	0.00065	1.36325E-05	0.001363248	0.999948528
PC37	0.00051	1.057E-05	0.001057003	0.999959098
PC38	0.00041	8.63602E-06	0.000863602	0.999967734
PC39	0.00033	6.80745E-06	0.000680745	0.999974541
PC40	0.00029	6.03312E-06	0.000603312	0.999980574
PC41	0.00023	4.72571E-06	0.000472571	0.9999853
PC42	0.00019	3.95669E-06	0.000395669	0.999989257
PC43	0.00017	3.61036E-06	0.000361036	0.999992867
PC44	0.00014	2.98744E-06	0.000298744	0.999995854
PC45	0.00008	1.71543E-06	0.000171543	0.99999757
PC46	0.00005	9.89799E-07	9.89799E-05	0.99999856
PC47	0.00004	8.65658E-07	8.65658E-05	0.999999425
PC48	0.00003	5.64538E-07	5.64538E-05	1

4.4. PEARSON'S CORRELATION

A correlation matrix from Pearson's correlation of the identified pathfinder elements with other trace elements were performed in order to infer the underlying geology of the regolith material from which samples were collected. From Table 4-4, it can be seen that Bi has a strong positive correlation with Ag (r = 0.815), Te (r = 0.829), Cd (r = 0.787), Mo (r = 0.728), Se (r = 0.758), Tl (r = 0.805), Sb (r = 0.810), and W (r = 0.770). Silver (Ag) shows a strong positive correlation with Te (r = 0.944), Cd (r = 0.937), Mo (r = 0.825), Se (r = 0.806), Tl (r = 0.884), Sb (r = 0.884), and W (r = 0.846); while Tellurium (Te) shows a stronger correlation with Cd (r = 0.909), Se (r = 0.901), Tl (r = 0.928), Sb (r = 0.894), and W (r = 0.879). Molybdenum (Mo) has demonstrated a strong correlation with Se (r = 0.909), Tl (r = 0.924) and Sb (r = 0.883). These elements are characteristics of sedimentary rocks, mafic volcanic and quartz vein (Carranza et al. 2013; Sadeghi et al. 2015). Gold (Au) as an indicator element shows a slightly positive correlation with the identified pathfinder elements such that $0.298 \leq r \leq 0.247$. The relationship between Au and its respective pathfinder elements indicates that gold mineralization is associated with either mafic volcanic rock/intrusions and quartz vein.

Table 4-3: Pearson correlation of pathfinder elements against other trace elements (only those showing positive correlation)

	Au	Bi	Ag	Te	Cd	Mo	Se	Tl	Sb	W
Bi	0.24704									
Ag	0.29002	0.8154								
Te	0.29802	0.82873	0.94379							
Cd	0.2846	0.78669	0.93739	0.95734						
Mo	0.26831	0.72841	0.82512	0.88626	0.90889					
Se	0.26727	0.75755	0.86103	0.91802	0.90136	0.90997				
Tl	0.27753	0.80529	0.88436	0.95113	0.92799	0.92426	0.94542			
Sb	0.2622	0.81037	0.88437	0.90337	0.89427	0.8827	0.90821	0.92633		
W	0.25208	0.77024	0.84611	0.87929	0.87927	0.88932	0.87619	0.90096	0.90187	
As	0.02007	0.02582	-0.0646	-0.0625	-0.0722	0.12494	0.08	0.06454	0.22773	0.1685
Ge	0.06249	-0.1226	0.0581	-0.0076	-0.0198	-0.2568	-0.1954	-0.1945	-0.3043	-0.2731
In	0.06278	-0.1236	0.05771	-0.0076	-0.0205	-0.2572	-0.1967	-0.1971	-0.3082	-0.2759

4.5. FACTOR ANALYSIS (FA)

The first three factors (Table 6) have shown that Al, Ba, Ca, Ce, Cr, Cs, Cu, Fe, Hf, K, La, Li, Mg, Mn, Na, Nb, Ni, P, Pb, Rb, S, Sb, Sc, Sr, Th, Ti, V, Zr, Be, Co, Ga, Sn, U, and Y contributed mainly to Factor 1. The contribution of As, Bi, and Mo to Factor 1 is minor loadings, while in the opposite direction a very minor loading is contributed from Au, Ag, Te, Ge and In(Fig. 8). Au, Ag, Bi, Mo, Sb, Te,W, Cd, Se, and Tl make a major contribution to Factor 2 with minor loadings from Ta while in the opposite direction a very minor loading is contributed from Cu. Factor 3 is contributed to by large positive loadings from Ta. Au (Table 6) have been noted to be associated with both factors

Table 4-4: Sorted Varimax rotated loadings for the first six factors

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	h2	u2	com
Au	-0.08	0.3	0	0	0.09	0.02	0.11	0.89367	1.3
Ag	-0.1	0.96	-0.07	-0.12	0.01	0.01	0.94	0.05503	1.1
Al	0.99	0.05	0.08	0.06	0.05	-0.02	1	0.00061	1
As	0.63	0	0.01	0.07	0.77	0	1	0.00453	1.9
Ba	0.99	0.05	0.07	0.05	0.05	-0.02	1	0.00074	1
Bi	0.1	0.83	0.01	-0.25	-0.02	-0.01	0.77	0.22917	1.2
Ca	0.99	0.05	0.05	0.06	0.05	-0.01	1	0.00424	1
Ce	0.99	0.05	0.11	0.05	0.05	-0.01	1	0.00149	1
Cr	0.99	0.05	0.04	0.06	0.05	0	1	0.00229	1
Cs	0.96	0.05	-0.04	0.05	0.06	0.13	0.95	0.05018	1.1
Cu	0.57	-0.61	0	0.26	0.21	0.03	0.82	0.18316	2.6
Fe	0.98	0.05	0.04	0.06	0.05	0.13	0.99	0.00589	1.1
Hf	0.97	0.05	0.21	0.05	0.05	0	1	0.00275	1.1
K	0.99	0.05	0.09	0.06	0.05	-0.03	1	0.0008	1
La	0.98	0.05	0.14	0.05	0.05	-0.01	1	0.00296	1.1
Li	0.99	0.05	0.06	0.06	0.06	0.02	1	0.00369	1
Mg	0.99	0.05	0.03	0.06	0.05	0.01	1	0.00074	1
Mn	0.99	0.05	0.1	0.06	0.05	-0.01	1	0.00214	1
Mo	0.2	0.91	0.06	0.21	-0.01	0	0.92	0.08196	1.2
Na	0.99	0.05	0.09	0.05	0.05	-0.05	1	0.00291	1
Nb	0.95	0.05	0.29	0.05	0.04	-0.05	1	-0.0002	1.2
Ni	0.99	0.05	0.02	0.06	0.05	0.02	1	0.00276	1
P	0.99	0.05	0.07	0.06	0.05	-0.01	1	0.00446	1
Pb	0.99	0.05	0.06	0.06	0.06	-0.01	0.99	0.01102	1
Rb	0.99	0.05	0.1	0.05	0.05	-0.01	1	0.00139	1
S	0.98	0.05	0.04	0.06	0.06	0	0.98	0.01976	1
Sb	0.26	0.93	0	-0.05	0.08	0	0.95	0.04879	1.2
Sc	0.99	0.05	0.03	0.06	0.06	0.05	1	0.00341	1
Sr	0.99	0.05	0.05	0.05	0.05	-0.02	1	0.00307	1
Ta	0.3	0.01	0.93	0.01	-0.01	-0.04	0.96	0.04116	1.2
Te	-0.04	0.99	0.01	-0.01	-0.05	-0.01	0.98	0.02123	1
Th	0.99	0.05	0.1	0.05	0.05	0.02	1	0.00385	1
Ti	0.99	0.05	0.09	0.06	0.05	-0.01	1	0.00091	1
V	0.99	0.05	0.06	0.06	0.05	0.01	1	0.00121	1
W	0.23	0.9	0.05	0.02	0.03	0	0.87	0.127	1.1
Zn	0.66	0.06	0.05	0.69	0.08	0	0.93	0.06714	2
Zr	0.99	0.05	0.13	0.05	0.05	-0.03	1	0.00043	1.1
Be	0.9	0.04	0.39	0.05	0.04	0.1	0.97	0.02564	1.4
Cd	-0.03	0.97	0	0.05	-0.07	0	0.95	0.04799	1
Co	0.99	0.05	0.01	0.06	0.05	0.03	1	0.00355	1
Ga	0.99	0.05	0.09	0.06	0.05	0.01	1	0.00079	1
Ge	-0.98	-0.05	-0.05	-0.06	-0.05	0.11	0.98	0.02002	1
In	-0.99	-0.05	-0.03	-0.06	-0.05	0.09	1	0.00263	1
Se	0.14	0.94	0.02	0.15	-0.03	-0.01	0.92	0.08165	1.1
Sn	0.77	0.04	0.58	0.04	0.05	0.2	0.97	0.02721	2
Tl	0.15	0.96	0.02	0.1	-0.05	-0.01	0.96	0.03717	1.1
U	0.97	0.05	0.05	0.05	0.06	0.04	0.95	0.0462	1
Y	0.98	0.05	0.15	0.05	0.05	0	1	0.00171	1.1

4.6. LINEAMENTS

Mapping geologic lineaments is important for mineral exploration because of their potentials for harboring ore bodies that are carried and deposited. In this study, lineaments were extracted from hillshade images which enhance these patterns. The image result of hillshade process (Figure 4-1 to Figure 4-4) showed linear features, such as valley lineament or the mountain peak on satellite images looked more prominent than surrounding area. The delineation of lineaments was supported with some processing tools such as the creation of a hillshade image, with the definition of azimuth of the illumination source and altitude. All of hillshade result imageries were displayed one after another and all linear features were digitized (Figure 4-5). The lineaments were then integrated with contour maps areas to determine gold potential areas.

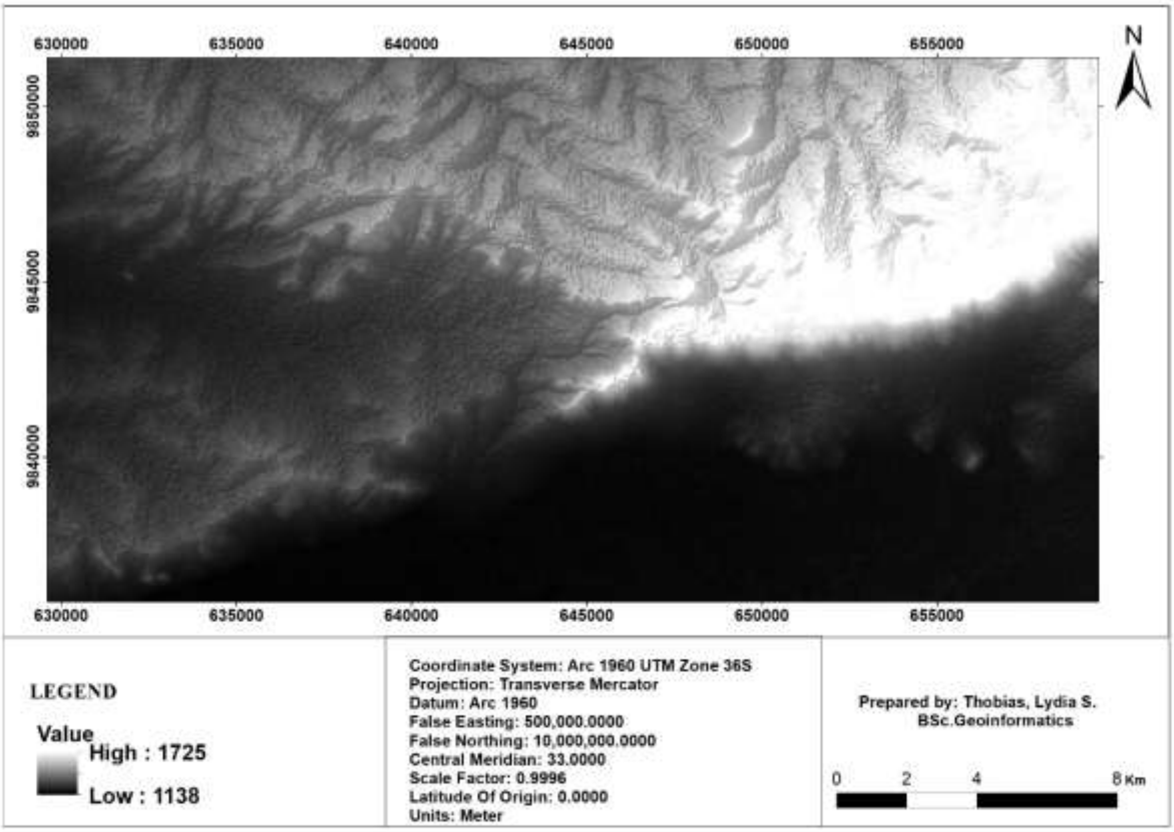

Figure 4-1: SRTM-DEM of Mara Greenstone Belt (Source: USGS)

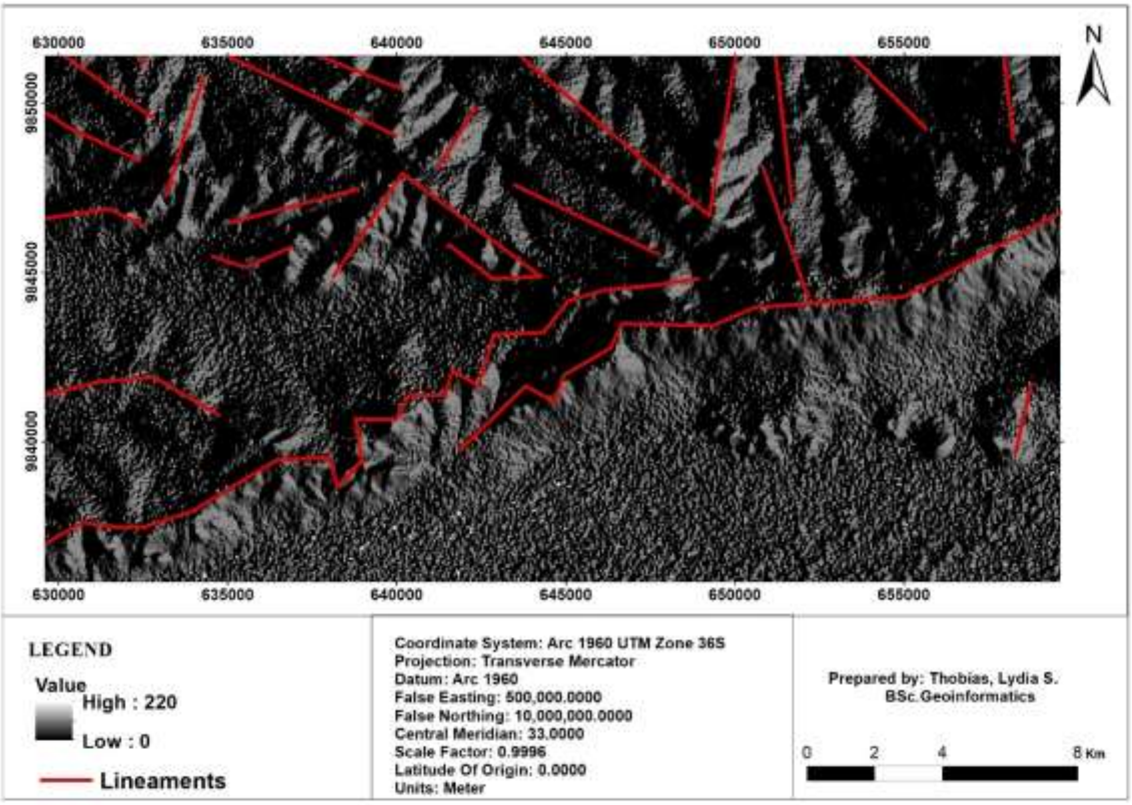

Figure 4-2: Lineaments overlaid with hill shade of azimuth 100 and altitude 60

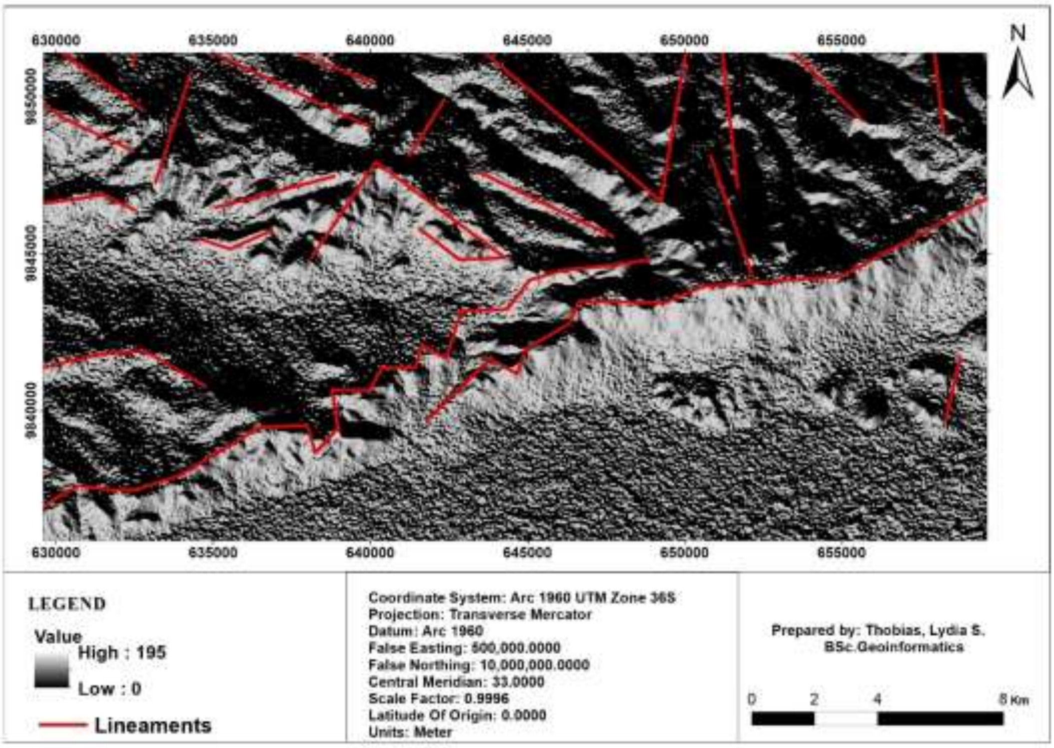

Figure 4-3: Lineaments overlaid with hill shade of azimuth 200 and altitude 50

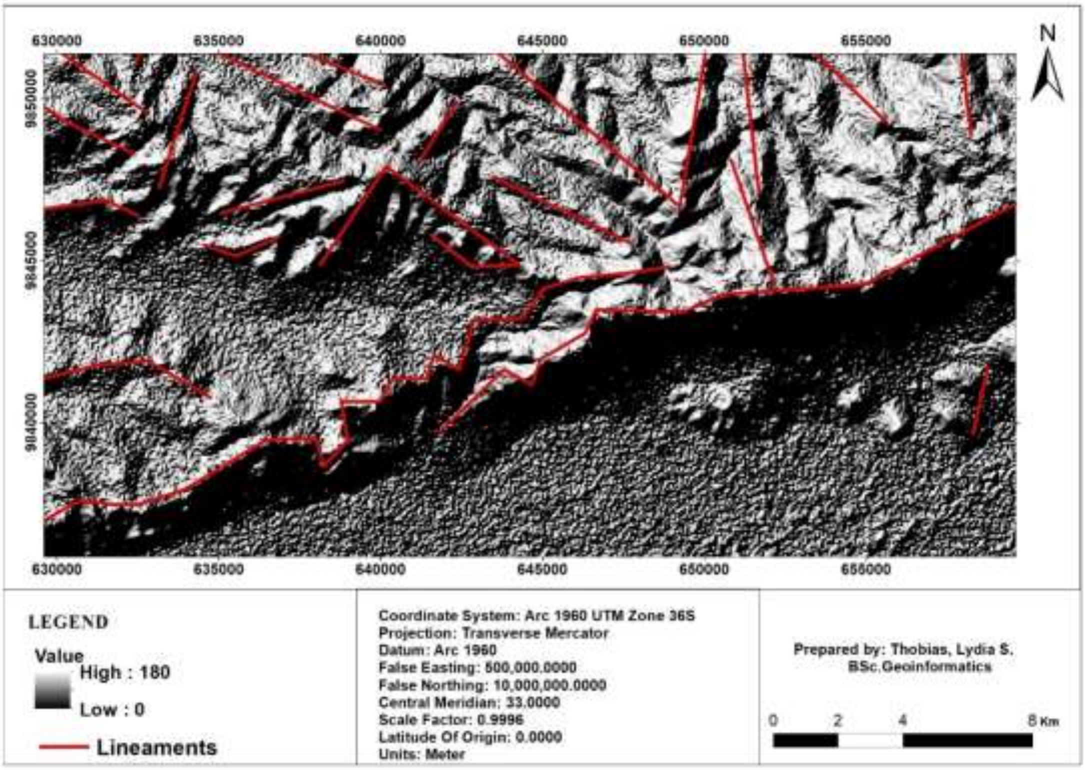

Figure 4-4: Lineaments overlaid with hill shade of azimuth 315 and altitude 45

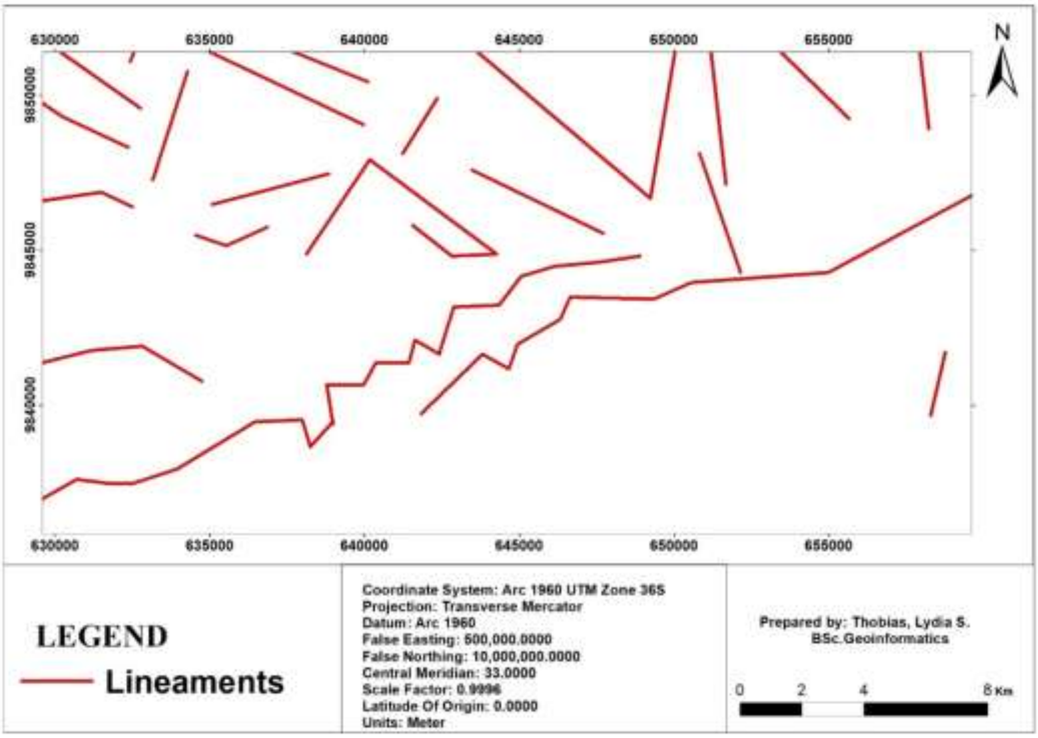

Figure 4-5: Extracted Lineaments

4.7. GOLD ANOMALIES AND TRENDS

Trends and patterns of gold mineralization and its associate elements are successfully visualized when contour plot maps and gridded maps are produced for all the geochemical elements tested Elements that were clustered in the subclusters of the gold group and identified to be potential pathfinders were plotted. Au, Se, Te and Tl (slightly) show anomalies trending ESE–WNW(Fig. 4-6 to 4-9), perpendicular to the major faults and parallel to the shear zones (Fig. 4-10).

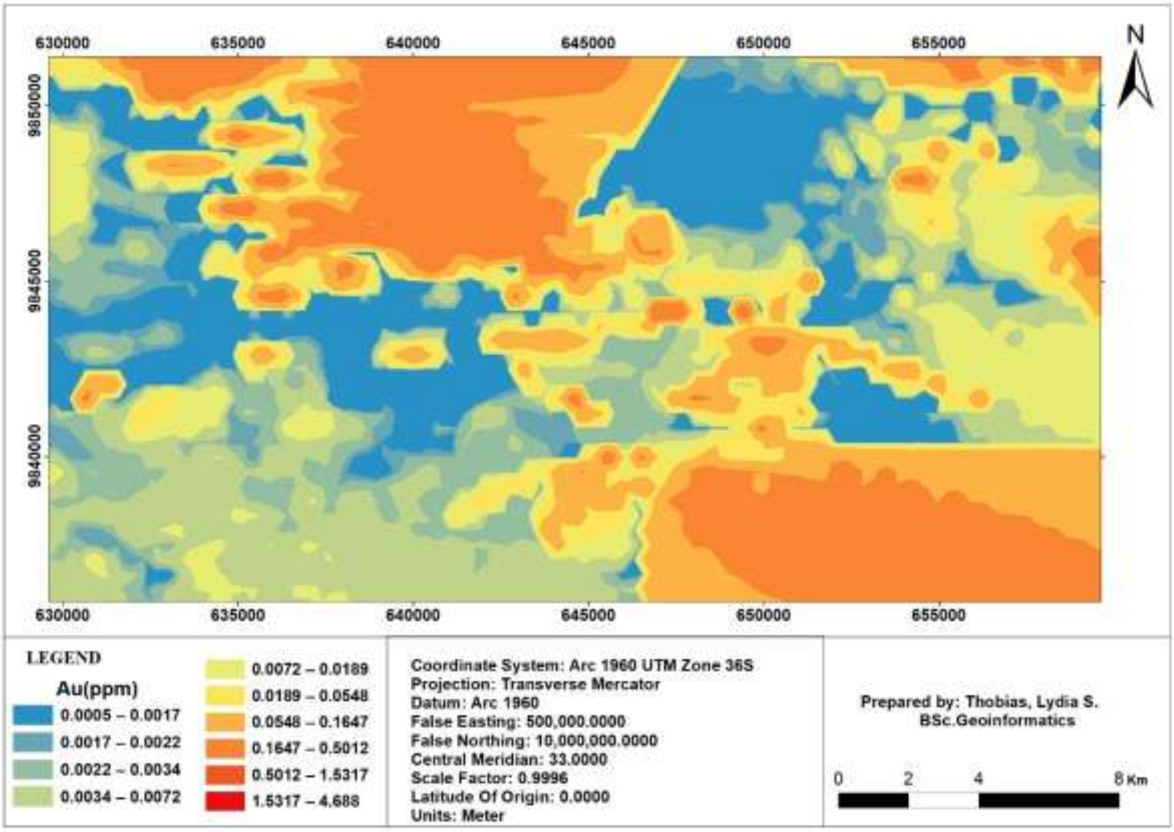

Figure 4-6: Spatial distribution map showing anomalies trend for Gold (Au)

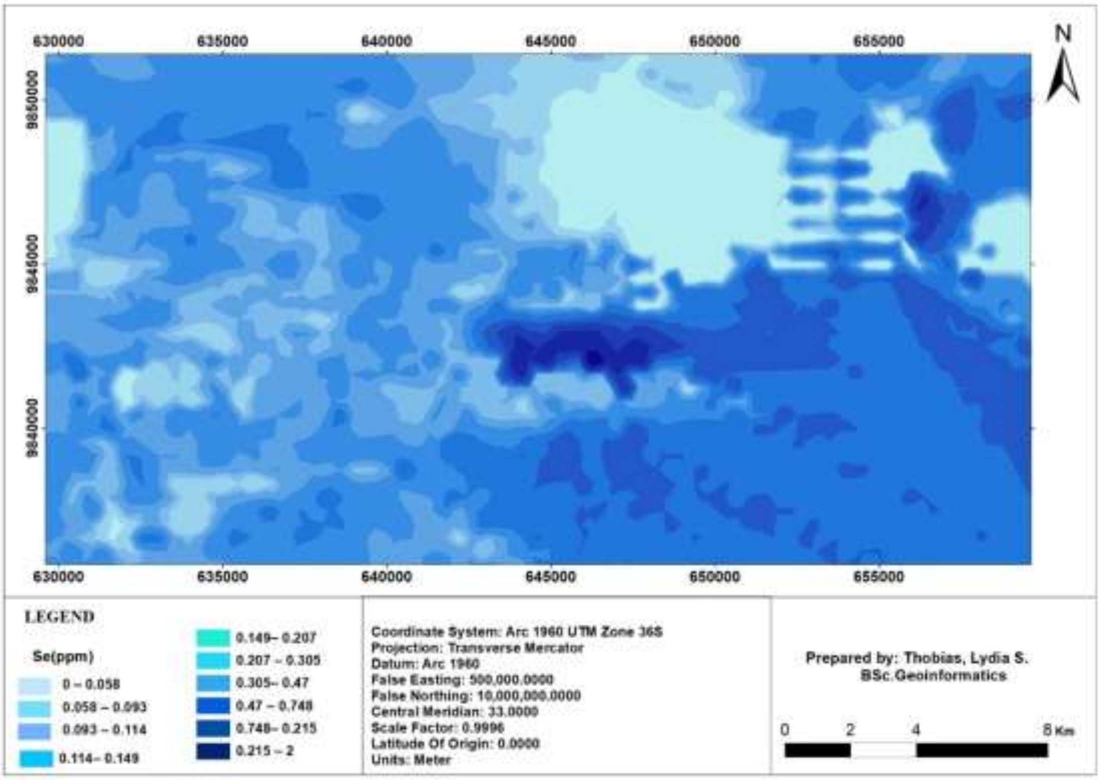

Figure 4-7: Spatial distribution map showing anomalies trend for Selenium (Se)

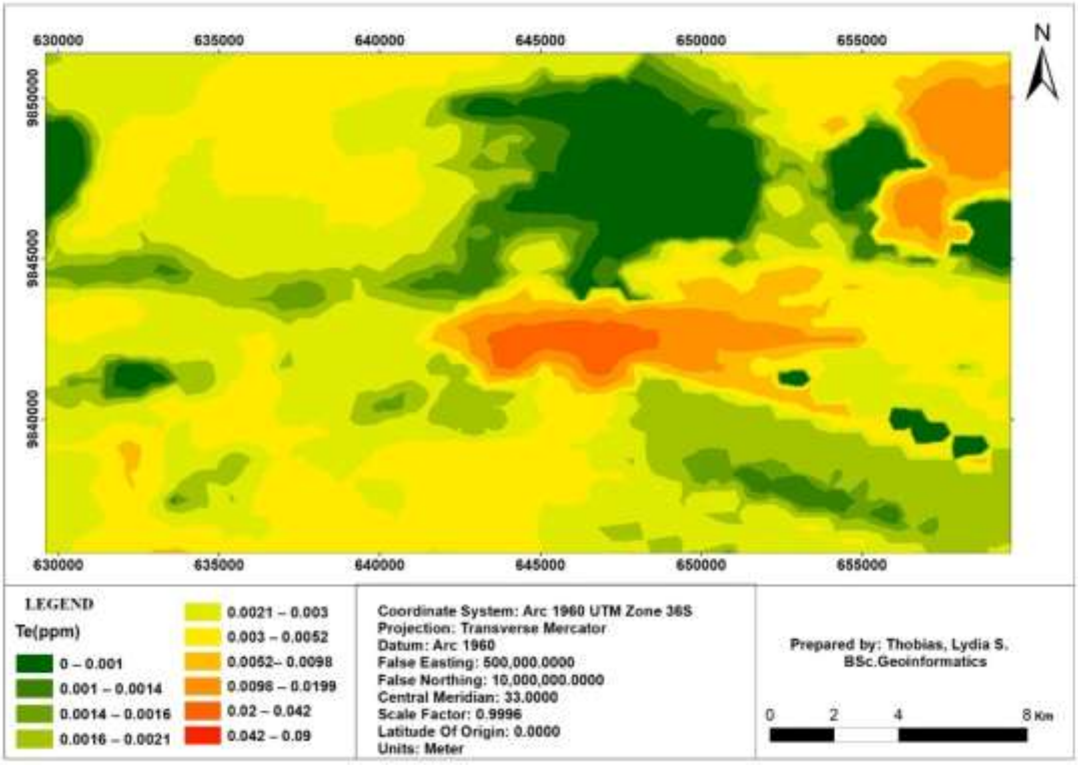

Figure 4-8: Spatial distribution map showing anomalies trend for Tellurium (Te)

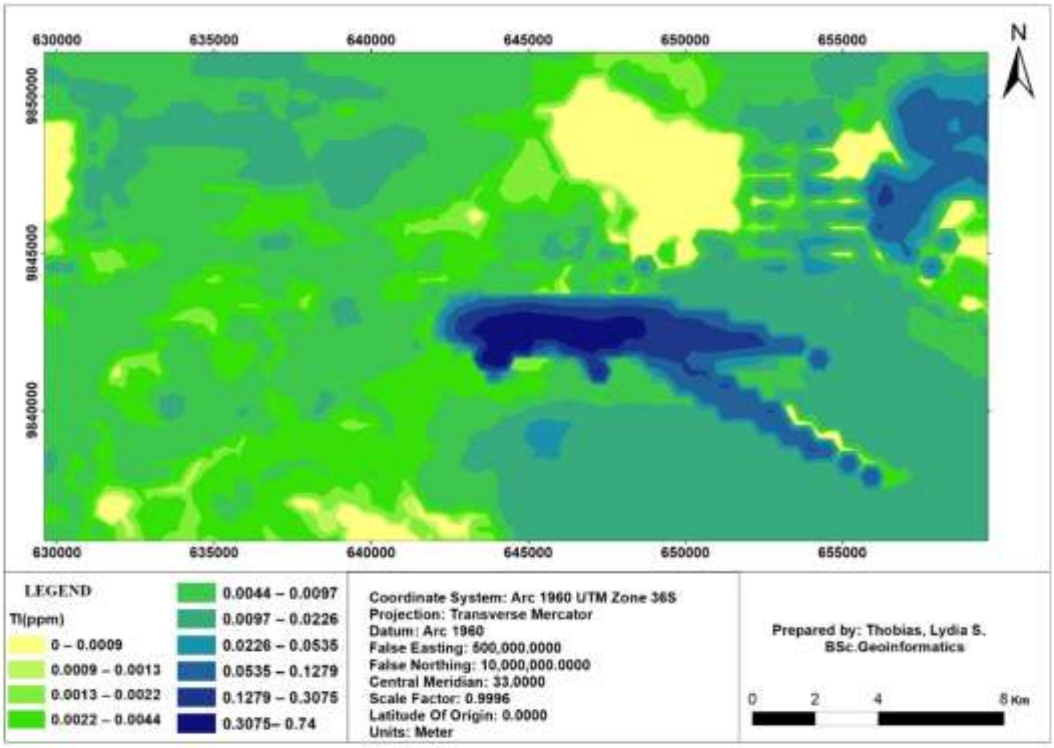

Figure 4-9: Spatial distribution map showing anomalies trend for Thallium (Tl)

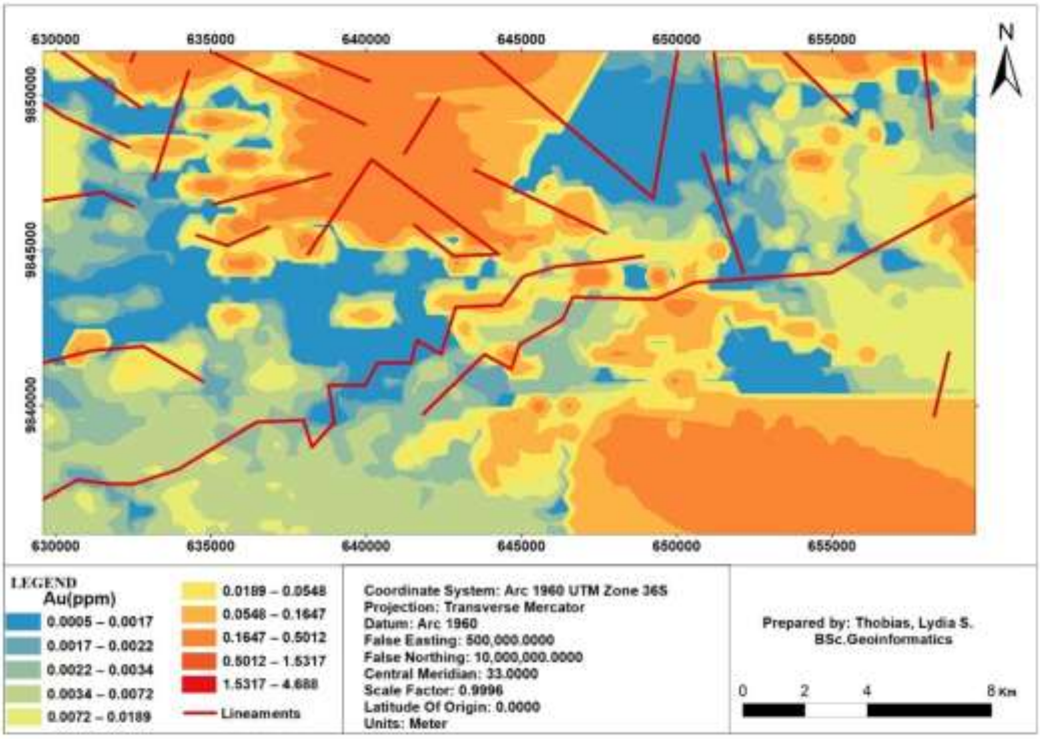

Figure 4-10: Spatial distribution map showing anomalies trend for Gold (Au) overlaid with lineaments

4.8. SUPERPOSITION OF GOLD CONTENT WITH SLOPE

The SRTM-DEM leads to the construction of slope map which shows spatial variation of the slope and enables rapid identification of flat zones. The superposition of gold content and slope maps show that the surface geochemistry which gold content are found around low lands and that the gold content are also very high around sloppy areas. The high gold content observed in hilly or sloped areas could be linked to different geological processes, hence show that the area has undergone tectonic activity or deformation, which could have led to the formation of mineralized veins or structures that host gold deposits. The presence of faults, shear zones, or other geological features associated with gold mineralization can be more pronounced in such areas with higher relief, potentially explaining the elevated gold content

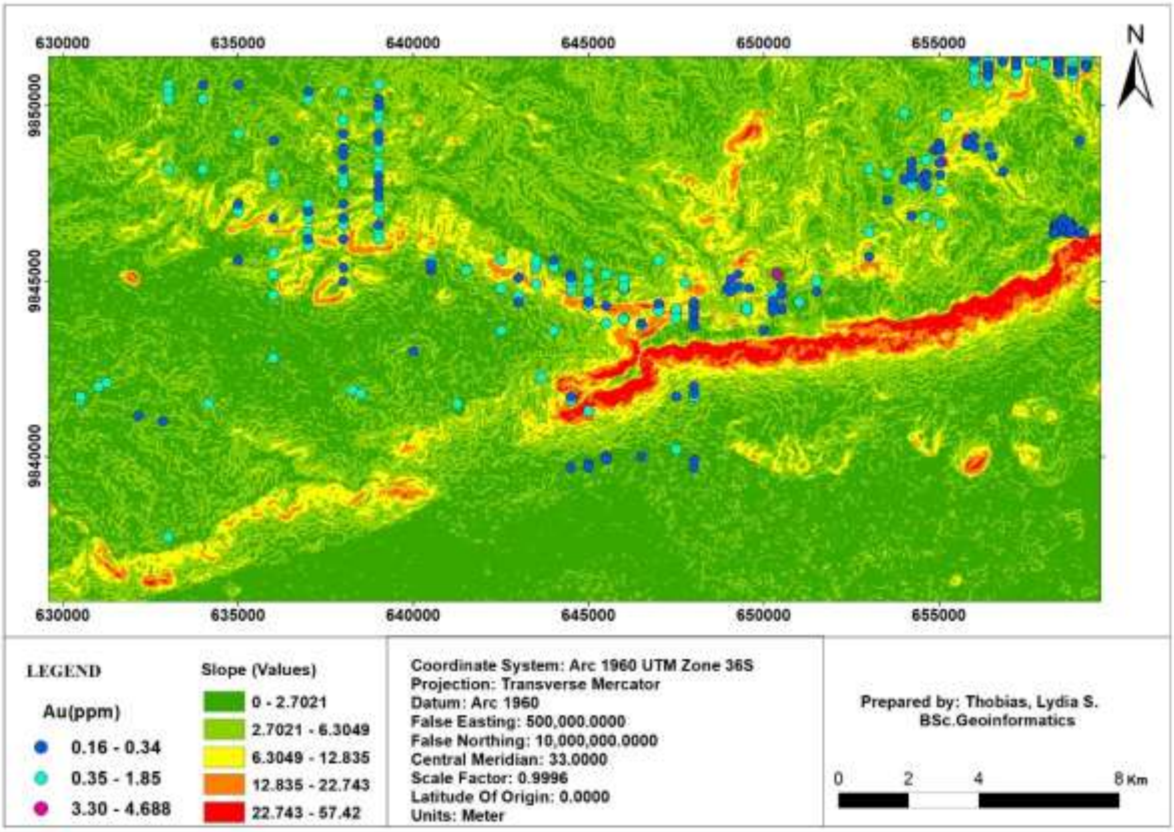

Figure 4-11: Superposition of Gold content with slope

4.9. VALIDATION OF THE RESULT

The overlaid spatial analysis map which was obtain through geostatistical method which is kriging interpolation and the mineral occurrences data shows that the areas with slightly or areas with high value of gold (ppm) are highly associated with the occurrence of minerals as shown in figure 12

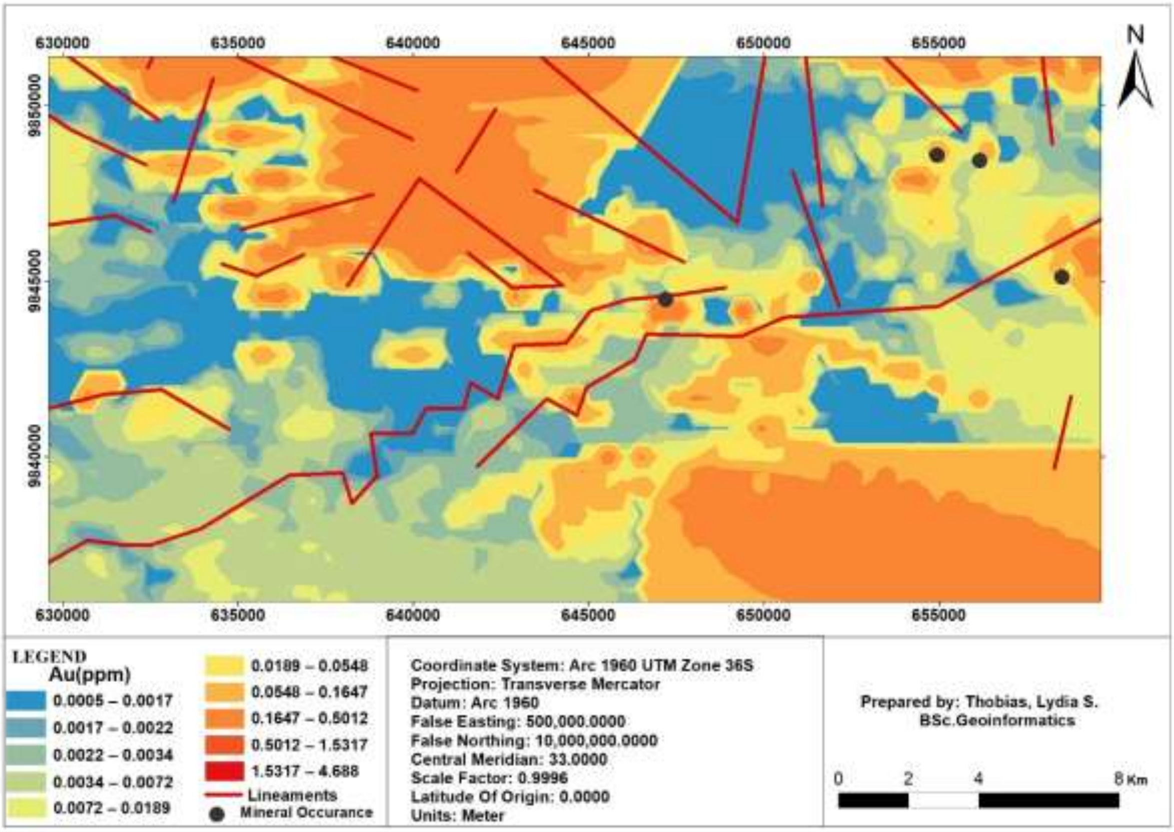

Figure 4-12: Spatial Analysis of Gold anomaly with mineral occurrence

The spatial distribution of gold anomalies with mineral occurrences, the superposition of gold content with slope, and the generated targets for gold potential is a common approach in mineral exploration and resource assessment. This can provide valuable insights into the relationships between geological features, topography, and gold mineralization.

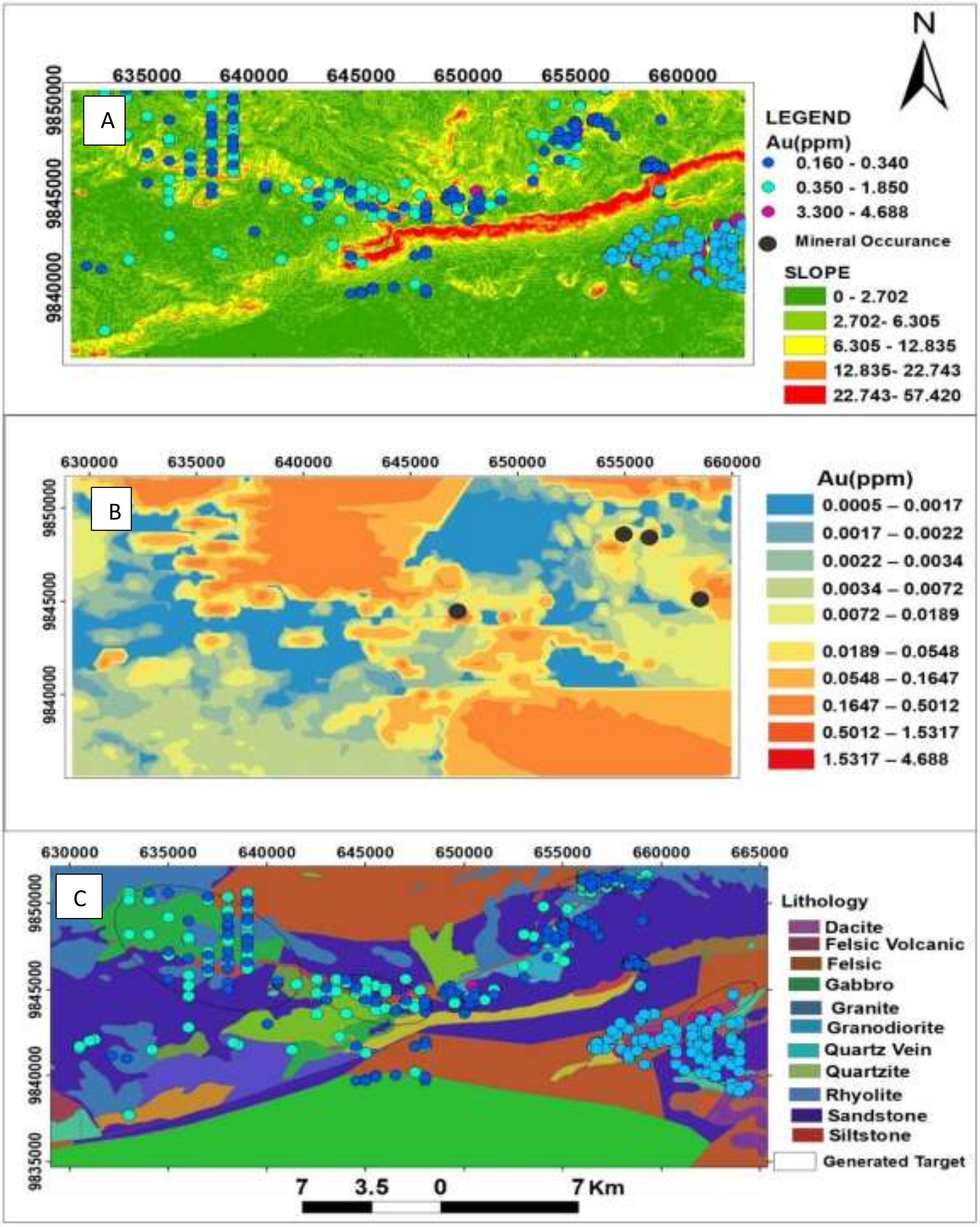

Figure 4-13: A comparison map showing A) superposition of gold content with slope, B) Spatial distribution of gold anomalies overlaid with mineral occurrences and C) Generated targets for gold potential

4.10. GENERATED TARGETS

Five targets have been identified in the northern part and Eastern part of the area. Line segments were drawn on the maps to represent anomalous trends and distributions (Figure 4-12) generally at the boundary between high and low concentrations (Jang et al. 2020; Kreuzer et al. 2020). The generated targets are trending ESE–WNW (Figure 4-10) perpendicular to the major faults and dolerite dyke intrusions (with a NE–SW trend). However, a few spots of higher gold concentrations observed in the part where there is no anomalous trend might be related to mineralization. From Figure 10, it has been noted that gold mineralization may be associated with tonalite, tonalitic orthogneiss (Thomas 2010) and hornblende gneiss interlayered with quartzite, with shearing being the controlling factor for mineralization emplacement.

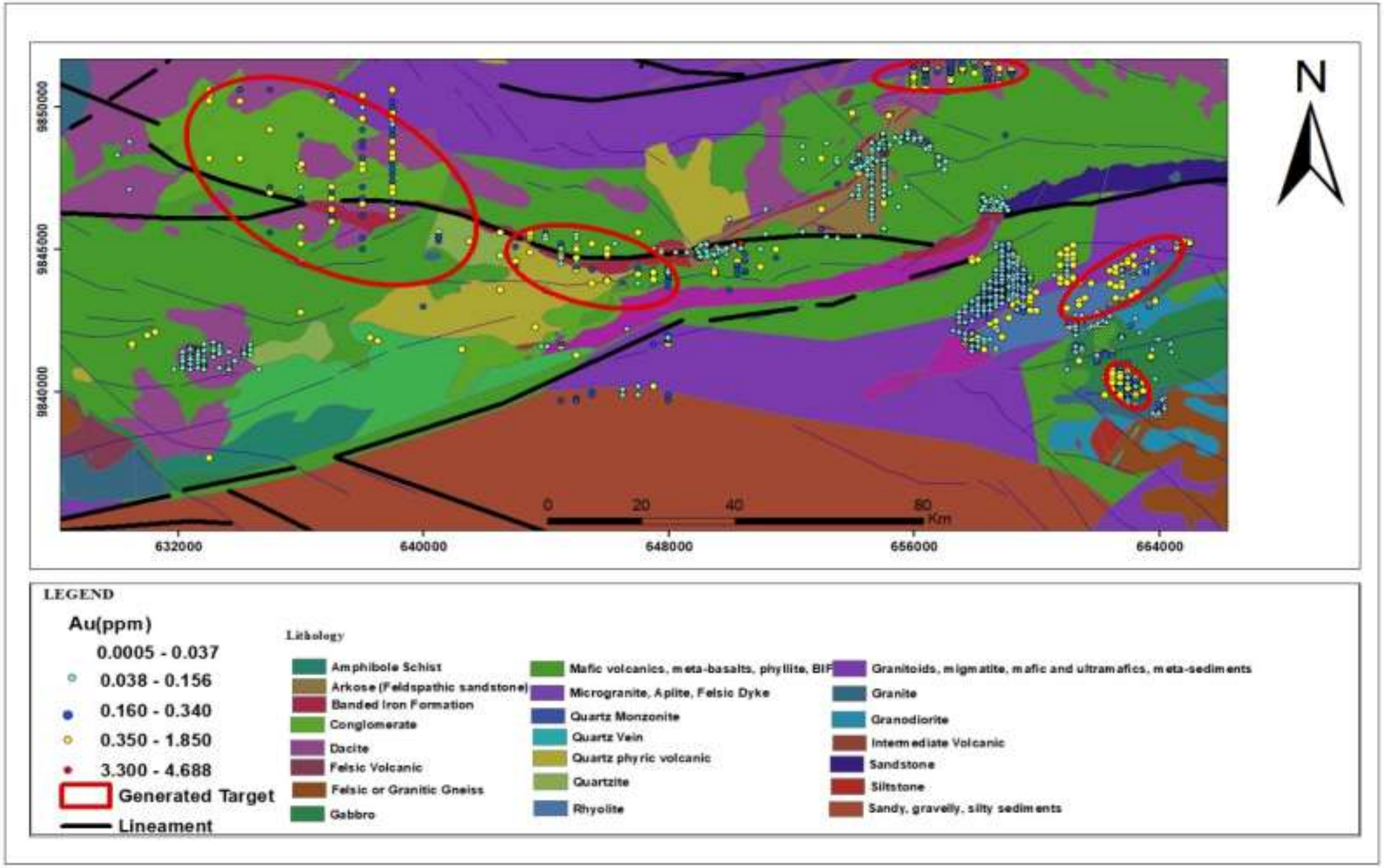

Figure 4-14: Geological Map showing target generated

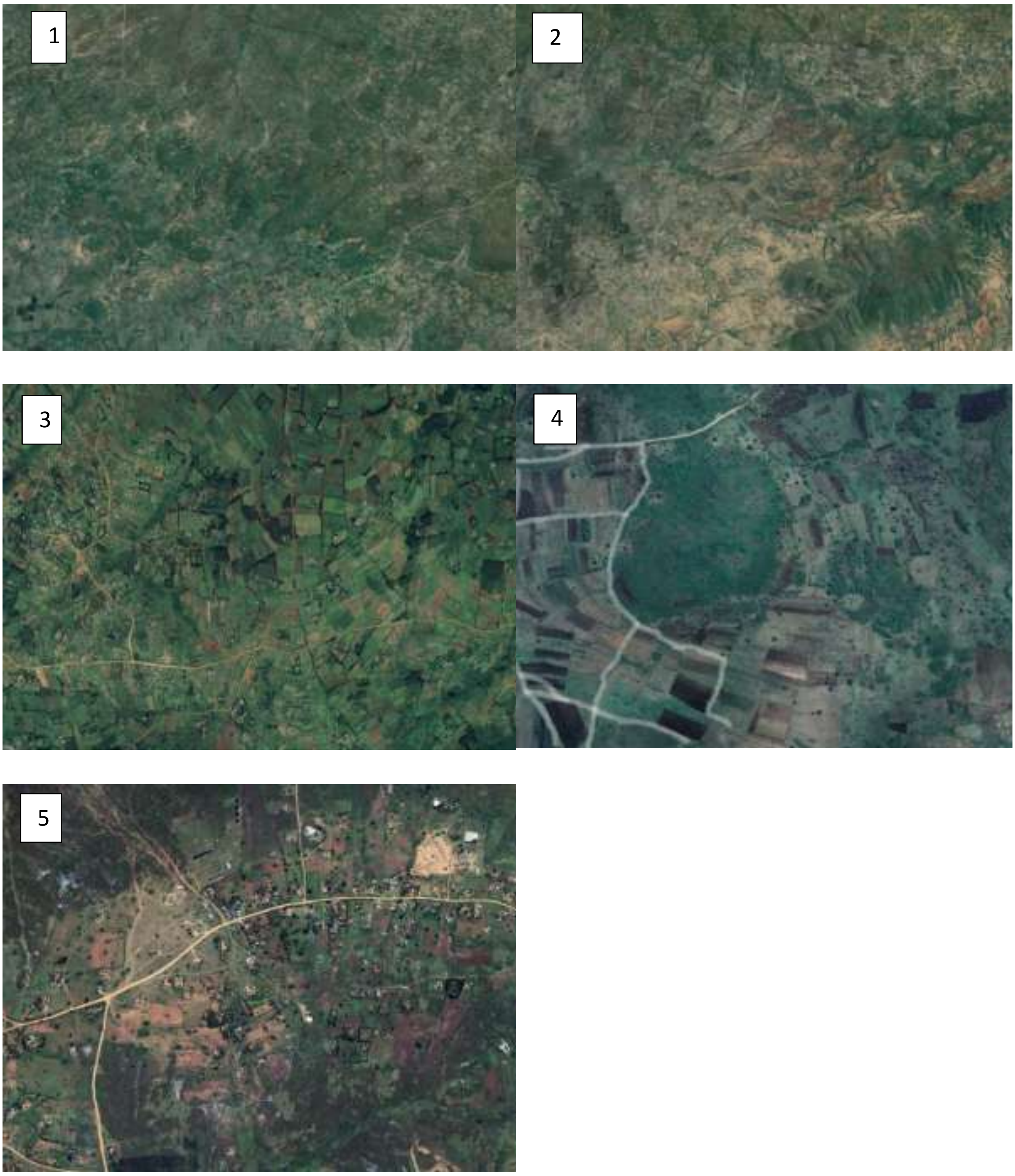

Figure 4-15: Identified targets for gold potential (Source: Google Earth Images)

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.0. OVERVIEW

The chapter is based on the findings from the results and analysis chapter in order to respond to the study's research questions, which are summarized in the conclusion. This chapter contains the conclusion as well as recommendation

5.1. CONCLUSION

The integration of geostatistical and GIS methodologies provides a robust and effective approach for processing and interpreting geochemical data in mineral exploration. By combining these techniques, researchers and geologists can overcome the challenges of limited sampling, spatial gaps, and the removal of background variations. The methodology allows for the interpolation of data points, enabling a more comprehensive understanding of the spatial distribution of geochemical elements. Through the visualization of interpolated data on maps, trends, anomalies, and areas of interest can be identified and analyzed.

The research objectives focused on developing techniques for interpolating geochemical data, removing background variations, integrating geostatistical tools with GIS platforms, identifying lineaments and anomalous trends, determining zones with high concentrations of gold and associated elements, and generating contour maps and gridded maps. These objectives were achieved through the implementation of the developed methodology.

5.2. RECOMMENDATIONS

- Further research should be conducted to validate and refine the developed methodology by applying it to different geological settings and datasets. This will help assess its applicability and performance across various scenarios.

- Collaboration between exploration companies, researchers, and GIS professionals should be encouraged to leverage their collective expertise and resources in integrating geostatistical and GIS methodologies effectively.

- Continuous advancements in geostatistical and GIS software should be monitored and utilized to enhance the capabilities of the methodology. Regular updates and training in the software platforms are recommended to stay up to date with the latest functionalities and features.

- The methodology should be integrated into the standard workflows of exploration companies to improve decision-making processes and optimize exploration strategies. Training programs and workshops can be organized to familiarize professionals with the methodology and its practical implementation.

- Ongoing monitoring and evaluation of the methodology should be carried out to assess its effectiveness and identify areas for improvement. Feedback from users and stakeholders should be collected to further refine and enhance the methodology.

REFERENCES

Barth, M.G. (1990). The Geology of the Tanzanian Craton. Geological Society, London, Special Publications, 46(1), 155-165.

Bentahar, M., et al. (2020). Applications of Principal Component Analysis in Satellite Image Processing. In Data Analysis in Management with SPSS Software (pp. 173-192). Springer.

Borg, G. and Schleton, P. (1997). The Geology of the Geita Greenstone Belt, Northern Tanzania: Implications for Structural Analysis and Tectonic Evolution. Journal of African Earth Sciences, 24(3), 357-378.

Duffield, J. E. (2007). Cluster analysis applied to mineral prospectivity mapping. Natural Resources Research, 16(1), 1-18.

Frutuoso, E. (2021). Principal Component Analysis (PCA). In Computer Vision and Image Processing: A Practical Approach Using CVIPtools (pp. 135-145). CRC Press.

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis (6th ed.). Pearson Education.

Kocal, G. (2004). Detection of linear structures in multispectral images with Haar wavelet transform. Computers & Geosciences, 30(5), 505-511.

Koike, K., Hosono, T., Shigematsu, H., Tsutsumi, H., Sassa, K., & Ueno, S. (1995). An automatic lineament extraction algorithm for the detection of surface fractures using aerial photographs and satellite images. Computers & Geosciences, 21(7), 855-862.

Kwelwa, S.D., Pernicka, E., Meisel, T., Grevel, C., and Muhongo, S. (2013). Archaean to Palaeoproterozoic Magmatism and High-Grade Metamorphism in the Tanzanian Craton: The Nyanzian-Kavirondian Event and Its Overprint. Precambrian Research, 228, 322-352.

Lillesand, T.M., Kiefer, R.W., and Chipman, J.W. (2004). Remote Sensing and Image Interpretation (6th ed.). John Wiley & Sons.

Liu, J., et al. (2014). Geochemical distribution and source apportionment of heavy metals in sediments of the Yellow River Estuary and nearby coastal areas. Marine Pollution Bulletin, 85(1), 448-456.

Liu, J., et al. (2019). Geochemical anomaly separation using hierarchical cluster analysis and principal component analysis in a polymetallic ore field. Ore Geology Reviews, 104, 1-13.

Liu, J., et al. (2020). Geochemical anomaly separation using a combination of hierarchical cluster analysis and principal component analysis: A case study of the Shuangjianzishan Pb-Zn deposit in South China. Journal of Geochemical Exploration, 213, 106432.

Manya, S. and Maboko, M.A.H. (2008). Petrogenesis of the Archaean Granitoids of the Dodoman Belt, Tanzania: The Geodynamic Significance. Journal

Masoumi, M., Pour, A.B., and Pour, A.H. (2017). A Review of Lineament Extraction Methods and Techniques: Approaches, Sensors, and Limitations. Geoscience Frontiers, 8(6), 1303-1319.

Moradi, M. (2015). An Introduction to Applied Multivariate Analysis with R. Springer.

Raghavan, V., Rajendran, S., and Rajendran, C.P. (1995). Lineament Extraction Techniques for Geological Applications – A Review. Journal of the Indian Society of Remote Sensing, 23(2), 95-107.

Raghavan, V., Raju, P. L. N., & Rao, B. (1995). Manual extraction of lineaments from aerial photographs and satellite imagery. Photogrammetric Engineering and Remote Sensing, 61(4), 437-441.

Rajesh, N. (2004). Principles of Remote Sensing. New Age International.

Rencher, A. C. (2003). Methods of Multivariate Analysis (2nd ed.). Wiley-Interscience.

Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th ed.). Pearson Education.

Van der Meer, F. D. (2012). Multi- and hyperspectral geologic remote sensing: A review. International Journal of Applied Earth Observation and Geoinformation, 14(1), 112-128.

Wang, Z., Chen, C. H., & Thompson, D. W. (1990). Line extraction in digital images using the Hough transform. Image and Vision Computing, 8(2), 107-115.

Zuo, R., et al. (2018). Multivariate statistical analysis of groundwater quality in a coastal area with extensive saltwater intrusion. Science of The Total Environment, 636, 1521-1529.