

https://dspace.org/

Using machine learning to identify the spatial extent of acid mine drainage from gold mining activities in Gauteng, South Africa

Bullah, Nikita

2025

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/50153

Bullah, Nikita. (2025). Using machine learning to identify the spatial extent of acid mine drainage from gold mining activities in Gauteng, South Africa. [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/50153

Downloaded from a DSpace repository. https://dspace.org/



Using machine learning to identify the spatial extent of acid mine drainage from gold mining activities in Gauteng, South Africa

Nikita Bullah (2303383)

Supervisor: Dr Iqra Atif

Co-supervisor: Dr Muhammad Ahsan Mahboob

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science

October 2025

School of Geography, Archaeology and Environmental Studies

Private Bag 3. Wits 2050, South Africa Enquiries: GEOGRAPHY: TEL: +27 11 717-6503 • Fax: +27 11 403-7281
ARCHAEOLOGY: TEL: +27 11 717-6045 • Fax: +27 11 339-1620

WITS UNIVERSITY

FULL NAME : Nikita Bullah

STUDENT NUMBER : 2303383

YEAR : 2025

COURSE NAME : Master of Science

SUPERVISORS : Dr Iqra Atif, Dr M Ahsan Mahboob

DATE OF SUBMISSION : 19 March 2025

TOPIC (in full) : Using machine learning to identify the spatial extent of acid mine drainage from gold mining activities in Gauteng, South Africa.

DECLARATION

I Nikita Bullah (Student number 2303383) am a student registered for MSc Geography, Archaeology, and Environmental Studies in the year 2025

I hereby declare the following:

* I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong

* I confirm that the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise

* I have followed the required convention in referencing the thoughts and ideas of others.

* I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature

Date 19/03/2025

Wits University Faculty of Science post-graduate student AI declaration

I understand that the use of generative AI tools (such as ChatGPT or similar) without explicitly declaring such use constitutes a form of plagiarism and is classified by Wits University as academic misconduct.

I declare that in the course of conducting the research towards my degree or in the preparation of this thesis/dissertation/research report (select one by marking with an X):

I **did not** make use of generative AI tools [X]

I **did** make use of generative AI tools for the following (tick all that apply):

1. Idea Generation (research problem/design, hypothesis) [ ]
2. Sourcing Related Work (summarising, identifying sources) [ ]
3. Methods and Experiment Design (experiment setup, model tuning) [ ]
4. Data Analysis (presentation, coding, interpretation) [ ]
5. Theoretical Development (theorem proving, conceptual analysis) [ ]
6. Code Development (generating algorithms, writing scripts) [ ]
7. Presentation (rendering graphics, formatting) [ ]
8. Editing (grammar, readability) [ ]
9. Writing (text generation, document structuring) [ ]
10. Citation Formatting (structuring, organising) [ ]

If other uses were involved, please specify below:

Generative AI tool used (list all)	**Used for?**

If generative AI tools were used as an integral part of the experimental design or in the direct execution of my research, I confirm that details of this use are clearly outlined in the relevant experimental/methodology chapters of my thesis/dissertation/research report.

Student number: 2303383

Candidate signature:



Date: 24/10/2025

ACKNOWLEDGEMENTS

This research would not have been possible without the help and supervision of Dr Iqra Atif and Dr Muhammad Ahsan Mahboob. I would like to express my gratitude to the Sibanye-Stillwater Digital Mining Laboratory (DigiMine) at the University of the Witwatersrand for providing access to the handheld XRF equipment and for offering the necessary training required for its operation. Many thanks also goes out to the managers at each sample point that provided me with permission to carry out my research within their facilities. Many of these areas were privately held tourist locations. Lastly, I would like to thank my mother who assisted me with collecting data at each sample point.

ABSTRACT

Acid mine drainage (AMD) poses a significant environmental and public health risk, particularly in mining-intensive regions. This study examines the spatial extent of AMD contamination in the Bloubankspruit and Crocodile River systems in Gauteng, South Africa, by analysing heavy metal concentrations in surface soils along a 20 km stretch downstream of the abandoned Kromdraai gold mine. Soil samples were collected at six locations at 4 km intervals and analysed using X-ray fluorescence (XRF) spectroscopy. The results indicate that AMD contamination is most severe within the first 4 km, with sulphur concentrations reaching 6 911 mg/kg, exceeding the acceptable threshold of 4 000 mg/kg by 72,8%. Cobalt levels were elevated across all sample sites, with the highest recorded concentration at 49.2 mg/kg—over eight times the permissible limit of 6 mg/kg. Chromium and copper concentrations also surpassed regulatory standards at 40,6 mg/kg and 32,6 mg/kg, respectively. A statistical analysis of variance (ANOVA) confirmed significant variation in contamination levels across sites ($F = 30,23$, $p < 0,05$), with AMD pollutants decreasing exponentially downstream due to natural dilution and sedimentation processes. The study further demonstrates that the gentle topographical gradient (0,0097) of the region limits surface runoff, constraining AMD migration beyond the immediate vicinity of the mine. The findings emphasize the urgent need for targeted remediation efforts, such as bioremediation and passive treatment systems, to mitigate the long-term environmental impact of AMD. This research provides critical insights into the transport dynamics of mining-related pollutants and supports the development of more effective water quality management strategies in affected river systems.

KEYWORDS

Acid Mine Drainage (AMD), Heavy Metal Contamination, Gold Mining Pollution, Geochemical Analysis, Water Quality Management, Soil Contamination, Environmental Remediation.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF ACRONYMS viii

1. INTRODUCTION 1

1.1 BACKGROUND 1

1.2 RESEARCH PROBLEM STATEMENT 4

1.3 AIM AND OBJECTIVES 4

1.3.1 Aim 4

1.3.2 Objectives 4

2. LITERATURE REVIEW 5

3. MATERIALS AND METHODS 26

3.1 STUDY AREA 26

3.3 DATA REQUIREMENTS 30

3.4 METHODS 30

4. RESULTS 39

5. DISCUSSION 56

5.1. INTERPRETATION OF RESULTS 56

5.2. COMPARISON WITH LITERATURE 59

5.3. ENVIRONMENTAL AND SPATIAL INFLUENCES 62

5.4. MACHINE LEARNING ANALYSIS 62

5.5. IMPLICATIONS FOR REMEDIATION 64

5.6. STUDY LIMITATIONS AND FUTURE RESEARCH 65

6. CONCLUSION 69

7. REFERENCES 73

8. APPENDIX 86

8.1 ETHICS WAIVER 86

8.2 RAW XRF COLLECTED DATA 87

LIST OF TABLES

Table 1: Water contaminants and risk types as per the SANS 241 and WHO guidelines (World Health Organisation, 2011; South African Bureau of Standards, 2022).....8
Table 2: Soil Screening Values (SSV) for different LULC types 20
Table 3: Data requirements for the implementation of this study 30
Table 4: Mean amounts of elements present per sample site.....43
Table 5: Skewness and kurtosis for each element 46
Table 6: ANOVA results representing all collected sample data 47
Table 7: 95% Confidence intervals for each element over each sample site.....49
Table 8: Accuracy matrixes for all cross classification results.....54
Table 9: Accuracy metrics for analysing the performance of cross classification SVM.....54

LIST OF FIGURES

Figure 1: Study area of the Bloubankspruit and Crocodile Rivers 28
Figure 2: Research design 29
Figure 3: 2km buffer of gold mines around the study area of the Bloubankspruit and Crocodile River 40
Figure 4: Land use and land cover around the Bloubankspruit and Crocodile River 41
Figure 5: Digital elevation model of the study area. 42
Figure 6: Elevation profile across all sample sites 42
Figure 7: Mean element concentrations over each sample site 44
Figure 8: Point based uncertainty map across all sample sites 50
Figure 9: General trend in AMD impact over all sample sites 52
Figure 10: Cross classification implementation through support vector machine learning algorithms 53
Figure 11: ROC curve and AUC value for predictive accuracy of SVM model 55

LIST OF ACRONYMS

AMD	Acid Mine Drainage
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area Under the Curve
FN	False Negative
FP	False Positive
GIS	Geographic Information Systems
KNN	K Nearest Neighbour
LULC	Land Use and Land Cover
ROC	Receiver Operating Characteristic
SANS	South African National Standard
SSV	Soil Screening Values
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
USGS	United States Geological Survey
WAWQI	Weight Arithmetic Water Quality Index
WHO	World Health Organization
WQI	Water Quality Index
XRF	X-Ray Fluorescence

1. INTRODUCTION

1.1 BACKGROUND

Acid mine drainage (AMD) is a problem that affects the quality of water systems on a global scale. The main cause of this originates from mining activity related to coal and heavy metals such as gold, copper, and silver due to the exposure of sulphide bearing rocks during the mining process. Drainage flows over sulphide minerals, such as pyrite, due to mining operations and the interaction of these minerals with water and oxygen causes the formation of sulphuric acid, which spreads into the water sources that surround mines (Mujuru, 2014; Curtis, 2021). AMD is therefore caused by the runoff of toxic waste from both active and inactive mines, and is stated to be one of the most damaging forms of water pollution (Mujuru, 2014; Curtis, 2021).

Issues of AMD are particularly present in South Africa, where there is a long history of gold and coal mining. These activities have resulted in many abandoned and unrehabilitated mine sites. Among the most affected regions, is the province of Gauteng (Curtis, 2021). This is due to its position of lying directly over the Witwatersrand basin, which was once one of the world's richest gold deposits. This directly established the foundation of the country's mining economy. Thousands of these sites are currently inactive and poorly maintained, which has caused the spread of AMD to surrounding river systems (Curtis, 2021). This poses a serious threat to water security, ecosystem health, and nearby communities. although there has been some remediation effort, the spatial monitoring of AMD remains uneven, and many contaminated zones go undetected for years.

The introduction of AMD into river systems has detrimental impacts on the ecosystems that reside and rely on the affected aquatic environments. Many aquatic organisms are unable to survive in water with high acidity as they are highly sensitive to changes in pH levels (Ngole-Jeme & Ndava, 2023). In addition to the increased levels of acidity, AMD also introduces the runoff of heavy metals to enter river systems. These metals, such as lead, mercury, and arsenic are highly toxic to organisms and may induce biological changes and mutations in the aquatic life that is exposed to heavy metals. The combination of sulphuric acid and heavy metals within aquatic environments cause drastic issues surrounding the biodiversity of water systems (Alvarenga, Guerreiro, Simões, Imaginário, & Palma, 2021; Ngole-Jeme & Ndava, 2023). The environment may become uninhabitable and cause animal and organism populations to

decrease. Furthermore, plants that rely on water from contaminated sources may become unable to withstand the harsh conditions. Exposure to heavy metals through water or subsequent soil that the water interacts with may hinder plant growth and cause harm to soil organisms which could cause long term damage to vegetation around contaminated water sources (Ighalo, et al., 2022). The harm that the contaminated water causes to soil and vegetation makes it unsuitable for use in agriculture. This is because contaminants cause crops to be unsuitable for consumption and alters soil chemistry. This reduces the microbial diversity of the area and impairs nutrient cycling, which makes vegetation unable to grow (Ighalo, et al., 2022).

AMD contains radioactive particles and toxic heavy metal concentrations which have extreme effects on people and the environment. This is because AMD affects river systems, which in turn affects the nearby soil and ecosystems that rely on these river systems (Curtis, 2021). This also impacts vegetation and agricultural crops that rely on river water to grow. As a result, the health and livelihoods of people within the community are affected by AMD (Mujuru, 2014). Mujuru (2014) suggests that AMD is most critical in African countries because the exploitation of natural and mineral resources is needed to boost economic growth and combat poverty. Therefore, the analysis of AMD within resource rich African countries is important as these areas face severe water contamination, and remediation is urgent. Seeing as South Africa is the most developed mining country in Africa, the evaluation of AMD is crucial (Mujuru, 2014).

The impact of AMD is particularly harmful to aquatic organisms such as macroinvertebrates and native fish species. This is because they are extremely sensitive to the ecosystem's pH and metal bioavailability (Williams & Turner, 2015; Ngole-Jeme & Ndava, 2023). An increase in acidity and presence of heavy metals make the environment inhabitable for these species. When the chemicals associated with AMD settle within sediments, long term ecological decline takes place. This means that vegetation becomes stressed, and soil quality rapidly changes. this is because of the altered state of the soil chemistry and the metal uptake of the vegetation (Williams & Turner, 2015). This destabilises the ecosystem and makes it inhabitable for plants and animals to live in. In the context of Gauteng, these impacts may also have socio-economic impacts. This is because AMD affected rivers often flow through urban areas. Hence, a decline in water quality directly impacts the livelihood, agricultural productivity, and human health of communities that rely of these water sources (Ngole-Jeme & Ndava, 2023).

AMD mapping in the past has relied mainly on hydrological modelling, geochemical simulations, or field-based analyses. The use of machine learning techniques has previously been used for other geographic contexts, such as contamination of groundwater and heavy metal contamination in soil (Barzegar, Moghaddam, Deo, Fijani, & Tziritis, 2018). Although the current application of machine learning in AMD has remained limited, it is promising for future research. Traditional geostatistical approaches often require dense field data and the assumption of spatial continuity or linear relationships (Szramowiat-Sala, 2023). By using machine learning models, particularly the use of SVM, non-linear classification problems can be addressed. Additionally, heterogenous datasets can be integrated, such as ones that use a combination of environmental factors like land use, topography, and spectral signatures (Barzegar, Moghaddam, Deo, Fijani, & Tziritis, 2018).

The research on this topic is important as it allows for an understanding of geochemical processes that develop between sulphide minerals, air, and water. This understanding may prompt mining companies to mitigate processes that lead to AMD and adopt more efficient and sustainable practices. Remediation methods for rivers that have already been affected by AMD may also be addressed. AMD research could also allow for a deeper understanding of ecological impacts and how to approach future contamination. The levels of contamination could also be forecasted through predictive modelling of future AMD. By documenting and researching these problems, there is a higher likelihood that environmental regulations and policies are implemented.

Current AMD monitoring systems heavily rely on field-based water sampling, remote sensing, and a combination of active and passive treatment systems (Barzegar, Moghaddam, Deo, Fijani, & Tziritis, 2018). Active systems include remediation efforts such as chemical neutralisation and passive systems include efforts such as constructed wetlands (RoyChowdhury, Sarkar, & Datta, 2015). Currently, monitoring is spatially fragmented and limited by the availability of monitoring stations. Accessibility to mining sites and the high cost of regular sampling are also challenges. Informal settlements or inactive mining sites also lack consistent data coverage or formal monitoring (Chalkley, 2024). Because of these challenges, there is a clear need for scalable, data efficient, and predictive tools. Utilizing these tools can help in identifying at risk areas without the presence of dense monitoring infrastructure (Chalkley, 2024). Machine learning methods such as SVM offers an approach

that leverages available datasets to model and map AMD risk more completely (Szramowiat-Sala, 2023).

1.2 RESEARCH PROBLEM STATEMENT

South Africa has experienced a water scarcity issue since 2015 due to various environmental and climatic conditions. Seeing as water resources are already in short supply, the risk of AMD contamination within water systems is extremely severe (Bischoff-Mattson, et al., 2020). Over the past few years, the Department of Water Affairs has identified the extent of the threat of AMD within Gauteng. This has specifically been done over the West Rand region, and the department has begun to introduce some short-term interventions (Bobbins, 2015). This however, proves to be ineffective without a long-term strategy to remediate and mitigate further damage to the already compromised health of river systems. A good understanding of the level of contamination within the Gauteng region needs to firstly be assessed in order to combat existing damage and prevent damage to Gauteng's River systems (Bobbins, 2015).

1.3 AIM AND OBJECTIVES

1.3.1 Aim

To identify the spatial extent of acid mine drainage from gold mining activities within Bloubankspruit and Crocodile River in Gauteng, South Africa.

1.3.2 Objectives

- To identify sample areas for data collection using a buffer analysis that considers how far downstream AMD effects reach.
- To screen for hazardous chemical elements within water resources, using XRF equipment at different proximity points from the mine.
- To identify the wider extent of contamination by training a Support Vector Machine learning model using the collected XRF points.

2. LITERATURE REVIEW

Acid mine drainage (AMD) occurs when waste from active or abandoned mines flows into a water system. The heavy metal composition of this waste pollutes the water that it flows into and alters the pH to be highly acidic (Dold, 2014; Skousen, Ziemkiewicz, & McDonald, 2019). AMD may contaminate surface water, as well as infiltrate ground water supply. Groundwater infiltration occurs when surface water seeps into soil and is absorbed by vegetation on the earth's surface. This process is highly dependent on the permeability of soil and the gradient angle of the earth's surface (Tomiyama & Igarashi, 2022). Water is able to move deeper into the earth's surface due to percolation, which describes the downward movement of water into a saturated zone due to gravity. Once water moves past this zone, it is able to infiltrate groundwater supply (Tomiyama & Igarashi, 2022).

A study by Acharya and Kharel (2020) review various coal mines within the United States of America. Groundwater infiltration due to waste from mining activity is studied in the States of in West Virginia, Pennsylvania, Ohio, and Indiana. This is done in order to understand the impact of AMD on communities and ecosystems. The mismanagement of mine waste has caused the degradation of drinking water in the previously mentioned states (Acharya & Kharel, 2020). This is due to the contamination of groundwater in the surrounding areas of mines. Furthermore, the seepage of the contaminants within soils is highly problematic as it can have harsh effects on the ecosystems that rely on these sources of water (Skousen, Ziemkiewicz, & McDonald, 2019; Tomiyama & Igarashi, 2022).

Hence, it is crucial to identify ways to mitigate and prevent the spread of AMD. AMD occurs when wastewater from mines comes into contact with sulphur-bearing minerals. These minerals are exposed from rocks as a result of mining activity. The chemical reaction between these minerals, water, and oxygen causes the release of high amounts of sulphuric acid into the water that flows into the surrounding rivers (Skousen, Ziemkiewicz, & McDonald, 2019). This increases the level of contamination that takes place. Furthermore, chemical and bacterial processes can cause other chemicals to be discharged from various types of rocks. Examples of these chemicals include iron, copper, lead, zinc, cadmium, arsenic, aluminium, and manganese (Skousen, Ziemkiewicz, & McDonald, 2019). The combination of acidity and metal toxicity accelerates the ecological degradation in affected river systems. importantly, the spatial variability of these chemical processes is influenced by factors such as mineralogy of exposed rocks, microbial activity, and hydrological flow. The understanding of these spatial

dynamics is essential for accurately mapping AMD. This motivates the use of advanced machine learning algorithms such as SVM to capture complex patterns in river systems.

Recent advancements in environmental research reveals that the traditional understanding of AMD currently underrepresents how complex the process of contamination is. For example, microbes, such as Acidithiobacillus ferrooxidans speed up the oxidation process which increases the severity of AMD (Zhan, et al., 2019). If multiple species of microbes are present within an environment, it increases the complexity of AMD and how acidic the river system becomes over time (Chen, et al., 2016). The pyrite that causes AMD also behaves differently based on the surrounding environmental characteristics. For example, chalcopyrite and arsenopyrite behave differently, which influences the quantity of acid that is formed. This depends on their structure and reactive properties (Chen, et al., 2016). It is therefore extremely important to acknowledge all environmental variables when studying AMD as the level of contamination is widely linked to the ecosystems and chemical activity per region.

Acid mine drainage (AMD) is therefore, a persistent problem within the province of Gauteng due to its large historic and current mining sector. The Gauteng City-Region Observatory (GCRO) reports on over 270 tailing waste sites in the 400 km^2 region of the Witwatersrand Basin alone (Khanyile, 2024). Coal and gold mines are specifically attributed to major effects in AMD and the deterioration of water quality according to Baloyi et al. (2023). The process of gold mining contains the exposure of large amounts of sulphide-bearing rocks, which inevitably becomes exposed to water and oxygen to create sulphuric acid (Lemos, et al., 2020). Coal naturally contains sulphur, such as pyrite within its structure. Much like with gold mining, the exposure of this to water and oxygen creates sulphuric acid which flows into rivers and is known as acid mine drainage (Acharya & Kharel, 2020).

Baloyi et al. (2023) notes that active mines are conscious of keeping water tables to a minimum and mitigating contamination through the use of pumps. As a result, more serious threats of AMD lie within abandoned and inactive mines. A study done in the Yudong River basin of Guizhou Province, China outlines that the lack of management of abandoned coal mines has resulted in an increase of ecological and environmental damage in comparison of operating mines in the area (Chen, Feng, & Liang, 2021). This is highly due to the fact that once these mines are no longer functioning, there are no effective operations in place to prevent the seepage of AMD from occurring. It is therefore imperative that decision makers implement measures to prevent harmful chemicals and elements from entering drainage networks. This is

because AMD has an effect on many different areas within the agricultural sector and biodiversity of a region.

For example, contaminated water has the potential to contaminate crops within the agriculture sector. This is because when acidic water and soil interact, the pH of the soil is lowered, which decreases the nutrient availability that plants rely on. Furthermore, plants ideally survive in neutral as opposed to acidic conditions, therefore, the fertility of the soil is decreased when exposed to AMD according to Baloyi et al. (2023). The heavy metals from AMD that leach into soils also have additional impacts to plants that are exposed to contamination. Heavy metals accumulate in plant tissues, which affects their growth and development over time (Skousen, Ziemkiewicz, & McDonald, 2019). Large quantities may even cause plants to die (Simate & Ndlovu, 2014). If crops are able to still be cultivated after exposure to AMD, it may travel down the food chain and affect the health and livelihoods of the people and animals relying on this resource (Liao, et al., 2016; Baloyi, et al., 2023). This means that human and animal health is at a high risk when in direct contact with water soil, plants, and groundwater that have had exposure to the high concentrations of heavy metals. A study by Fernández-Caliani et al. (2019) assesses the human health risks related to AMD contaminated soils around the mine site of Tharsis, Spain. This research identifies elevated amounts of lead of arsenic through crops in the area that are available for human consumption. Residents of the area were exposed to an average elevated amount of 27,7% and 42,6% of arsenic and lead respectively (Fernández-Caliani, Giráldez, & Barba-Brioso, 2019). The exposure of humans to the ingestion of arsenic has shown an increased risk of cancer. The risk decreased when exposed through dermal contact and the least amount of risk was associated with contact by inhalation (Fernández-Caliani, Giráldez, & Barba-Brioso, 2019). This pattern is repeated when analysing the extent of neurological and kidney damage associated with long term exposure to lead.

Compared to plant and human risks, aquatic life is the most impacted by AMD. This is because even a small exposure to heavy metals could kill organisms directly. A longer exposure could cause aquatic lift to mutate or stunt biological development (Talukdar, Kalita, Basumatary, Saikia, & Sarma, 2017; Skousen, Ziemkiewicz, & McDonald, 2019). Talukdar et al. (2017) discusses how the aquatic ecosystem is affected due to AMD caused by coal mining in the region of Meghalaya, India. The lower pH combined with the exposure of heavy metals damages the gills of fish, which directly relates to the respiratory function of most aquatic animals in the region. This lowers the survival rate of fish within the area and decreases the

biodiversity of the ecosystem, which causes an imbalance in the biological chain (Talukdar, Kalita, Basumatary, Saikia, & Sarma, 2017). Furthermore, heavy metals are able to accumulate within the tissue of organisms and animals that consume it and can be passed through the food chain. This process is known as bioaccumulation.

Below is a table that outlines the parameters that may affect water quality due to AMD, and the limits at which they are acceptable.

Table 1: Water contaminants and risk types as per the SANS 241 and WHO guidelines (World Health Organisation, 2011; South African Bureau of Standards, 2022)

Macro Parameters	Risk Type	SANS 241 Limit	WHO Limit
Ammonia (NH_3)	Aesthetic Operational	≤ 1,5 mg/L ≤ 0,3 mg/L	N/A
Chloride (Cl)	Aesthetic	≤ 250,0 mg/L	N/A
Fluoride (F)	Chronic health	≤ 1,5 mg/L	≤ 1,5 mg/L
Nitrate (N)	Acute health	≤ 11,0 mg/L	≤ 11,0 mg/L
Sodium (Na)	Aesthetic	≤ 200,0 mg/L	N/A
Sulphate (S)	Aesthetic	≤ 250,0 mg/L	N/A
Zinc (Zn)	Aesthetic	≤ 3000,0 µg/L	N/A
Micro Parameters			
Aluminium (Al)	Operational	≤ 200,0 µg/L	≤ 100,0 µg/L
Antimony (Sb)	Chronic health	≤ 20,0 µg/L	≤ 20,0 µg/L
Arsenic (As)	Chronic health	≤ 10,0 µg/L	≤ 10,0 µg/L
Barium (Ba)	Chronic health	≤ 1300,0 µg/L	≤ 700,0 µg/L
Boron (B)	Chronic health	≤ 2400,0 µg/L	≤ 2400,0 µg/L
Cadmium (Cd)	Chronic health	≤ 3,0 µg/L	≤ 3,0 µg/L
Chromium (Cr)	Chronic health	≤ 50,0 µg/L	≤ 50,0 µg/L
Copper (Cu)	Chronic health	≤ 2000,0 µg/L	≤ 2000,0 µg/L
Iron (Fe)	Aesthetic	≤ 300,0 µg/L	N/A
Lead (Pb)	Chronic health	≤ 10,0 µg/L	≤ 10,0 µg/L
Manganese (Mn)	Aesthetic	≤ 100,0 µg/L	≤ 100,0 µg/L
Mercury (Hg)	Chronic health	≤ 6,0 µg/L	≤ 6,0 µg/L
Nickel (Ni)	Chronic health	≤ 70,0 µg/L	≤ 70,0 µg/L
Selenium (Se)	Chronic health	≤ 40,0 µg/L	≤ 40,0 µg/L
Uranium (U)	Chronic health	≤ 30,0 µg/L	≤ 30,0 µg/L

*N/A is due to the conclusion of no health-based guideline being proposed

Besides the severe environmental risks that AMD proposes, it also sparks major public health concerns. This is an issue globally as toxic chemicals could lead to chronic illnesses. This could occur through contaminated drinking water, contaminated agriculture, or direct contact with the source of AMD polluted river systems. The exposure of AMD brings about many burdens

within the healthcare system. Table 1 above indicates that many parameters relating to contaminated water introduces chronic health problems within communities that rely on polluted water sources. The long-term exposure to certain chemicals such as arsenic can result in a higher risk of cancers within communities (Shakoor, et al., 2017). Additionally, arsenic can also bring about different liver issues, such as cirrhosis. When communities are exposed to both arsenic and cadmium, kidney disease and dysfunction may become apparent (Sankhla, Kumari, Nandan, Kumar, & Agrawal, 2016; Shakoor, et al., 2017). An increased risk of cardiovascular diseases has also been linked to arsenic exposure (Shakoor, et al., 2017). Similarly, prolonged exposure to lead results in risks of high blood pressure, which can also cause cardiovascular risks (Sankhla, Kumari, Nandan, Kumar, & Agrawal, 2016). Lead and mercury are also linked to neurological damage such as memory loss, cognitive deficits, and impaired motor skills within people that are exposed to these contaminants over longer periods of time (Sankhla, Kumari, Nandan, Kumar, & Agrawal, 2016). This is directly linked in developmental issues in children that live within these communities (Sankhla, Kumari, Nandan, Kumar, & Agrawal, 2016). Lastly, prolonged fluoride exposure may cause weakened bones that are prone to fractures, as well as bone deformities or skeletal fluorosis (Solanki, Agarwal, Gupta, Gupta, & Shukla, 2022).

The effects of AMD in turn, affects the economy of an area. This is because extensive funding will need to go towards healthcare and the remediation of contaminated water. The costs of AMD remediation are substantial according to a study by Naidu et al. (2019). This study identifies the top four countries with the highest AMD impact in the world. South Africa is third, with the United States and Canada coming in at first and second respectively. Although there are no figures relating to the AMD remediation costs in South Africa, remediation in the US amounts to 10 billion USD and environmental liability costs in Canada range between 1.5 to 3.5 billion USD (Naidu, et al., 2019). Hence, it can be concluded that the cost of AMD remediation in South Africa is considerably high.

A study by Botezan et al. (2020) reveals that after monitoring three mining areas in the Apuseni Region, Romania, mining activities have caused many socio-economic impacts. The remediation and waste management of the mines in this region has not been a priority for the government for a long time. Therefore, waste has seeped into soils and slowly migrated further away from the mines (Botezan, et al., 2020). This affects the economy of the region because the surrounding areas are now unable to engage in agricultural activity. The tourism of the area

is also non-existent as the environment is now barren due to the unhealthy exposure to contaminated water (Botezan, et al., 2020). The people residing around these areas were relying on employment opportunities from the mines that have since shut down. A study by Siyongwana and Shabalala (2019) identifies the impacts that mine closure has on a community in Pilgrim's Rest, Mpumalanga. Closure of this mine resulted in 27% of residents to relocate out of the community due to a decrease in employment opportunities. Furthermore, 63% of the local residents were left unemployed after the closure of the mine. This has led to high crime rates and an increase in alcohol and drug abuse within the community (Siyongwana & Shabalala, 2019).

AMD is a critical issue in South Africa specifically because of its rich mining history. The country has about 6100 abandoned mines with 2322 being classified as high risk in both AMD and safety of local communities (Mhlongo, 2023). This is due to the lack of management and remediation methods of these mines after closure. Large scale operations have been present in mining minerals such as gold, coal, and diamonds since the 1850s (Thompson, 2014). Many of these mining sites continue to expose the surrounding environment to heavy metals and sulphuric acid, even after 100 years. These contaminants continue to leach into the surrounding soils and water (Amponsah-Dacosta & Mhlongo, 2016). Because South Africa has an abundance of metals and minerals, rapid economic growth has occurred, and many more mines have opened up since. The mining sector in South Africa has an annual industry turnover of 1 trillion rand and makes up 6% of the country's GDP as of 2024 (Minerals Council South Africa, 2024). Although very basic and general information is available about mining in South Africa, there is a local research gap when it comes to specific case studies.

Many mines have closed since the 1850s and because there is little incentive to address the environmental impacts of pollution, closed mining sites still contribute to AMD decades after closure (Thompson, 2014). This is not the case in many large mining industries around the world. In China, AMD remediation policies have been implemented since the 1980s (Li, Jiao, Xiao, Chen, & Chang, 2015). It is important to note that China is the world's largest producer of gold and coal so AMD levels are extremely high around mining operations (Chen, et al., 2018; Liu, Bai, Shi, Zhou, & Liu, 2019). Pollution due to AMD is regularly monitored and different strategies and regulations are implemented over time to address high contamination levels. These regulations are implemented on both local and national scales. This supports the idea that contamination should be constantly monitored and that remediation efforts will

naturally take a long period of time to be implemented effectively (Li, Jiao, Xiao, Chen, & Chang, 2015). It is vital to mitigate pollution as many communities rely on water from AMD contaminated streams.

Historically active and currently active mines are in close proximity to communities of lower socio-economic status (Edwards, et al., 2014; Thompson, 2014). A study taking place in rural communities in Limpopo surveys how people living in the community overcomes the issue of contaminated water in the area. 28,2% of residents within the area perceived that consuming contaminated water has severe health risks. The identified risks include malnutrition, cholera, fever, and vomiting among other symptoms. To overcome this, residents keep containers such as buckets full of water from distribution tanks for domestic use. This prevents health risks from consumption, however, risks through physical contact or inhalation are not addressed. This group is the most vulnerable to the severe health problems that AMD causes. They do not have adequate access to health care, nor do they have access to education that might inform them of the impacts of AMD. Therefore, it is important to identify methods of mitigating the further exposure of communities to the long-term effects of AMD (Edwards, et al., 2014; Thompson, 2014).

AMD contamination can be examined through traditional methods such as pH testing, monitoring metal concentrations, and conducting visual surveys (Martín-Crespo, et al., 2018). When AMD occurs, the pH of the contaminated water drops significantly. It becomes more acidic particularly sue to the oxidation of sulphide minerals. Monitoring this can be done through the use of portable pH meters or pH testing kits (Martín-Crespo, et al., 2018; Oberholster, Schoeman, Truter, & Botha, 2022). Metal concentrations are monitored by extensive lab testing when examining contaminated water. The monitoring of metal concentrations is important because it allows for a better understanding of the extent of AMD (Martín-Crespo, et al., 2018). This also allows for a better understanding of the extent of damage that communities and ecosystems face due to the associated health risks. Visual surveys are also important when examining the extent of AMD because contaminated water changes significantly in colour and odour. This helps in identifying where contamination is present, however, quantitative data is necessary in supporting these observations to conclude the accuracy of the level of contamination (Martín-Crespo, et al., 2018; Oberholster, Schoeman, Truter, & Botha, 2022). Additionally, there is no presence of fish or other aquatic life in AMD contaminated water due to the toxic and acidic properties that make the

environment inhabitable (Martín-Crespo, et al., 2018). A visual survey could therefore indicate the first signs of contamination.

Apart from water testing, soil and sediment sampling is also important when examining the effects of AMD. This is because of the way that contaminated water infiltrates the immediate soil that is in contact with contaminated water (Oberholster, Schoeman, Truter, & Botha, 2022). The soil that is in immediate contact with contaminated water may also provide a more accurate long-term analysis of contamination than surface water does. This is because contaminants are deposited and accumulated from water to soil over time (Liao, et al., 2016). Sediment cores and soil profiles are able to determine the amounts of contamination that has been historically present within soil samples. This is because each layer of soil is able to record metal accumulation through an analysis of the stratigraphy, or vertical profile of the sample. This method is effective as it is able to display contamination throughout decades and centuries (Veerasingam, Vethamony, Murali, & Fernandes, 2015). There are a few ways in which soil can be tested for heavy metal contamination. The first way would be by grab sampling. This method requires collecting a surface level sample to take back to a controlled environment for testing (Vicente-Beckett, Taylor McCauley, & Duivenvoorden, 2016). A study by Vicente-Becket et al. (2016) utilises this method when reviewing soils associated with AMD in Queensland, Australia. This study notes that it is a quick, easy, and efficient way of concluding surface level contamination at a particular point, however, it is important to note that contamination levels vary in depth, so this method may not be entirely accurate (Vicente-Beckett, Taylor McCauley, & Duivenvoorden, 2016). Furthermore, this method may not be ideal for examining and mapping larger areas. Because of this, they then extract core data per sample point to accurately analyse how far downstream AMD travels. This is also effective in identifying how long contamination has occurred over the region.

The next method is a process of composite sampling, where multiple samples are taken over an area, and is then mixed to form one wholistic representation of the area (Bartram, et al., 2014). This gives an overall idea of the level of contamination in a region; however, it does not acknowledge where different contamination hotspots may be. It also does not acknowledge where contamination is very low within the area of study (Bartram, et al., 2014). These concerns may be solved by means of point sampling. This kind of sampling requires soil to be examined randomly at different points within an area (Heung, et al., 2016). This allows for spatial variability to be identified which provides insight into which points contain higher and

lower levels of contamination (Heung, et al., 2016). Because of the random selection of sample points, the concentration gradients may not be reflected accurately. This could lead to uneven data distribution (Heung, et al., 2016). The most effective method of soil sampling in this case, would be the use of incremental sampling. This method of sampling examines soil at predetermined intervals in order to create a composite of soil contamination along an area (Malone, Minasny, Odgers, & McBratney, 2014; Yang, Christakos, Guo, Xiao, & Huang, 2017). It is advantageous because it provides a more systematic and accurate analysis of the study area. The disadvantages are that it is very time consuming and uses very expensive and advanced equipment for collection (Yang, Christakos, Guo, Xiao, & Huang, 2017). In order to manage data outliers and errors, multiple readings should be taken at each sample point.

The mobility of heavy metals through soil is important in understanding the extent of long-term impacts of AMD within the environment. The mobility and leaching of heavy metals in soil increases when the pH of the soil is more acidic (Komonweeraket, Cetin, Aydilek, Benson, & Edil, 2015). The type of the soil also influences the retention quality of heavy metals within soil. Clay rich soils usually have high cation exchange capacities that encourage the storing of heavy metals within the soil. They also prevent the leaching of heavy metals into water and vegetation (Jiang, Wang, Yu, Cao, & Yan, 2018). The presence of organic matter content would further reduce metal leaching through retaining metals by binding them in organic complexes (Jiang, Wang, Yu, Cao, & Yan, 2018). Toxic metals have the ability to leach important minerals from soil where an environment includes high pH levels, low cation exchange capacities, and low organic matter content. These metals include as iron, copper, zinc, lead, and cadmium (Jiang, Wang, Yu, Cao, & Yan, 2018). This has direct severe consequences on the health of ecological and agricultural activity that relies on the soil. A study by Alengebawy et al. (2021) has established that heavy metal contamination in agriculture contributes to decreased crop yields. This is because of the ability that metals like cadmium have to absorb organic matter and nutrients from soil, which negatively impacts the growth and development of plants (Alengebawy, Abdelkhalek, Qureshi, & Wang, 2021). Lead could also cause DNA damage and hinder seed germination processes. Additionally, zinc blocks the transport of nutrients within the plant and decreases the ability of photosynthesis (Alengebawy, Abdelkhalek, Qureshi, & Wang, 2021).

Apart from chemical changes, AMD also has the ability to alter the physical structure of soil it comes into contact with. Contaminated soil generally has a lower pH and may lead to the

compression and poor aggregation of soils. This would mean that less aeration and filtration occur which leads to a lower fertility rate in the soil (Delgado & Gomez, 2017). This causes agriculture to struggle to grow. A study by Delgado and Gomez (2017) explains that soil management is extremely crucial in ensuring that soils hold the same properties when they are being used for important activities, such as agriculture. This is because the level of infiltration, aeration, retention, and trafficability is highly dependent on the physical composition of soil. Soil composition is also dependent on the chemicals that are in contact with the soil (Delgado & Gomez, 2017). Aeration describes the oxygen content that is available within soils, which is important for plant root respiration (Ben-Noah & Friedman, 2018; Xin, Tang, Liu, Zhang, & Tian, 2019). The infiltration of metals such as iron or aluminium within soil decreases the amount of oxygen that plants have access to. This has the capacity to clog pores and cause waterlogging within soils. This is detrimental to the growth of plants due to the restriction of oxygen. As a result, plant roots are damaged and activity for necessary soil organisms is reduced (Xin, Tang, Liu, Zhang, & Tian, 2019).

The occurrence of heavy metals in soil also reduces the capability of soils to hold and retain water and nutrients. This is directly linked to the cation exchange capacity of soils. Metals such as zinc, cadmium, and lead are able to compete with essential nutrient cations such as calcium, magnesium, and potassium (Van Groeningen, Christl, & Kretzschmar, 2021; Nachana'a Timothy, 2019). The trafficability of soils relate to the ability of soils to withstand compaction from surface activities (Wieder & Shoop, 2018). The presence of metals within soils reduces trafficability through the formation of crusts on the soils surface (Wieder & Shoop, 2018). These make soils much more prone to compaction and lowers soil porosity and drainage capabilities. The aeration of soils is also lowered by the formation of crusts. This makes it extremely difficult for plant to develop and survive (Wieder & Shoop, 2018). Long term exposure to heavy metals has the potential to condense and accumulate soil and sediment particles together so that soil becomes more of a clay. This affects the ecosystems that rely on the soil as the conditions have now become uninhabitable (Delgado & Gomez, 2017).

The topography of a region largely influences the effect that AMD has on its surrounding environment. The elevation of a region may provide insight into where toxic waste is able to migrate due to water flowing downstream. Areas that contain a steeper slope gradient will have lower contamination levels in comparison to areas of a gentler or flatter gradient. This is because steeper slopes have higher runoff rates. Alternatively, flatter land is attributed to water

pooling as there is nowhere for water to move (Galván, Olías, Cánovas, Sarmiento, & Nieto, 2016). This causes an accumulation of contaminants within soils of flatter land. This is helpful in predicting where AMD would be most present, which is easily done when creating a digital elevation model (DEM). This is analysed through identifying the difference between a high and low lying area and the distance between the two (Galván, Olías, Cánovas, Sarmiento, & Nieto, 2016). If there is a high difference between a short distance, the slope gradient is high, and runoff of contamination is more likely. A low difference between a long distance means that it is likely that contamination will be accumulated (Galván, Olías, Cánovas, Sarmiento, & Nieto, 2016).

It is important to note that the satellite data availability also presents challenges for the quality of the DEM that is produced. While multispectral imagery from platforms like Sentinel-2 is generally sufficient for vegetation and surface water analysis, its effectiveness is limited by cases such as cloud cover, sensor noise, and revisit frequency (Modiegi, Rampedi, & Tesfamichael, 2020). This is most relevant in persistently cloudy seasons or during rapid environmental change. This is particularly problematic in areas downstream of mine dumps where AMD-related features like stressed vegetation may be transient or obscured (Modiegi, Rampedi, & Tesfamichael, 2020). Without high-frequency cloud-free imagery, temporal validation of AMD dynamics becomes difficult, limiting the ability to distinguish between persistent contamination and temporary false positives (Modiegi, Rampedi, & Tesfamichael, 2020).

Satellite spectral resolution also restricts the ability to detect some AMD signatures. For example, Sentinel-2 lacks shortwave infrared (SWIR) bands. This prevents the sensors from identifying certain mineralogical changes that are associated with AMD (Modiegi, Rampedi, & Tesfamichael, 2020). Hyperspectral sensors could address this limitation, but are largely inaccessible due to cost, limited coverage, or data licensing restrictions. As a result, there is an imbalance between data accessibility and spectral specificity, which limits the precision of AMD detection (Modiegi, Rampedi, & Tesfamichael, 2020).

A study taking place in Southern Brazil specifically notes that during rainfall, heavy metals are transported downstream from a coal mine very rapidly. This process allows for contamination to occur further away from the source and promotes soil nutrient leaching. It also expands the spatial scope of the ecological impacts (Galhardi & Bonotto, 2016). Contaminants may also stay present within the environment over long periods of time. Therefore, it is imperative to

evaluate the impact of AMD during dry seasons so that rainfall does not dilute the results. This would mean that a more accurate outcome is produced on how heavy metals are transported downstream. Seasonal flooding influences both the spread and persistence of contamination over an area which could result in either long term or short term AMD damage (Galhardi & Bonotto, 2016; González, Cánovas, Olías, & Macías, 2020).

Expanding on the study above, AMD severity varies considerably under different climatic and geological contexts. In Gauteng, the geology is highly dominated by rocks that are rich in pyrite and uranium. This directly increases the potential of the amount of acid that is created when exposed to oxygen and water (Minnaar, 2020). Furthermore, the climate of Gauteng is classified as semi-arid and has distinct wet and dry season. During the rainy season, there is an increase in water infiltration which promotes sulphide oxidation and metal leaching. The dry season on the other hand, amplifies water evaporation and the concentration of solutes. Therefore, the interaction of geology and climate should be fully considered in order to understand AMD in the geographic context of Gauteng (Minnaar, 2020).

Gauteng has an extremely high presence of historical and currently functioning gold mines. These types of mines are seen as main contributors of AMD and uranium issues. The other mine types, such as coal or non-metallic mines, do not have a significant presence within the province. The AMD risk associated with these mines are also medium to low risk (Minnaar, 2020).

Flash flooding is able to cause short term AMD contamination within a downstream area. This is due to the process of flushing contaminants from already affected upstream soil. After floodwaters recede, however, the system may return to its usual uncontaminated state (González, Cánovas, Olías, & Macías, 2020). This is because the presence of AMD contamination is not persistent within that system (González, Cánovas, Olías, & Macías, 2020). Alternatively, in events of persistent flooding or continuous heavy rainfall, AMD will become severe in downstream areas. This is due to the persistent exposure of the contaminants within soils (González, Cánovas, Olías, & Macías, 2020).

Myers (2016) maps the transportation of AMD downstream within a mine facility in Minnesota, USA. Through doing this, streamflow and elevation is taken into account in order to accurately map out the extent and severity of AMD in the correct direction. Hence, it is important to incorporate elevation data when reviewing the transportation of AMD. Myers

further emphasises the same argument as Galhardi and Bonotto (2016) where they note that heavy metals that stem from AMD have the ability to bind and settle in sediments. They may also remain dissolved in water for extended periods of time. This means that they can have long term impacts on the surface and ground water of a region (Myers, 2016). Migration over time could occur through the slow movement of ground water, biological consumption by plants and animals, or hydrological transport (Myers, 2016).

One way of examining soil samples is through the use of X-ray fluorescence (XRF). This method can be used with incremental sampling and is effective in analysing the soil samples. This is because of its non-destructive nature to the environment and its ability to identify trace elements within soil and sediment (Mejía-Piña, Huerta-Diaz, & González-Yajimovich, 2016; Declercq, et al., 2019). XRF utilises high-energy x-ray beams in order to measure the concentration of elements of the surface that the XRF is in contact with. The XRF measures 1 to 2cm beneath the surface being examined (Young, Evans, Hodges, Bleacher, & Graff, 2016; Declercq, et al., 2019; Potts & Sargent, 2022). It works particularly well with solid particles such as rock, soil, and sediments, however, using this method to measure water quality is not effective as elements are constantly moving within a liquid (Young, Evans, Hodges, Bleacher, & Graff, 2016; Potts & Sargent, 2022). XRF technology is able to measure light and heavy metals, as well as trace concentrations within a sample. This is because of its high sensitivity (Declercq, et al., 2019). Further advantages of XRF include the rapid analysis of XRF technology.

Data can be available within seconds at the sample point (Potts & Sargent, 2022). Furthermore, sampling can be done in situ. This means that extraction does not need to take place for laboratory analysis because results are available instantaneously (Declercq, et al., 2019; Potts & Sargent, 2022). This also means that this device is not destructive or disruptive to the environment in question. It is still however, important to note that depth sampling is important. This is because it provides an indication of how contaminants migrate over time (Malone, Minasny, Odgers, & McBratney, 2014). Analysis for depth sampling is done by extracting cores from different points to create a vertical profile of contamination (Malone, Minasny, Odgers, & McBratney, 2014). A challenge with XRF technology includes the fact that environmental influences may influence the accuracy of the readings (Declercq, et al., 2019). These influences include sediment grain size, soil type, or moisture content within soils. In

order to mitigate this, a few samples should be taken at a point in order to identify which readings are false or erroneous.

According to a study done by Yang et al. (2017), readings were done 5 times over a sample point within a 100m radius of the water contamination site. The process of re-taking readings have been proven to prevent errors in data from being included in the research for AMD. therefore, having more data available can increase the accuracy of the research being done (Yang, Christakos, Guo, Xiao, & Huang, 2017).

It is important to explore the environmental behaviour of metals such as arsenic, cobalt, chromium, copper, manganese, nickel, lead, sulphur, and zinc. This is because they are highly variable and are directly influenced by both their mobility in the soil-water system and their detectability using techniques such as X-ray fluorescence (XRF). A failure to consider these differences can lead to misinterpretation of contamination extent or severity in AMD studies.

Metal mobility is largely governed by factors such as pH, soil texture, and organic matter content. For instance, arsenic tends to adsorb strongly to iron oxides under oxidising conditions but becomes more mobile under waterlogged conditions (Aguilar-Carrillo, Velázquez, Gutiérrez, & Reyes-Domínguez, 2022). This can result in vertical leaching beyond the initial contamination site, reducing the reliability of surface based XRF readings. On the other hand, chromium tends to bind tightly to soil particles. Chromium is soluble and migrates readily in acidic and oxidising environments, which is common in AMD impacted zones (Pytlakowska, 2016). This poses a challenge for accurate surface detection.

In contrast, copper and zinc are moderately mobile in acidic soils, which are typical of AMD environments. Both metals are more likely to remain in the upper soil horizons and are thus more reliably detected by surface XRF methods. However, their bioavailability and retention can vary significantly depending on clay content and cation exchange capacity (Tighe, et al., 2018; Aguilar-Carrillo, Velázquez, Gutiérrez, & Reyes-Domínguez, 2022). Nickel and cobalt exhibit intermediate mobility and are influenced by pH conditions (Siebecker, Chaney, & Sparks, 2018). Cobalt, in particular, tends to be more mobile under waterlogged conditions, which may lead to deeper migration and lower detectability at the surface (Turcotte, Jamieson, & Parsons, 2025).

Lead is well known for its low mobility in most soil types due to strong absorption to organic matter and mineral surfaces. It accumulates near the surface and persist over long periods,

making it highly responsive to XRF detection (Aguilar-Carrillo, Velázquez, Gutiérrez, & Reyes-Domínguez, 2022). However, its immobility also means it may not indicate current AMD activity but rather older contamination (Aguilar-Carrillo, Velázquez, Gutiérrez, & Reyes-Domínguez, 2022). Manganese is redox-sensitive and can become highly mobile under waterlogged conditions or in acidic soils. As such, hotspots can indicate recent AMD-driven water movement, but concentrations may fluctuate seasonally (Jensen, 2015).

While sulphur is a key indicator of AMD due to pyrite oxidation, XRF detection is limited to solid-phase sulphur and is not sensitive to dissolved sulphate in soil moisture or groundwater (Vöröš, et al., 2018). Therefore, reliance on XRF alone may underestimate sulphur content in areas where leaching has occurred.

These differences in mobility and persistence have critical implications for XRF-based spatial mapping. Elements like lead and copper, which are more stable and surface-retained, provide reliable indicators of historic deposition. In contrast, arsenic, chromium, cobalt, and manganese may migrate below the detection range of surface XRF. Therefore, in this study, elemental results are interpreted in the context of known AMD geochemistry and supported by spatial proxies such as terrain, proximity to tailings, and vegetation stress to avoid over-reliance on concentration data alone (Vöröš, et al., 2018).

A study done by Raji et al. (2021) reviews uranium pollution levels within the Rietspruit River in South Africa. This study follows the river for about 60 km downstream until reaching the Vaal River. The concentration of uranium in the most densely contaminated part of the river remains constant for about 2 km before a notable reduction is seen. Therefore, contamination should be assessed at an initial 2 km interval from the mine to understand the level of contamination that is produced by a mine. Even though the mine being examined has been decommissioned, the erosion of the dump still contributes to waste entering water systems. This causes health risks to people and environments that are dependent of these water bodies (Botezan, et al., 2020; Raji, Hoffmann, Ngie, & Winde, 2021). For example, if the river were to stop flowing during dry months, contamination could potentially affect various different ecosystems such as wetlands. Furthermore, farming, crop irrigation, animal watering, and religious activities in the area all make use of the uranium contaminated water. This increases the presence of chronic health illnesses over the area (Botezan, et al., 2020; Raji, Hoffmann, Ngie, & Winde, 2021).

It is important to cross reference heavy metal levels with local and international standards in order to quantify the level of contamination of AMD. The national and international thresholds for elements related in water contamination is depicted in Table 1 above. These standards are taken from official documents by the South African Bureau of Standards and the World Health Organisation (World Health Organisation, 2011; South African Bureau of Standards, 2022). Additionally, the Department of Environmental Affairs in South Africa has identified element thresholds that indicate soil contamination. Soil screening values (SSV) vary depending on whether the water in contact with the soil is used for human consumption or protecting the ecosystem health (SSV1). This is seen in the table below. Values also vary within different environments, such as residential and industrial environments (SSV2). It is important to note that SSV1 does not account for different types of land use and SSV2 does not account for the health of humans and the ecosystem (Department of Environmental Affairs, Republic of South Africa, 2010). The values that are used as the relevant thresholds for AMD, therefore relies on the surrounding environment of the sampling sites. This can be determined by examining the on-site properties of the environment as well as the LULC map of the study region.

Table 2: Soil Screening Values (SSV) for different LULC types

Parameter	**Units**	**SSV1**		**SSV2**	
		All Land-Uses Protective of the Water Resource	**Informal Residential**	**Standard Residential**	**Commercial/ Industrial**
Arsenic	mg/kg	5,80	20,00	45,00	150,00
Cobalt	mg/kg	6,00	300,00	640,00	5000,00
Chromium	mg/kg	6,00	6,00	13,00	40,00
Copper	mg/kg	16,00	1140,00	2400,00	19500,00
Lead	mg/kg	20,00	110,00	230,00	1900,00
Manganese	mg/kg	245,00	740,00	1550,00	12750,00
Nickel	mg/kg	620,00	620,00	1290,00	10600,00

Sulphur	mg/kg	4000			
Zinc	mg/kg	2000,00	9300,00	19400,00	160000,00

Because many elements are being tested for in the river system, it is important to identify which of the tested elements hold the most importance in the contamination of drainage basins. This provides more practical insights as to how severely contaminated the water is. A study by Patel et al. (2023) makes use of various different parameters present within a river system to measure water quality in the south of the Gujarat region, India through Artificial Neural Network (ANN) models. This study uses a Weight Arithmetic Water Quality Index (WAWQI) method, modified from Horton's formula, to assess the water quality of the river.

A study by Betrie et al. (2013) compares the use of artificial neural networks (ANN), support vector machine (SVM), k-nearest neighbour (KNN), and M5P model trees in the prediction of copper concentrations in acid mine drainage. It is concluded that SVM contains the highest accuracy of prediction. This is further validated by a study in Indonesia that compares the accuracy of KNN and SVM in testing drinking water quality (Danades, Pratama, Anggraini, & Anggriani, 2016). The results of this conclude that SVM is most effective with a 92,4% accuracy rate and 12/15 data set match. Hence, the use of SVM would be most effective in classifying the levels of contamination within Gauteng's River systems.

Over the past two decades, machine learning has become a valuable tool in environmental science, especially for modelling complex spatial processes like heavy metal contamination, AMD, and land degradation (Hsieh, 2022). Among supervised learning algorithms, SVM is widely used due to its ability to handle high-dimensional, non-linear, and imbalanced datasets. SVMs classify data by constructing an optimal hyperplane and are well-suited for modelling complex environmental phenomena (Hsieh, 2022). In order to implement SVM, preprocessing is a critical step in spatial machine learning workflows. It typically involves collating DEMs, satellite indices, and geochemical data into a unified raster grid, followed by data normalization, balancing, and outlier detection (Zhong, et al., 2021). Such steps allow machine learning models to generalize spatial patterns, especially in processes influenced by topography or runoff.

In addition to SVM, other classifiers like Random Forest (RF), k-Nearest Neighbours (KNN), and Logistic Regression are also beneficial. RF handles non-linearity well but may overfit and

lacks interpretability in high-dimensional contexts. KNN is intuitive but sensitive to sampling bias, and Logistic Regression, while transparent, struggles with complex interactions (A. Ramezan, A. Warner, & E. Maxwell, 2019; Sharma, Saharia, & Ramana, 2024). Comparative studies often find that SVM strikes a good balance between accuracy and robustness. The rise of geospatial machine learning platforms has enhanced the scalability of these methods. Tools like Google Earth Engine, ArcGIS Pro, and open-source environments such as Python's scikit-learn and R allow for efficient processing of satellite and terrain data (Bezabih, Glaety, Wako, & Worku, 2024; Rahimoon, et al., 2025). Despite this, relatively few studies have integrated field-based XRF data into machine learning workflows.

While machine learning and geochemical methods are widely used internationally for environmental assessment, few have applied them specifically to legacy AMD from gold mining in urbanised and geologically complex regions like Gauteng, South Africa. This study addresses that gap by combining XRF-based soil chemistry with remote sensing and terrain data to develop a model of AMD contamination.

Examples of practical uses of machine learning may include a study by Sakala et al. (2019) in South Africa. This study demonstrated the use of a deep artificial neural network (ANN) to assess groundwater vulnerability to sulphate contamination in the Witbank coalfield. The presence of sulphur is a key indicator of AMD due to the way that AMD forms (Sakala, Fourie, Gomo, & Coetzee, 2019). Their model incorporated hydrogeological parameters from the DRIST framework and predicted sulphate concentrations at borehole locations with high accuracy. The DRIST framework is specific to analysing depth to water level, recharge, impact of the vadose zone, soils and topographic slope for AMD research (Sakala, Fourie, Gomo, & Coetzee, 2019). Compared to traditional overlay methods, the ANN model provided more detailed spatial differentiation of AMD risk, highlighting the advantages of non-linear machine learning algorithms in capturing complex environmental interactions. The methodology is especially relevant for the Gauteng goldfields, where legacy mining infrastructure has significantly altered surface and groundwater dynamics (Sakala, Fourie, Gomo, & Coetzee, 2019).

Similarly, Wei et al. (2019) explored the estimation of heavy metal contamination in soils using hyperspectral imagery in combination with Light Gradient Boosting Machines (LightGBM). Their study introduced a soil specific modelling approach that recognizes the heterogeneity in spectral responses across different soil classes (Wei, et al., 2019). The models demonstrated

strong predictive performance, particularly for arsenic and copper, achieving R-squared values above 0,70 in validation datasets. The authors emphasized that accounting for local variability in soil chemistry can significantly enhance ML model performance. These insights are particularly pertinent for the gold mining belt of Gauteng, where complex mineral collections introduce additional modelling challenges (Wei, et al., 2019).

In a more directly relevant South African context, Mogashane et al. (2023) conducted a bibliometric and thematic review of AMD-related research in South Africa between 1998 and 2022. Their analysis revealed a dominant focus on chemical treatment technologies, while monitoring and prediction remain underrepresented (Mogashane, Maree, Mokoena, & Tshilongo, 2025). The authors identified a critical need for integrated approaches that combine remote sensing, ground-based measurements, and data-driven modelling to improve the detection and prediction of AMD events. This underscores the contribution of the present study, which leverages machine learning and satellite imagery to map the spatial presence of AMD associated with gold mining in Gauteng (Mogashane, Maree, Mokoena, & Tshilongo, 2025).

Spatial analysis is crucial when reviewing contamination over large distances, and especially when there are many factors that need to be integrated into the analysis. These factors may include types of pollutants, hotspots, land use and land cover data, elevation, and the proximity of points to contamination sources (Worlanyo & Jiangfeng, 2021). Understanding factors like these assists in creating an understanding of the uneven distribution of AMD and contamination. For example, a higher level of contamination may be present in rivers that run through agricultural areas (Obi, Ugwuishiwu, & Nwakaire, 2016; Karaouzas, 2018). This may be because of the runoff of different chemicals and pesticides from farming activities. Furthermore, high levels of contamination could be located nearby industrial areas as pollutants are usually produced by these activities (Karaouzas, 2018). It is also important to identify slope properties to understand the path of AMD contamination and where runoff or pooling is apparent. This creates an understanding of how contamination varies and where concentrations are higher in order to establish remediation strategies.

Common remediation strategies may be implemented after it is revealed that a large amount of AMD is present within an area. Methods for remediation include neutralization with lime, bioremediation, and passive treatment systems. The neutralisation of AMD with lime is one of the most common methods of remediation. It involves the introduction of lime or limestone

into contaminated water in order to reduce the high amounts of acidity in the water (RoyChowdhury, Sarkar, & Datta, 2015; Rambabu, et al., 2020). This is because when sulphuric acid reacts with lime, natural salts are introduced into system. This raises the pH to become less acidic and more alkaline (RoyChowdhury, Sarkar, & Datta, 2015). Although this method is widely used and works very rapidly, there is a high cost associated with it. This method also creates a lot of sludge over time and therefore requires a lot of maintenance (RoyChowdhury, Sarkar, & Datta, 2015). The process of bioremediation relies on the use of micro-organisms and plants to reduce the toxicity of the contaminants. This process transforms contaminants into less harmful substances (Rambabu, et al., 2020). For example, certain bacteria are able to oxidise both iron and sulphur. This makes the metals less harmful and increases the pH of the water, making it less acidic (Rambabu, et al., 2020).

Some plants also absorb heavy metals from contaminated water and are able to store these metals in their tissues. This is seen in wetland species such as reeds and cattails (Naidu, et al., 2019; Rambabu, et al., 2020). These plants absorb metal contents from water and soils and store the harmful chemicals within their roots or stems (Guo & Cutright, 2017). If metals are persistent over time, this could lead to the process of bioaccumulation within the plants (Guo & Cutright, 2017; Nguyen, Soda, Kanayama, & Hamai, 2021). Additionally, surface absorption could also take place if wetland plant species contain porous root surfaces. The likelihood of absorption increases in acidic environments due to the solubility of metals such as arsenic and lead (Guo & Cutright, 2017). A high amount of organic content also increases the likelihood of metal absorption within wetland plants. This is because organic matter has the capability to bind with metals and increase their availability to plants (Nguyen, Soda, Kanayama, & Hamai, 2021).

Apart from the environmentally friendly aspect of this method, the process of bioremediation is also cost effective and may be applied to both small- and large-scale operations. Unfortunately, the process of bioremediation takes a long period of time occur. This process would therefore be unsuitable for sites that need immediate intervention (Rambabu, et al., 2020). Furthermore, environmental factors such as water temperature, pH, nutrient availability, and climate change may affect the effectiveness of this method. Different plants and organisms may also be unavailable at different environments. Therefore, extensive research will need to be done per location (Naidu, et al., 2019; Rambabu, et al., 2020).

Passive treatment systems utilise the natural surrounding environment in order to remediate contaminated water systems. An example of this is constructed wetlands, which mimics the natural process of filtration that wetlands use. This system filters out contaminants and neutralises acidity by means of plants, bacteria, and organic matter (RoyChowdhury, Sarkar, & Datta, 2015; Naidu, et al., 2019). Another example is the introduction of vertical flow systems. this system allows water to flow through a sequence of layers that promotes the neutralization of water. These layers include limestone, microbial populations, and organic material. Neutralisation by these layers occur through chemical reactions and biological treatment (RoyChowdhury, Sarkar, & Datta, 2015; Naidu, et al., 2019). Methods like these are sustainable, cost effective, and are highly effective over a long period of time. However, these processes are also slow to establish, vulnerable to failure due to poor maintenance, and only work to treat small amounts of AMD at a time (RoyChowdhury, Sarkar, & Datta, 2015).

3. MATERIALS AND METHODS

3.1 STUDY AREA

The Gauteng region contains a very high concentration of gold mining activities. These mines are concentrated over the Witwatersrand basin, towards central and southern Gauteng. The mines in this region have been present since the 1880s (Bobbins, 2015). Because of the dense concentration of mining activities, AMD contamination is highly apparent within this region (Bobbins, 2015; Baloyi, et al., 2024). This contamination is even more apparent when considering the migration through the various amounts of rivers and streams within this area. AMD specifically impacts the Blesbokspruit, Klip River system, Wonderfontein, Natalspruit, Tweelopiespruit, and Hartebeespoort Dam. The Vaal River system is at the highest risk as AMD has affected this system for over a century, since the 1900s (Bobbins, 2015). Because of the high concentration of rivers in close proximity to the mines, the effects of AMD are able to be carried downstream and impact a much larger area of land (Steyn, et al., 2019).

In order to isolate the effects of AMD, a more remote area needs to be analysed. The Bloubankspruit and Crocodile River flows within two kilometres of abandoned gold mining activity. This means that tailings run directly through this area. The Bloubankspruit and Crocodile River were selected due to their close proximity to historically operating gold mining areas on the West Rand. These mines have a well-documented exposure to AMD (Mvandaba, Kapangaziwiri, & Shadung, 2019). The Bloubankspruit in particular has been identified as one of the most visibly impacted streams, with highly acidic discharges emerging from the nearby western basin (Mvandaba, Kapangaziwiri, & Shadung, 2019). The western basin is a known AMD hotspot. The Crocodile River, which receives flow from several tributaries, also has elevated concentrations of sulphates and heavy metals (Mvandaba, Kapangaziwiri, & Shadung, 2019). This is because the tributaries that flow into it is deeply affected by AMD. These systems are characterised by their ecological vulnerability and historical failure of containing AMD (Hobbs, 2017). Both of these systems therefore represent critical case studies within Gauteng and is a good case study to evaluate.

Seeing as this area has minimal intervention from other mining activity, the extent of AMD from gold mining activities can be accurately measured (Kgotso, 2018). To further support this point, the surrounding areas of this river are cultivated activities, shrubland, grassland and forested land. There is no evidence of heavy industrial activity taking place along the river, or

large amounts of waste flowing into the river, except for the waste originating from AMD (Kgotso, 2018). Although there are no contaminants flowing into the river, the contaminated river water does flow out of the river into other areas. Both the Bloubankspruit and Crocodile River systems flow through areas where local communities rely on surface water for domestic use, irrigation, and livestock farming (Hobbs, 2017). The Crocodile River, in particular, feeds into parts of the Hartbeespoort Dam catchment, which supports downstream agricultural production and recreation (Hobbs, 2017). Contaminated runoff entering these systems threatens not only aquatic biodiversity but also crop viability and human health through heavy metal uptake in irrigated soils (Alvarenga, Guerreiro, Simões, Imaginário, & Palma, 2021). Settlements such as Carletonville, Krugersdorp, and downstream farming communities are at risk of long-term exposure due to AMD infiltration into water supplies (Hobbs, 2017). Despite ongoing remediation efforts, episodes of acid and metal pollution continue to be reported in these catchments, pointing to the need for spatially resolved monitoring tools to guide environmental management and public health interventions (Ighalo, et al., 2022).

The rock type of this area directly relates to the level of contamination. This is because the underlying rocks of the Witwatersrand basin contains mainly sedimentary and volcanic rocks (Curtis, 2021). These formations are rich in pyrite, which is a primary source of AMD when exposed to water and oxygen (Curtis, 2021).

The mine that is being evaluated as the source of AMD is the Kromdraai gold mine. This mine has been inactive since 1914; however, it is considered one of the most productive mines in its active years. These years span between 1885 and 1894 (Park, 2014). The site is now a tourist attraction. Although this gold mine has been closed for over 100 years, it still contributes to a high amount of AMD. This is due to the fact that tailings are still being discharged from the mine (Park, 2014). The idea that waste is so prominent after over a century of the mine being closed means that poor waste management is present within the area. It provides an understanding of how the inactive status of mines does not mean that contamination is not occurring anymore.

Only perennial rivers were reviewed when deciding on a study site. This is because data is collected during dry months. This minimises the impact that rainfall may have on transporting the contaminants downstream. Seeing as the region follows a tropical climate, rainfall is prominent within summer months. Furthermore, non-perennial rivers have a slightly different impact on the ecology of the environment. This is because contaminants are more severely

present within soil when the river dries up (Mosley, Fitzpatrick, Palmer, Leyden, & Shand, 2014). AMD within Gauteng region is also mostly present within permanent rivers, hence, a more wholistic understanding of waste in river systems can be concluded through this study.

The Bloubankspruit flows into the Crocodile River, which eventually flows into the Hartbeespoort Dam. This means that the river flows in a north easterly direction. Data should therefore be collected from the point of contamination until the dam. This is visualised in Figure 1 below. The transportation of heavily contaminated waters into the Hartbeespoort Dam could mean increased health risks in aquatic life, human health, and agriculture surrounding the region (Rimayi, Odusanya, Weiss, de Boer, & Chimuka, 2018). It is important to understand the quality of water of the Hartbeespoort Dam because it is used for a wide variety of human activity such as canoeing, swimming and fishing.

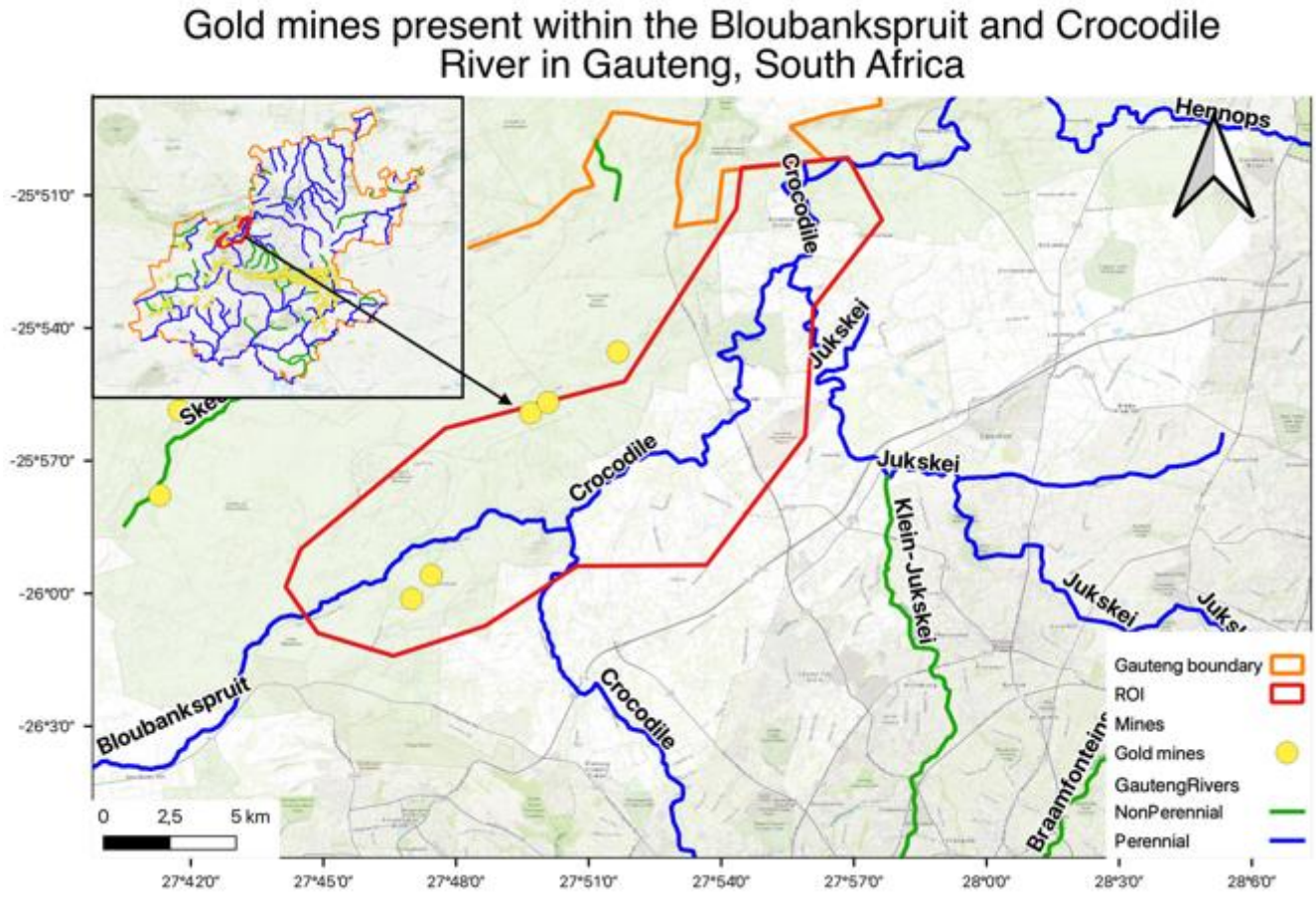


Figure 1: Study area of the Bloubankspruit and Crocodile Rivers

3.2 RESEARCH DESIGN

The model below depicts a flowchart of each step in this research process. This begins with acquiring different environmental datasets and ends with advising on different recommendations and intervention methods for acid mine drainage.

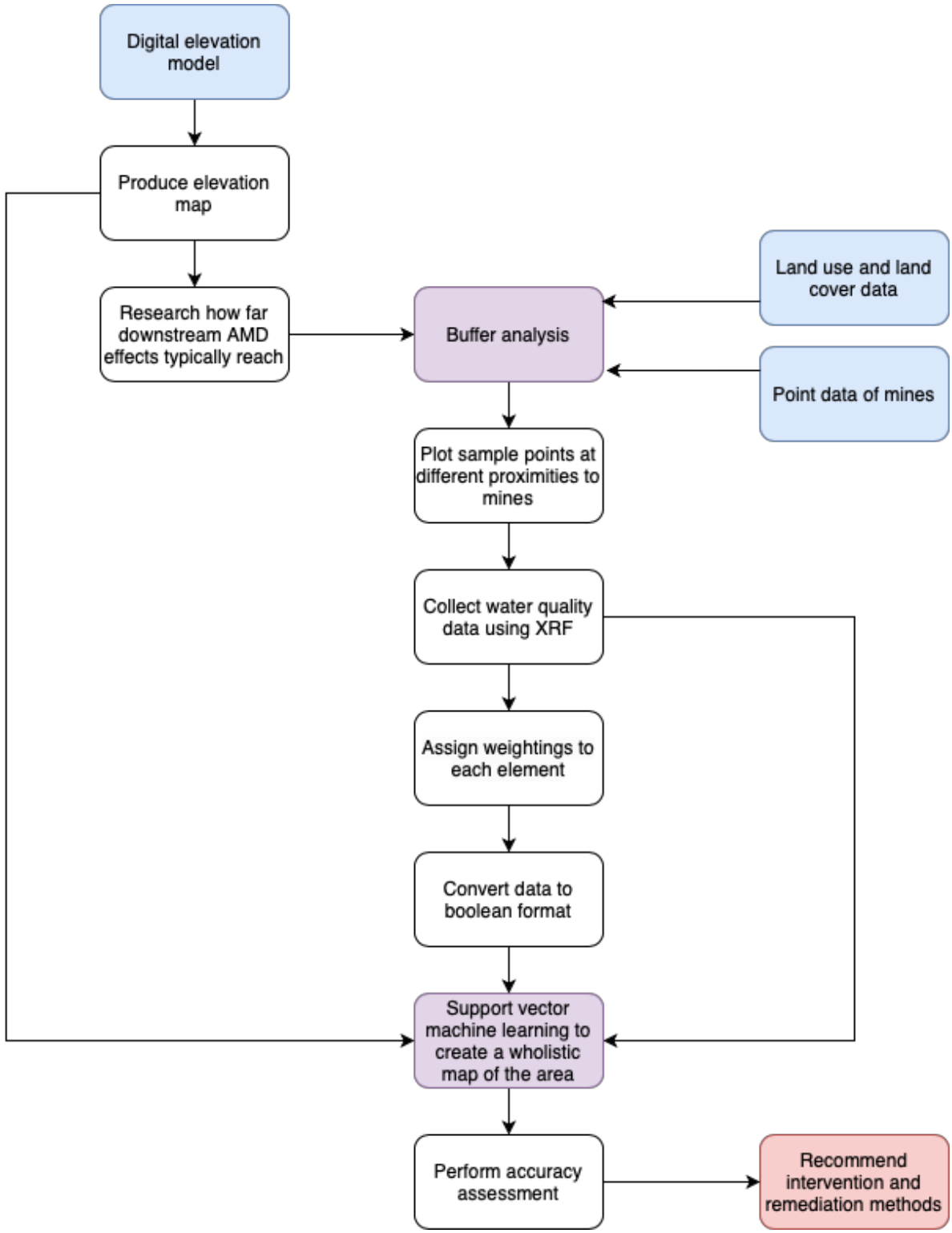


Figure 2: Research design

3.3 DATA REQUIREMENTS

Table 3: Data requirements for the implementation of this study

Data Set	**Temporal Resolution**	**Spatial Resolution**	**Source**	**Web Links**
Digital Elevation Model	2014	30m	The United States Geological Survey	https://earthexplorer.usgs.gov
Mineral Resources West Rand	2006	Point data; Vector shapefile	Council for Geoscience South Africa	https://maps.geoscience.org.za/download/mineral-shapefiles.php#
Mineral Resources Rustenburg	2014	Point data; Vector shapefile	Council for Geoscience South Africa	https://maps.geoscience.org.za/download/mineral-shapefiles.php#
Types of River Systems in South Africa	2006	Line data; Vector shapefile	Department of Water and Sanitation	https://www.dws.gov.za/iwqs/gis_data/river/All.html
Land Use and Land Cover Data for South Africa	2024	20m	Department Forestry, Fisheries and the Environment	https://egis.environment.gov.za/data_egis/data_download/current
XRF Field Data	2024	Point data	Collected data	

3.4 METHODS

Since this study focuses on river water quality, it is essential to first identify and visualize the flow of rivers within the study area. This is done by producing a Digital Elevation Model (DEM) to review the elevation and slope data for the region. This creates an understanding of the direction in which the river will flow, therefore transporting AMD waste products downstream.

Different proximities from the mines were identified through utilising a buffer analysis. From this, sample points are plotted at appropriate distances. This was based on previous studies that show that the intensity and concentration of AMD contaminants decrease with distance from

the pollution source. It is important to note that terrain and land use can significantly alter this result, which is why the use of the DEM is important. Therefore, sample points were strategically plotted with regard to distance and topographic flow.

Land use and land cover data is also considered when identifying sample locations. This is because waste from different agricultural and industrial activities may contribute to the quality of water within river systems. this includes runoff from agricultural areas, such as nutrients or pesticides, and runoff from industrial zones, which may include heavy metals or chemicals. After this analysis, chemical composition screening takes place with XRF equipment. This was chosen because of its suitability for field based elemental analysis. Data collection through the use of XRF equipment begins by first calibrating the machine. This is done by pointing the XRF at a reference material. Usually this is provided with the equipment and consists of a homogenous sample of an alloy (Oyedotun, 2018). This is done at each sample site and is crucial in ensuring that the data collected is accurate. It also indicates whether the machine is working as it should. Multiple samples are taken at each site. This will account for variability and ensure that readings are validated for accuracy.

The correctness of the collected data further contributes to the accuracy of the data. This can be ensured by repeating the reading over the same area, under the same conditions, using the same equipment (Oyedotun, 2018). This decreases the chances of large data variability over a site. Another method is reproducibility which uses a different method of extracting data from a sample and compares it to the XRF analysis (Oyedotun, 2018). This method is not ideal as the sample sites in question are not able to be disturbed, therefore, extracting samples for a different type of analysis is not appropriate in this study. The validation of XRF results can be ensured through constant calibration checks as previously mentioned, and statistical validation. This validation will be undertaken through looking at the analysis of variance (ANOVA) of the collected data. ANOVA statistically quantifies the level of variation present within and across data points. This is used to check the quality of the data (Henson, 2015). Although all sites are sampled consistently, each locations are assessed by identifying LULC, vegetation, proximity to mines and other activities, slope gradient. These observations assist in contextualizing why element values vary.

ANOVA is used to determine the variation of means within three or more sample groups. It observes whether differences in data are due to random chance or if they hold significance (Henson, 2015; Stoker, Tian, & Kim, 2020). A null hypothesis, denoted by H_0, means that all

group means are equal. H_1 is used when the null hypothesis is rejected and there is a difference in at least one of the sample means (Stoker, Tian, & Kim, 2020). After this is calculated, the F statistic is produced. This is a calculation of variance between groups divided by variance within groups. It is used to indicate whether differences between the mean values of groups are statistically significant (Stoker, Tian, & Kim, 2020; Emerson, 2022). This value is then compared to the critical value, which is usually calculated at a significance level of 0,05. The significance value corresponds with an f distribution table. The degrees of freedom, which is the number of values assessed, is also compared in order to conclude whether the null hypothesis is accepted or rejected (Stoker, Tian, & Kim, 2020). If the f statistic is higher than the critical value, it means that variation between groups is higher than variation within groups. In the context of this research, the variation between groups would refer to the variation between sample sites. The variation within groups would refer to the variation between all of the collected samples at one specific site (Stoker, Tian, & Kim, 2020). A p-value is also sometimes produced. A p-value less than 0,05 indicates that there is a significant difference between the datasets (Stoker, Tian, & Kim, 2020).

There are different ways in which ANOVA is able to be performed. The first is one way ANOVA, which compares the means of multiple groups that contain one independent variable (Emerson, 2022). The second is two-way ANOVA where there are usually two independent variables. The third and last is two-way ANOVA with replication, in which the same factors are measured multiple times (Emerson, 2022). The third way will be used for this study as it meets the criteria. The ANOVA calculation will include two independent variables, which are the location and the elevation of the study areas. The dependent variables would be the element concentrations present at the sample sites. These values are measured multiple times across the whole study area, which makes replication ideal for the analysis (Assaad, Hou, Zhou, Carroll, & Wu, 2015; Emerson, 2022). Two-way ANOVA with replication is also the most robust because element combinations are tested multiple times, which increases the accuracy of the statistical analysis.

Screening with XRF is able to take place within soil that is in direct contact with contaminated water bodies. The chemicals being assessed are Arsenic (As), Cobalt (Co), Chromium (Cr), Copper (Cu), Manganese (Mn), Nickel (Ni), Lead (Pb), Sulphur (S), and Zinc (Zn). This is recommended by the Department of Environmental Affairs, Republic of South Africa (The Department of Environmental Affairs, Republic of South Africa, 2010). The acceptable

thresholds of these elements should follow SSV1 because it pertains to the lowest thresholds of elements and specifically accounts for human and ecosystem health.

Ideally, samples would be collected and treated before readings are taken, however, due to the fact that data collection was done at privately owned locations such as in parks and lodges, permissions need to be given for soil extraction. This was not given, and owners of the relevant areas had stated that the environment is not allowed to be disturbed. Therefore, readings need to be taken when the soil is in situ. Samples are collected at surface level, which is where the most AMD contamination is present. It is also important to note that deeper sampling cannot be carried out as the environment needs to be left undisturbed.

Data could not be collected in publicly owned areas due to the lack of accessibility along the river. The previously mentioned study done by Raji et al. (2021) states that the most contaminated part of the river remains constant in contamination for about 2 km before a notable reduction is seen. Therefore, contamination should at least be assessed at a 2 km interval from the source of the mine to understand the level of contamination present due to AMD. Seeing as there is limited accessibility to the Crocodile and Bloubankspruit rivers, samples will need to be taken at the highest intervals acceptable to identify contamination, which ideally would be at 2 km intervals, however, due to the challenges in accessibility, an interval of 4 km will be used. Because data collection is done in situ, there is minimal safety hazards present when handling the samples (McIntosh, et al., 2016). The only necessary precautions are to wear gloves to reduce contamination by contact and to clean the XRF after touching contaminated soil after each sample is taken. There is little to no effect on the environment after collecting samples as it is left undisturbed.

A 4 km interval is not ideal; however, this will still have an impact of delivering good results as the overall trend can be captured over the study area. Through using a larger interval, general trends are also able to be mapped and identified. This interval is also ideal when examining a region with a relatively flat slope as movement and runoff is not common. On the other hand, local variations could be missed, finer intervals such as 1 km or 500 meters might be necessary for higher precision. This is however, most important when examining environments with varying terrains as it allows analysis to effectively account for variations in contamination levels. While monitoring heavy metal concentrations in the Red River, Vietnam, Nguyen et al. (2016) makes use of 30 m intervals over 50 sample points, which increases the accuracy and allows for a better understanding of the environment. Hence, more samples at smaller intervals

would be extremely beneficial. Because of the sparseness of the collected data, a confusion-based accuracy assessment and cross validation need to be done. This involves training the SVM model multiple times and comparing the results. This is impactful in delivering fair results and identifying hotspots through analysing the robustness of the model (Çakir, Yilmaz, Oral, Kazanci, & Oral, 2023).

A study by Patel et al. (2023) makes use of various different parameters present within a river system to measure water quality in the south of the Gujarat region, India. This study makes use of a machine learning technique to create more wholistic results. The methods of mapping water quality through machine learning includes different calculations to determine the Water Quality Index (WQI) of the different sample points (Patel, et al., 2023). The calculation of most importance from this study is the unit weight (W). Each element being monitored is given a level of importance according to how much they contribute to the worsening water quality of the river systems. The same calculation can be used in soil sampling to determine the weighting of different elements. W is calculated by dividing the proportionality constant (k) of the parameter with the standard value (S) of each parameter (Patel, et al., 2023). This calculation can be seen in equations 1 and 2 below.

Equation 1: $W = \frac{k}{S}$

Where k is calculated as follows in Equation 2: $k = \frac{1}{\Sigma\frac{1}{S}}$

These equations are reworked and simplified to create equations 3 and 4 below. The value of S now pertains to the deviation of the average value of the sampled element to the acceptable threshold as provided by the Department of Environmental Affairs, Republic of South Africa

Equation 3: $Deviation\ of\ average\ to\ standard\ = \frac{Average}{Acceptable\ standard} + 1$

Equation 4: $Weight = \frac{Deviation\ of\ average\ to\ standard}{\sum Deviation\ of\ average\ to\ standard} \times 100$

To be able to easily view the results of the weighted sum, the data is scaled from 1-10 as seen from Equation 5 below. This allows for the minimum value of AMD to be 0 and the maximum value to be 10.

Equation 5: $Scaling\ 1\ to\ 10 = \frac{Weighted\ sum - Min}{Max - Min} \times 10$

The collected data is then tabulated and converted to a Boolean format where samples within limits for water quality will be labelled “FALSE” for contamination, and samples over the limits will be labelled “TRUE” for contamination. The overall threshold of contamination will be calculated based on a weighted calculation of all of the acceptable standards per element.

After the weightings for all elements are concluded, the elements will be run through a support vector machine learning (SVM) algorithm, along with the elevation and mine point data files. To maximise accuracy over a small dataset, cross validation will be utilised at an 80/20 split for training and testing data. This split allows for the data to be easily divided into 5 different subsets. This is needed for the process of cross validation to occur, which involves classifying the data multiple times to increase the accuracy and robustness of the model (A. Ramezan, A. Warner, & E. Maxwell, 2019). These models will produce wholistic results of the contaminated river basin, in which the presence of AMD will be concluded. The SVM algorithm works by taking the vector points and elevation data and classifying the area based on the relationship of these two layers (Arabgol, Sartaj, & Asghari, 2016). The way that SVM works is by separating data through the use of 2D lines or 3D hyperplanes. The placement of these lines or planes is strategically positioned in order to maximise the distance from the plane to the different classification groups. The points that are closest to the line or plane is referred to as support vectors. This type of classification is generally used with classifying binary data (Gharakhanlou & Perez, 2023).

The SVM classifier has been selected for this study due to its strengths in handling small, high dimensional datasets (Sharma & Sharma, 2016). This is beneficial in environmental modelling where ground truth data is limited but numerous spatial features are available. Because the sample size in this study is constrained, the use of SVM is an optimal classification method to use (Sharma & Sharma, 2016). Additionally, SVM is robust to non-linear relationships. This is particularly relevant to AMD mapping due to the complex interactions of slope, land use, and proximity to mining infrastructure (Sharma & Sharma, 2016).

The use of SVM classification in geospatial contexts has been highlighted in many areas of study. For example, landslide susceptibility is concluded through the use of SVM classification within different mountainous regions in India. Although the study states that SVM is not the most ideal for larger sets of data, the accuracy of the model was quite high at 91,58% (Sharma, Saharia, & Ramana, 2024). Another use of SVM in geospatial studies includes Karimi et al. (2019) investigating urban expansion prediction in Guilford County, North Carolina, USA.

This study makes use of different datasets such as land use and land cover, transportation, and population to identify the likelihood of urban expansion. The study yielded a training accuracy metric of 98% and testing accuracy of 85% after converting the training datasets to a binary format. Therefore, based on the two studies discussed, binary data over small datasets yields the most accurate results. It is important to note that SVM classifies data based on the best decision boundary of the area, hence, exact data is not necessarily preserved (Arabgol, Sartaj, & Asghari, 2016).

The results of the data should be corroborated by the observations that are made at the different sampling sites. For example, if contaminant concentrations are high, there should be physical indications such as discoloration, odour, or vegetation die off to validate these results. In order to further validate the results produced by the SVM model, an accuracy report will be generated. The SVM model that was previously generated in GIS software will be used in R Studio. This will be beneficial in generating the report that contains all accuracy metrics. This report will include accuracy, precision, recall, and F1 metrics to ensure a detailed evaluation of the performance of the SVM model (Sethy, Barpanda, Rath, & Behera, 2020). The caret package will need to be imported in R Studio for this step. A confusion matrix is generated to assess four outcomes of the model. These are true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). For example, a true positive will occur when the model correctly identifies a contaminated area. A false positive will occur when the model identifies an area of contamination where there is in fact no contamination present. Similarly, a true negative will occur when the model successfully identifies an area where no contamination is present. A false negative will occur when the model identifies an area of no contamination where there is in fact contamination present.

The accuracy of the model is calculated by the ratio of TP and TN to the total amount of samples as seen in Equation 6. This is used to test the overall correctness of the model (Sethy, Barpanda, Rath, & Behera, 2020). While accuracy gives an overall success rate of the model, it can be misleading if the dataset is imbalanced. That is why other metrics are also important to review and analyse.

Equation 6: $Accuracy = \frac{TP+TN}{All\ samples}$

Precision calculates the ratio of TP to the sum of TP and FP as seen in Equation 7. It measures the proportion of true AMD affected predictions that were actually correct (Sethy, Barpanda,

Rath, & Behera, 2020). It indicates the reliability of the model. A high precision means fewer misclassifications of contaminated areas (Sethy, Barpanda, Rath, & Behera, 2020).

Equation 7: $Precision = \frac{TP}{TP+FP}$

Recall is calculated by the ratio of TP to the sum of TP and FN. This calculation provides insight into the sensitivity of true positives as seen in Equation 8. It reflects the model's ability to correctly identify actual AMD affected areas. It is especially important in risk mapping as missing contaminated sites could have dire consequences. High recall ensures that most contaminated areas are detected, even if it means slightly more false positives (Sethy, Barpanda, Rath, & Behera, 2020).

Equation 8: $Recall = \frac{TP}{TP+FN}$

The F1 score is lastly calculated as follows in Equation 9.

Equation 9: $F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)}$

This calculation works to exclude any outliers within the data to convey a more balanced reflection of the model's performance, where both false positives and false negatives are important (Sethy, Barpanda, Rath, & Behera, 2020). A high F1 score indicates a strong overall performance of the model.

In the context of AMD detection, high recall is crucial because missing contaminated areas could lead to continued ecological damage and inefficient remediation efforts. Precision is also important in order to prevent the false labelling of uncontaminated zones as polluted (Sethy, Barpanda, Rath, & Behera, 2020). In doing so, resources could be wasted, and monitoring efforts could be misled. Therefore, an F1 score is important as it balances both of these needs (Sethy, Barpanda, Rath, & Behera, 2020).

To further assess the binary classification performance of the SVM model, a Receiver Operating Characteristic (ROC) curve was generated using the pROC package in R (Narasimhan & Agarwal, 2017). The ROC curve illustrates the transaction between the true positive rate and the false positive rate across a range of classification thresholds. The Area Under the Curve (AUC) was calculated to provide an overall measure of the model's ability to distinguish between AMD-affected and non-affected areas (Narasimhan & Agarwal, 2017).

The AUC value, which ranges from 0 to 1, provides a single scalar measure of model performance. An AUC closer to 1 indicates excellent discriminatory ability, while a value near 0,5 suggests random performance (Narasimhan & Agarwal, 2017). A high AUC value, which is greater than 0,85, indicates strong classification performance and confirms that the SVM model performs well across different decision thresholds (Narasimhan & Agarwal, 2017).

After the successful analysis of the extent of AMD, intervention and remediation methods will be reviewed and discussed.

4. RESULTS

Figure 3 below depicts the area of study and points to which samples will be taken. These points are plotted according to the most accessible areas along the river. The first point is plotted within a two-kilometre radius of the Kromdraai mine. It is important that the first sample is taken within this distance as the highest concentration of AMD is usually stable and prevalent for the first two kilometres of the drainage site according to Raji et al. (2021). The other mines in the area do not fall within a 2 km radius of the river as seen in Figure 3. This means that the isolated impact of AMD along a river can be assessed. This increases the accuracy of the study. 10 individual samples were taken per sample point to ensure accuracy based on the article by Yang et al. (2017) whereby 5 samples were taken per point. Because this study examines a smaller area, more samples are taken to ensure that the SVM has sufficient data to work with and to increase the accuracy of the results. This means that a total of 60 measurements has been taken along the Bloubankspruit and Crocodile River in this study.

The observations at sample point 1 includes a strong smell of rotten egg because of a large amount of sulphur present within the river. This was also prevalent along points 2 and 3. The water was very murky and cloudy and was greenish brown in colour. There were also small amounts of foam that were present on the water surface at this sample point. The immediate surrounding area was a sculpture park, which is very accessible to the public.

The surrounding area of sample point 2 belongs to a lodge and spa which is also very accessible to the public. The staff at this site did warn that nobody should come into contact with the water from the river as it is severely contaminated. This was visually evident due to the high turbidity of the water, green-brown colour, and the stench being emitted from the water body.

Point 3 also had a strong smell, however, it was not as severe as in points 1 and 2. Water here was browner in colour and the level of turbidity remained high. This point belongs to a lakeside game lodge. Many guests visit this site for recreational activities such as hiking, cycling, canoeing, and open water swimming. Information from staff indicates that the water quality has been approved for swimming by a testing centre. Many animals are also free to roam around these areas. On the way to the testing site, giraffes, zebra, and springbok were seen in close proximity of the water source.

The turbidity at point 4 was slightly less, however, the water was still a greenish brown. The smell of sulphur also decreased a bit from point 3. The same observations are true for points 5 and 6. There are little to no variations in the visual survey along the last three points apart from the smell of sulphur being effectively gone by point 6.

Figure 3 below shows where each sample point is located, and which mines would affect the river based on the 2 km buffer.

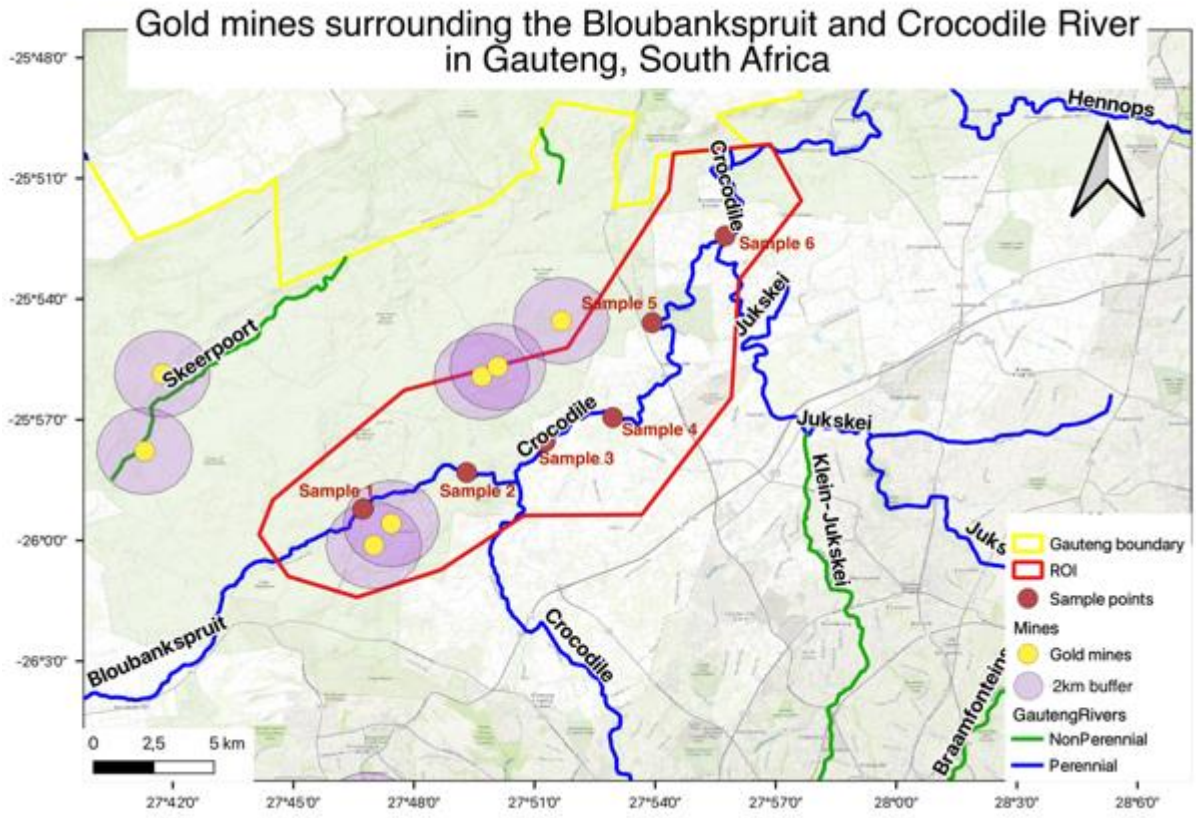


Figure 3: 2km buffer of gold mines around the study area of the Bloubankspruit and Crocodile River

Figure 4 below shows the different land use and land cover (LULC) data surrounding the area of study. The river is mostly surrounded by grassland with small amounts of cultivated land and human activity occuring nearby. It is therefore very unlikely that significant amounts of additional waste are entering the river. The built up areas are attributed to shopping areas, residential areas, and different eco-lodges, where insignificant amounts of pollution are emitted.

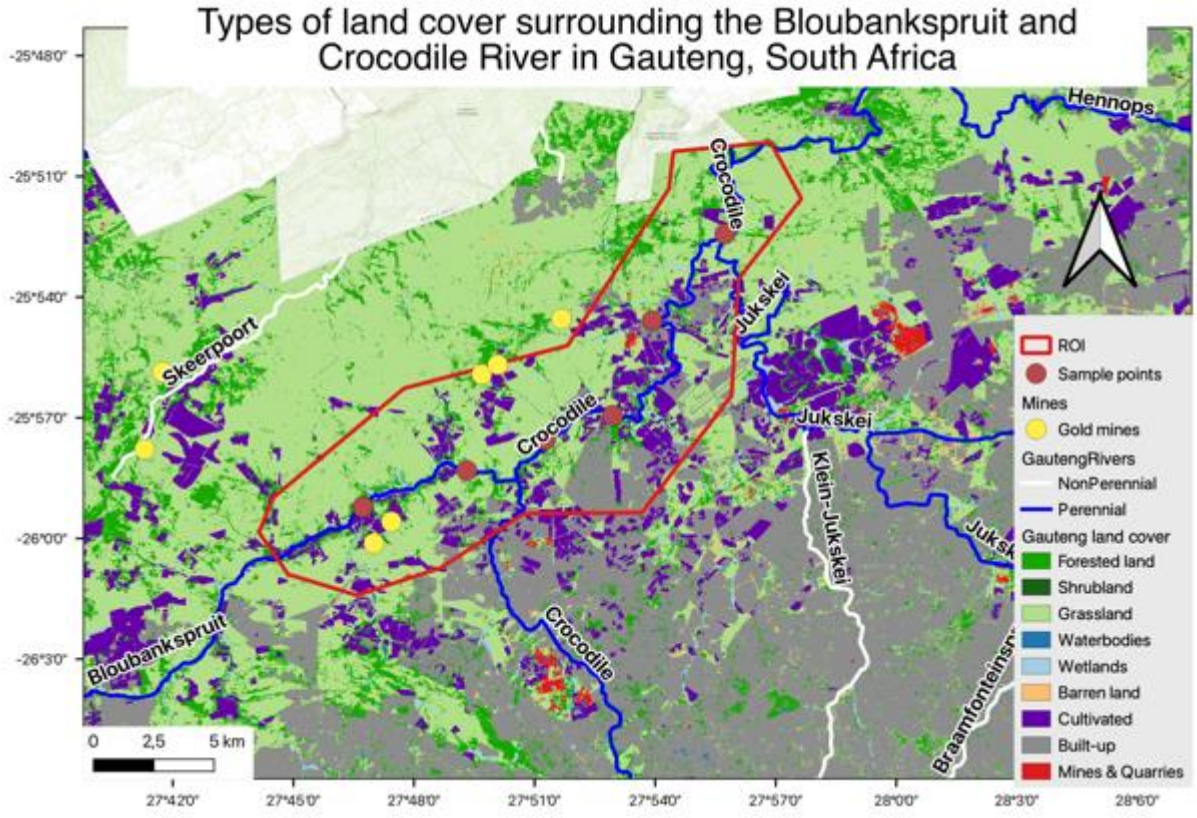


Figure 4: Land use and land cover around the Bloubankspruit and Crocodile River

Figure 5 below shows the elevation of the area. It is conclusive that the river flows in a north-easterly direction based on this. This information is further substantiated by the fact that it flows directly into the Hartbeespoort Dam which is located just a further bit north from the study area, in the North West province. AMD effects are generally present for 2 km within the radius of the mine before decreasing as stated by Raji et al. (2021). Therefore, the first sample point is plotted within a 2 km buffer of the mine being examined, and the rest of the points are plotted based on the accessibility of the area. This is done through the method of incremental sampling as mentioned by Yang et al. (2017), and the direction of water flow as seen in the elevation map below.

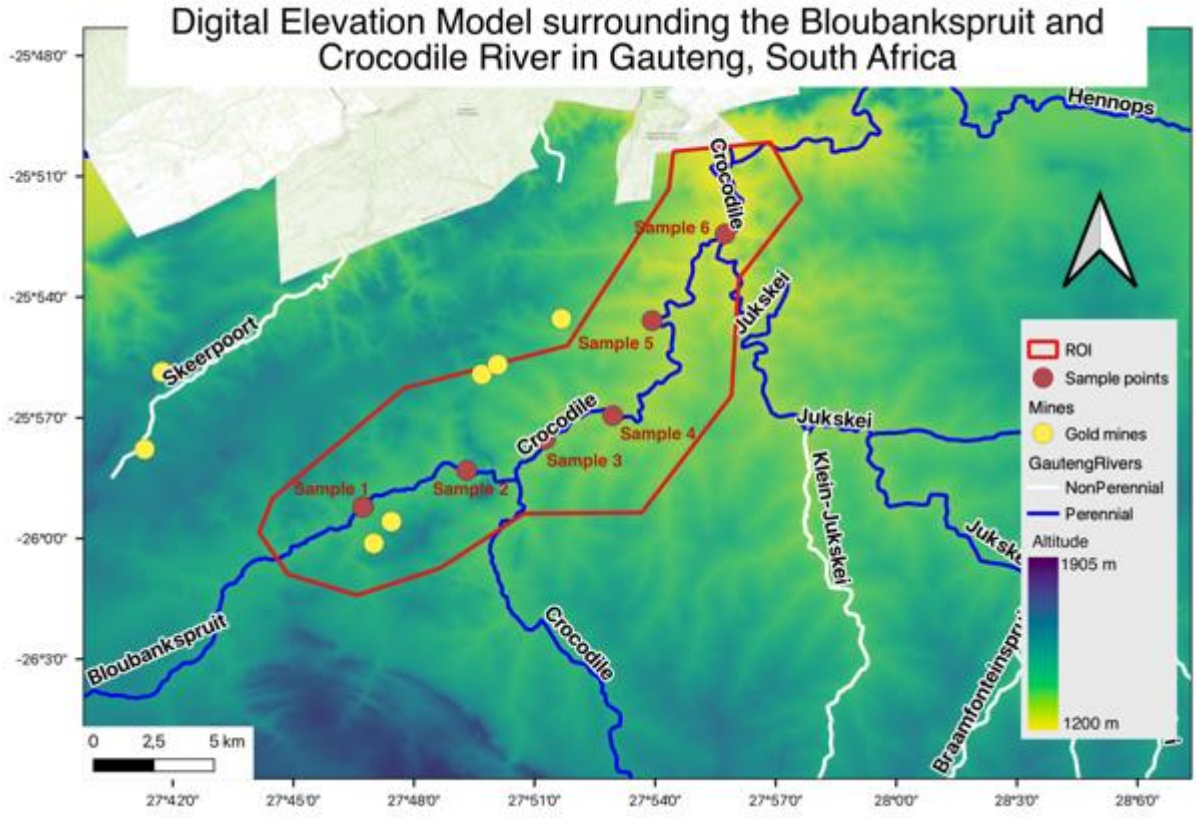


Figure 5: Digital elevation model of the study area.

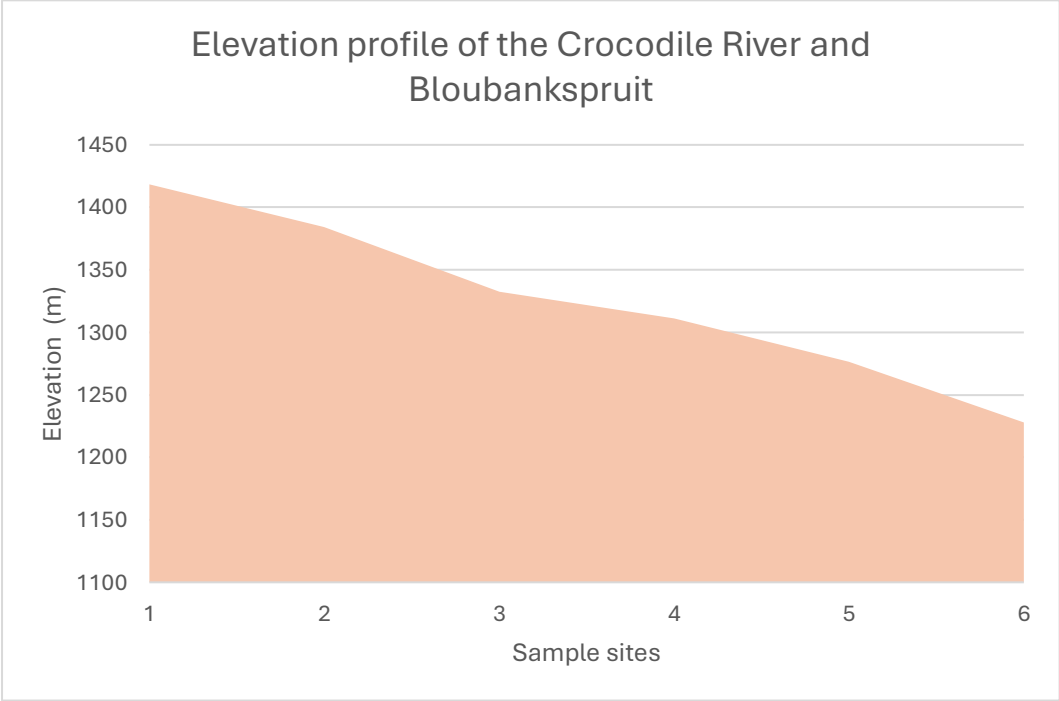


Figure 6: Elevation profile across all sample sites

The profile of the elevation data across the sample points is shown in figure 6 above. This graph indicates that a constant gradient is present throughout the study area. The gradient can be quantified by dividing the difference in elevation by the distance in kilometers between the first and last sample site. This calculation is seen in equation 10 below:

Equation 10: $m = \frac{y_2 - y_1}{x_2 - x_1}$

The highest point of the study area is 1418,4m and the lowest point is 1228m. The total distance between the first and last study point is 19670,579m. The result of the equation above is therefore 0,00967943, which signifies an extremely gentle gradient along the study area. This means that surface runoff of AMD does not generally occur within this area. This affects the distribution of AMD across the area as there will be a low rate of movement of contaminants between sample sites (Müller, Österlund, Marsalek, & Viklander, 2020). Therefore, contamination will be most present at the source of the mine. The decision to include slope as the primary terrain variable was based on its influence on surface runoff and erosion processes, which are key drivers in the transport of AMD within fluvial systems.

Due to the gentle gradient of the slope, a sample interval of 4 km is ideal. This is due to the increased probability that contamination is unlikely to migrate downstream at a rapid rate.

Table 4: Mean amounts of elements present per sample site

	As	**Co**	**Cr**	**Cu**	**Mn**	**Ni**	**Pb**	**S**	**Zn**
Acceptable threshold	**5,8**	**6**	**6**	**16**	**245**	**620**	**20**	**4000**	**2000**
Sample 1	2,85	40,10	40,60	32,60	860,70	41,00	11,35	6911,17	147,00
Sample 2	0,00	0,00	0,00	26,22	582,58	0,00	8,33	4572,54	50,55
Sample 3	0,00	16,30	0,00	13,20	573,00	1,10	0,40	3205,69	28,70
Sample 4	34,00	11,70	12,70	28,30	1015,90	13,70	7,20	562,91	277,40
Sample 5	0,00	9,70	0,00	15,42	378,28	0,00	3,33	2474,56	33,45
Sample 6	0,00	49,20	33,00	23,30	565,80	7,50	1,50	1293,04	52,90

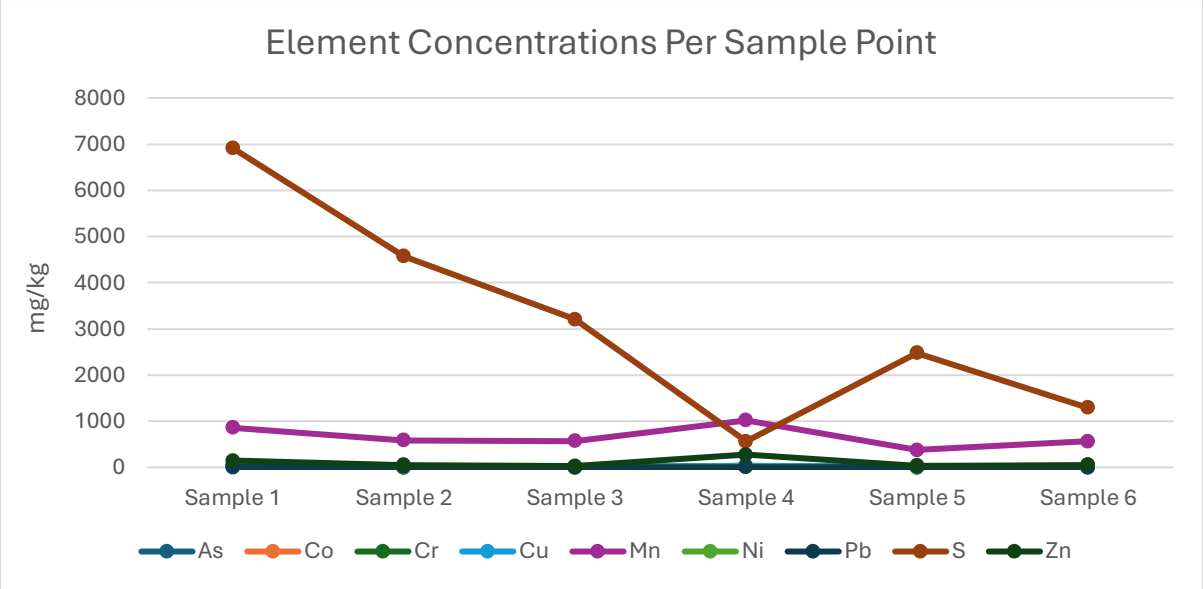


Figure 7: Mean element concentrations over each sample site

Table 4 and figure 7 above shows the mean concentrations of each element over all sample areas. When compared to the acceptable threshold, it is seen that arsenic values are extremely high at sample site 4. This is unusual as it is within the acceptable threshold along all other sites. Therefore, it can be concluded that this anomaly is not due to AMD. There is a lot of cultivated land that surrounds sample point 4. This means that it is highly likely that contamination is caused by chemical runoff from agricultural activities. The increase in arsenic could have extreme impacts on the quality of aquatic and plant life as well as affect human health. This would be done by the introduction of cardiovascular issues and cancer (Edmunds, Ahmed, & Whitehead, 2015).

Cobalt values surpass the acceptable threshold along all sample points except for point 2. This has a negative impact on aquatic life because an accumulated amount and long term exposure could be toxic to fish and other organisms (Barrio-Parra, et al., 2018). An extremely high amount of cobalt, as seen in points 1 and 6, could result in extremely negative impacts in human health. These impacts include the presence of cardiovascular effects, thyroid problems, and lung damage when consumed (Barrio-Parra, et al., 2018).

Chromium values along are over the threshold at sample points 1, 4, and 6. Because these values are so high, the damage to aquatic life is extremely severe. These levels of chromium could result in the death of species and affect reproduction within aquatic organisms (Atibu, et al., 2018). Furthermore, the risk to human health through consumption could lead to respiratory problems, skin rashes, kidney damage, and an increased risk of cancer (Atibu, et al., 2018).

Copper values are over the acceptable threshold at all sites apart from sample sites 3 and 5. Although values at sites 3 and 5 are below the threshold, they still sit very close to the border of the maximum acceptable amount. An increased amount of exposure to copper can be toxic to aquatic life by affecting their respiratory process (Atibu, et al., 2018). Furthermore, if levels are not managed or decreased, the long term exposure of humans to copper could lead to neurological damage, liver and kidney damage, and gastrointestinal issues (Atibu, et al., 2018).

Manganese values are all substantially above the acceptable threshold. This is extremely problematic as it may cause neurological damage in both aquatic and human health. This directly impacts motor skills and the overall health of ecosystems and communities that are reliant on the contaminated water (Barrio-Parra, et al., 2018).

Nickle values are all extremely low, which is good as the values are extremely below the threshold of acceptable amounts. This means that no health or environmental risks are present due to excessive nickel concentrations.

Lead amounts are all within the acceptable values, which is satisfactory as this element is highly toxic and presents significant risks to the environment and people that interact with it (Atibu, et al., 2018).

Sulphur values are extremely high along the first two sample sites. These values incrementally decrease over the next four sites, apart from an anomaly at sample 4. This makes sense as the process of AMD involves the flow of mine waste over sulphide materials to form sulphuric acid (Lemos, et al., 2020). An extremely high sulphur level, as exhibited in sample site 1 causes water to be acidic, which affects aquatic life. This is because many organisms are sensitive to the pH of their environment (Likus-Cieślik, Smoliński, Pietrzykowski, & Bąk, 2019). This may be attributed to its location downslope of historical mine tailings, which are known sources of sulphide-bearing materials. During rainfall events, these materials can oxidize and mobilize, leading to increased sulphur loading in adjacent soils and sediments (Likus-Cieślik, Smoliński, Pietrzykowski, & Bąk, 2019). The terrain at Site 1, characterized by low slope and limited vegetation cover, likely facilitates runoff accumulation and contaminant retention. Furthermore, communities that interact or ingest the water at points 1 and 2 may experience nausea, headaches, and an irritated respiratory system (Likus-Cieślik, Smoliński, Pietrzykowski, & Bąk, 2019).

Zinc values are all within the acceptable threshold by an extremely high margin. Therefore, there is no immediate risk to the health of humans or the environment. There is also no immediate concern regarding zinc toxicity within the environment because of the high margin.

Values across all elements seem to deviate at sample site 4. As per figure 4, the land use along this point consists of cultivated land and a built up areas. Therefore, these anomalies may be due to agricultural or municipal waste. This is further supported by the fact that a high amount of agricultural land and built up land are clustered around point 4, whereas the other points are surrounded by grassland and small amounts of cultivated land. It should also be noted that the presence of these elements are not migrated further downstream as the concentrations over all elements at points 5 and 6 are consistent with the values at points 1, 2 and 3. Hence, it may also be possible that these anomalies at point 4 are not a result of large amounts of contamination, but rather, smaller chemical activity or waste that is taking place at the study site. Because the readings are based on contamination in soil that is in direct contact with the river, and not river water itself, the concentration of elements may only be present within the soil structure. This means that it does not largely interact with river water that flows downstream.

The distribution of the data needs to be evaluated before ANOVA takes place. The results of this can be seen through the skewness and kurtosis of elements in the table below.

Table 5: Skewness and kurtosis for each element

	As	**Co**	**Cr**	**Cu**	**Mn**	**Ni**	**Pb**	**S**	**Zn**
Skewness	3,55	4,14	3,37	0,72	0,40	2,08	0,18	0,53	1,16
Kurtosis	11,09	21,19	10,29	2,79	-0,30	3,15	-1,43	-0,71	-0,06

To evaluate whether the assumptions of ANOVA are met, specifically the assumption of normality, skewness and kurtosis were calculated for each elemental variable. ANOVA assumes that the data is approximately normally distributed, as violations can lead to inaccurate conclusions about group differences. Skewness measures the asymmetry of a distribution, while kurtosis reflects the presence of extreme values (Ezeora, Heckenbergerova, & Musilek, 2015). In general, skewness values outside the range of -2 to 2 and kurtosis values above 7 are considered indicative of significant deviation from normality.

In this study, elements such as arsenic, cobalt, and chromium displayed high positive skewness at 3,55, 4,14, and 3,37 respectively, and excess kurtosis at 11,09, 21,19, and 10,29, indicating highly non-normal distributions. These distributions suggest that a small number of samples had extremely high concentrations, possibly linked to localized pollution or contamination events. Other elements, including lead, sulfur, and zinc, exhibited skewness values closer to zero at 0,18, 0,53, and 1,16, and relatively low kurtosis at -1,43, -0,71, and -0,06, suggesting approximately normal distributions. Manganese, nickel, and copper showed moderate skewness and kurtosis, with lead and nickel still within acceptable ranges.

Given that only a few variables show substantial non-normality, and that ANOVA is generally robust under such conditions, the decision was made to proceed with ANOVA without data transformation. This approach aligns with practices in environmental research, where elemental data often display natural skewness due to localized contamination, and where ANOVA can still provide valid comparisons across groups when applied carefully (Ezeora, Heckenbergerova, & Musilek, 2015). The use of ANOVA is therefore considered appropriate for the current dataset and research objectives.

Table 6: ANOVA results representing all collected sample data

ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Sample	28966937,30	5,00	5793387,47	30,23	4,01E-27	2,23
Columns	564392497,00	9,00	62710277,50	327,27	3,45E-212	1,90
Interaction	242184285,00	45,00	5381873,00	28,09	3,44E-113	1,39
Within	103471723,00	540,00	191614,30			
Total	939015443,00	599,00				

Table 6 above identifies the variation within the datasets at different sample points. The rows in the table above represent the differences between the sample points where the row is labelled ‘Sample’. ‘Columns’ represent differences between the elements being studied, and ‘Interaction’ represents depicts the combined results of the sample and column rows. ‘Within’ represents the variation within each group. The values of ‘SS’ in the columns pertains to the sum of squares and indicates the total amount of variation within the data (Kim, 2017). The column ‘df’ relates to the degrees of freedom within a dataset. This is calculated by counting the amount of independent values in the dataset (Kim, 2017). ‘MS’ stands for mean square and is a representation of the SS column being divided by the df column. This value represents the average amount of variation over the dataset (Kim, 2017).

The F statistic is calculated on the ratio of the mean square (MS) between groups with the MS within the sample groups. This calculation tests whether the mean between groups is significantly different (Kim, 2017). A higher value indicates more significant differences. The results above show a significant difference between sample groups, which indicates a drastic change in element concentrations over the whole study area. The variation between element groups is extremely high as well at an f-statistic value of 327,27. This is due to the different levels of relative toxicity and levels at which different elements naturally occur within the environment.

The p-value represents the probability that the observed results are due to random chance. A very small p-value is generally less than 0,05 and indicates that the differences are statistically significant and not due to random chance (Kim, 2017). Seeing as the p-values above are extremely close to 0, these results are highly accurate between sample sites as well as between elements.

The F crit value is also referred to as the F critical value and is calculated by identifying the critical value of F at a significance level, which is usually 0,05. If the value of F is greater than the F critical value, the null hypothesis is rejected and it is concluded that there is a significant difference between the groups (Kim, 2017; Emerson, 2022). There is difference of 28 between the F and F critical values for the variance between sample sites. This is a highly substantial amount and therefore indicates that the null hypothesis is rejected. This means that there is a high level of variance between sample sites and that contamination amounts experience a severe change between sample sites. There is also a difference of 26 between the f-statistic and the f critical values of the variance between elements at the sample sites. This further supports the significance of the variation of elements due to concentrations naturally occuring at differing levels.

To strengthen the reliability of the reported concentration values and address measurement uncertainty, 95% confidence intervals (CIs) were calculated for each element across all sampling points. These intervals were derived using the standard deviation and sample size, under the assumption of a normal distribution. The confidence intervals quantify the range within which the true mean concentration is likely to fall with 95% certainty (Ott, 2018). These results are summarized in the table below and provide a statistically sound basis for concluding significant variation in contaminant distribution across the study area.

Table 7: 95% Confidence intervals for each element over each sample site

Sample_pt.	As_95%_CI	Co_95%_CI	Cr_95%_CI	Cu_95%_CI	Mn_95%_CI	Ni_95%_CI	Pb_95%_CI	S_95%_CI	Zn_95%_CI
1	2,21	46,25	61,24	4,50	193,39	25,27	1,58	721,84	23,25
2	0,00	0,00	0,00	6,11	208,55	0,00	1,20	1368,18	17,42
3	0,00	20,90	0,00	3,42	246,53	2,49	0,90	1017,20	7,99
4	31,47	17,66	19,16	4,27	209,35	8,74	2,59	52,65	19,19
5	0,00	15,42	0,00	8,98	136,86	0,00	3,10	1409,96	12,06
6	0,00	74,38	50,02	12,29	242,00	8,70	3,39	427,88	13,84

To improve the statistical robustness of the findings and address concerns regarding reliability, confidence intervals and measurement uncertainties were incorporated into the analysis. For each element across all sampling sites, 95% CIs were calculated based on the mean and standard deviation per site. These intervals provide a range within which the true mean concentration is expected to lie with 95% certainty.

These values reflect the upper bound of uncertainty associated with each measurement, indicating the potential range within which the true concentration may lie. For example, Sample 1 shows relatively elevated levels of sulphur at 721,84 ppm, chromium at 61,24 ppm, and manganese at 193,39 ppm, suggesting possible contamination. In contrast, Sample 2 contains no detectable levels of arsenic, cobalt, or chromium but shows a very high sulfur content at 1368,18 ppm. This may be indicative of acid mine drainage influence. Notably, Sample 4 records the highest arsenic confidence interval at 31,47 ppm, while Sample 6 displays high values for cobalt at 74,38 ppm and chromium at 50,02 ppm. The variation across samples suggests spatial heterogeneity in contamination levels, possibly influenced by proximity to

mine tailings, local hydrology, or land use. These confidence intervals are critical for assessing the reliability of measurements and for informing risk assessments in contaminated landscapes.

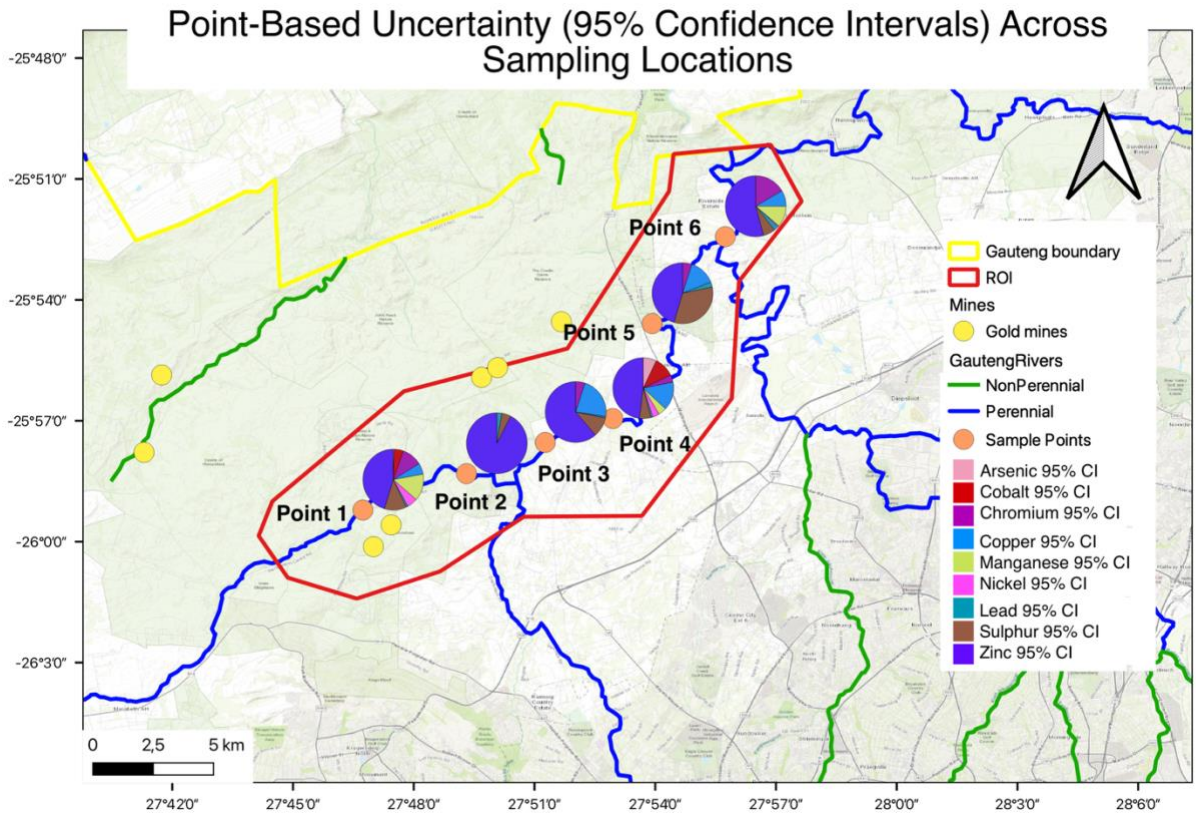


Figure 8: Point based uncertainty map across all sample sites

The uncertainty map above depicts the uncertainty that has been discussed. Zinc is clearly shown as having the greatest level of uncertainty. This means that the concentrations of zinc either fluctuate greatly across the region, or that there were sensitivity issues with the XRF that were unaccounted for. The other elements do not seem to vary as greatly across th esampling sites. The only anomaly would be sulphur at point 5, which has a much greater uncertainty than the other points. This is most likely due to sensitivity issues with equipment or values fluctuating due to additional contamination taking place at that point.

After data is collected at each sample point, the elements that are most hazardous are reviewed and all outliers are disregarded. The weighted sum is calculated as shown in equation 11 below. This calculation is based on the amounts of each element present at the sample point multiplies by a calculated weighting. The calculated weighting per element is calculated based on a calculation by Patel et al. (2023) at which the deviation of the average values across all sample

points is calculated per element. This is then divided by the standard deviation across all values for the respective element. This value is then converted into a percentage.

Equation 11: $Weighted\ sum = (9{,}68 \times As) + (21{,}29 \times Co) + (15{,}98 \times Cr) + (11{,}51 \times Cu) + (17{,}42 \times Mn) + (4{,}78 \times Ni) + (5{,}96 \times Pb) + (8{,}43 \times S) + (4{,}93 \times Zn)$

After the weighted calculation is established, the acceptable standard values are used in the weighted sum calculation and then scaled from 1-10 in order to identify the weighted threshold of AMD present within the river system. The result was a value of 5,72 which means that after all sample points are weighted and scaled, a value above 5,72 means that a high amount of AMD is present and intervention is needed. Alternatively, all values below 5,72 indicates a lower amount of AMD present that is not of immediate concern.

Figure 9 below shows exactly where AMD is severe by averaging the scaled results per sample point. The graph is inversely proportionate and shows a steep decline in AMD before the values gradually even out. AMD is extremely severe at sample point 1. The next point still shows a high amount of AMD; however, it is much closer to the threshold and will not need urgent intervention. The rest of the sample points are below the threshold, and the values decline at a more gradual rate than between points 1 and 2. In order to quantify the severity at which AMD declines, it is noted that the difference between points 1 and 2 is 4,1 scaled units whereas the difference between points 3 and 6 is 2,4 units. This steep decline in the presence of AMD along the study area supports the conclusions of the ANOVA analysis as values vary significantly both between and within sample sites.

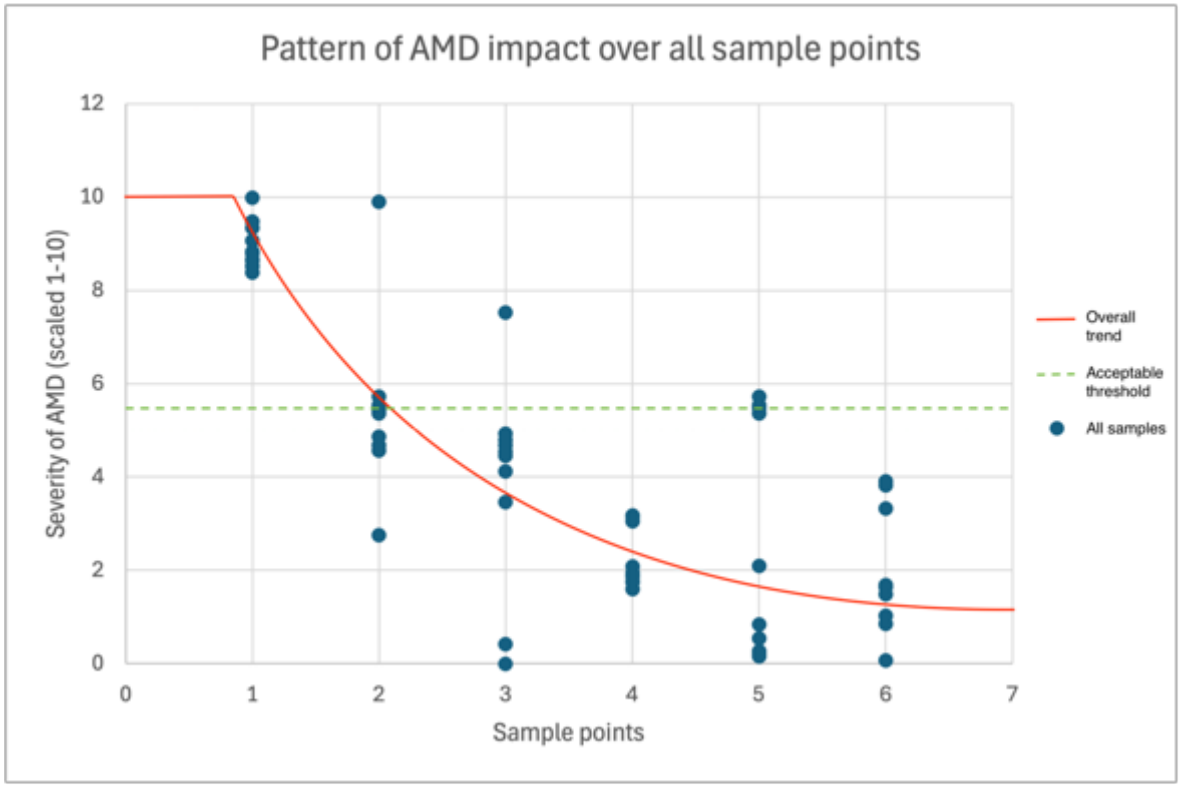


Figure 9: General trend in AMD impact over all sample sites

The results are further supported by the SVM classification in Figure 10 below. Figure 10 includes results from the cross classification of SVM. The dataset was split into 5 equal subsets, where four were used for training and the last subset was used for testing. This subset changed over each classification. The results produced extremely similar results indicating that the results produced have a high accuracy. It is also indicative of a robust SVM model. In the classification, it can be seen that points 1 and 2 have high amounts of AMD present. The severity of AMD seems to dissipate between points 2 and 3, with little to no AMD present over the last three points. This is further supported by the elevation data in figure 5 above, in which it is described that the low slope gradient of the area contributes to a small amount of surface runoff of contaminated matter.

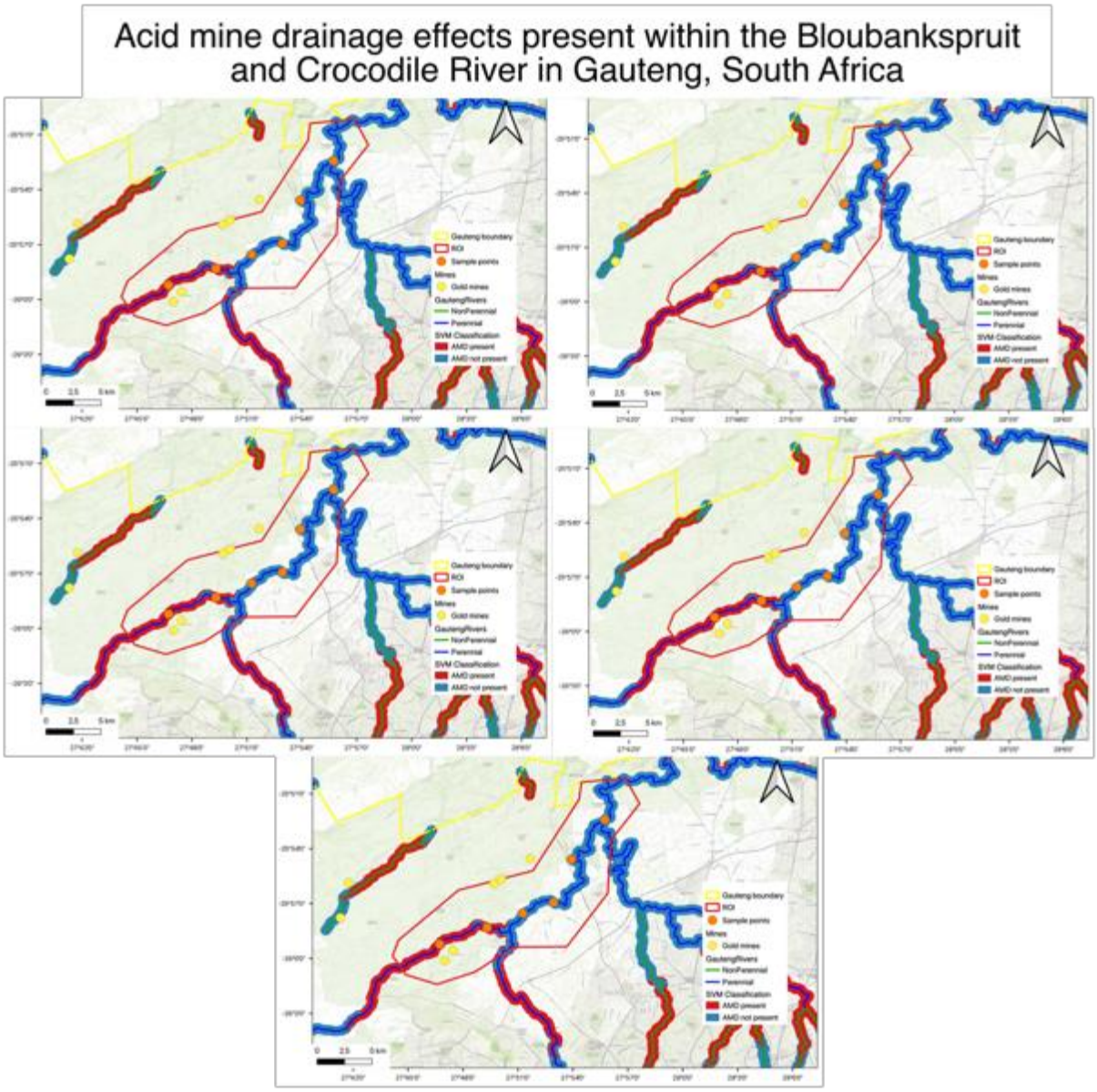


Figure 10: Cross classification implementation through support vector machine learning algorithms

The accuracy matrix shown below in Table 8 shows the amounts of true and false positives and negatives in the data per each classified SVM model. The amount of TP values pertain to how many samples were classified above the AMD threshold. This is both in the SVM model, and the weighted AMD calculation. FP shows how many samples were above the AMD threshold according to the AMD weighted calculation, but below the threshold in the SVM model. FN shows how many samples were above the AMD threshold according to the SVM model, but below the threshold in the AMD weighted calculation. Lastly, TN quantifies how many samples are below the AMD threshold in both the SVM model, and the weighted AMD calculation. It is shown that the majority of the data collected contains a large number of values

where AMD is not present. The performance of this model can be quantified by the accuracy metrics shown in table 9 below.

Table 8: Accuracy matrixes for all cross-classification results

SVM Model 1		
	Positive	Negative
Positive	True Positive (TP) = 9	False Positive (FP) = 5
Negative	False Negative (FN) = 4	True Negative (TN) = 42
SVM Model 2		
	Positive	Negative
Positive	True Positive (TP) = 11	False Positive (FP) = 2
Negative	False Negative (FN) = 3	True Negative (TN) = 44
SVM Model 3		
	Positive	Negative
Positive	True Positive (TP) = 10	False Positive (FP) = 4
Negative	False Negative (FN) = 4	True Negative (TN) = 42
SVM Model 4		
	Positive	Negative
Positive	True Positive (TP) = 12	False Positive (FP) = 2
Negative	False Negative (FN) = 4	True Negative (TN) = 42
SVM Model 5		
	Positive	Negative
Positive	True Positive (TP) = 10	False Positive (FP) = 3
Negative	False Negative (FN) = 2	True Negative (TN) = 45

Based on the table above, the following metrics are calculated:

Table 9: Accuracy metrics for analysing the performance of cross classification SVM

	Model 1	Model 2	Model 3	Model 4	Model 5	Average
Accuracy	0,85	0,92	0,87	0,90	0,92	0,89
Precision	0,64	0,85	0,71	0,86	0,77	0,77
Recall	0,69	0,79	0,71	0,75	0,83	0,76
F1 Score	0,67	0,81	0,71	0,80	0,80	0,76

The average accuracy of the AMD models is 0,89 which indicates that these results have a high level of correctness and are extremely reliable. The precision of the models combined are 0,77 which demonstrates that out of all positive predictions made, 77% of them were true. This value indicates that the results provide by the model is useful. The recall is 0,76 which indicates that the performance of the model is good. However, for the performance to be excellent, a

recall amount above 0,85 is required. The F1 score is 0,76 which indicates that the performance of the model is satisfactory. The accuracy metrics have the potential to improve if more data is provided to the model (Chandra & Bedi, 2021).

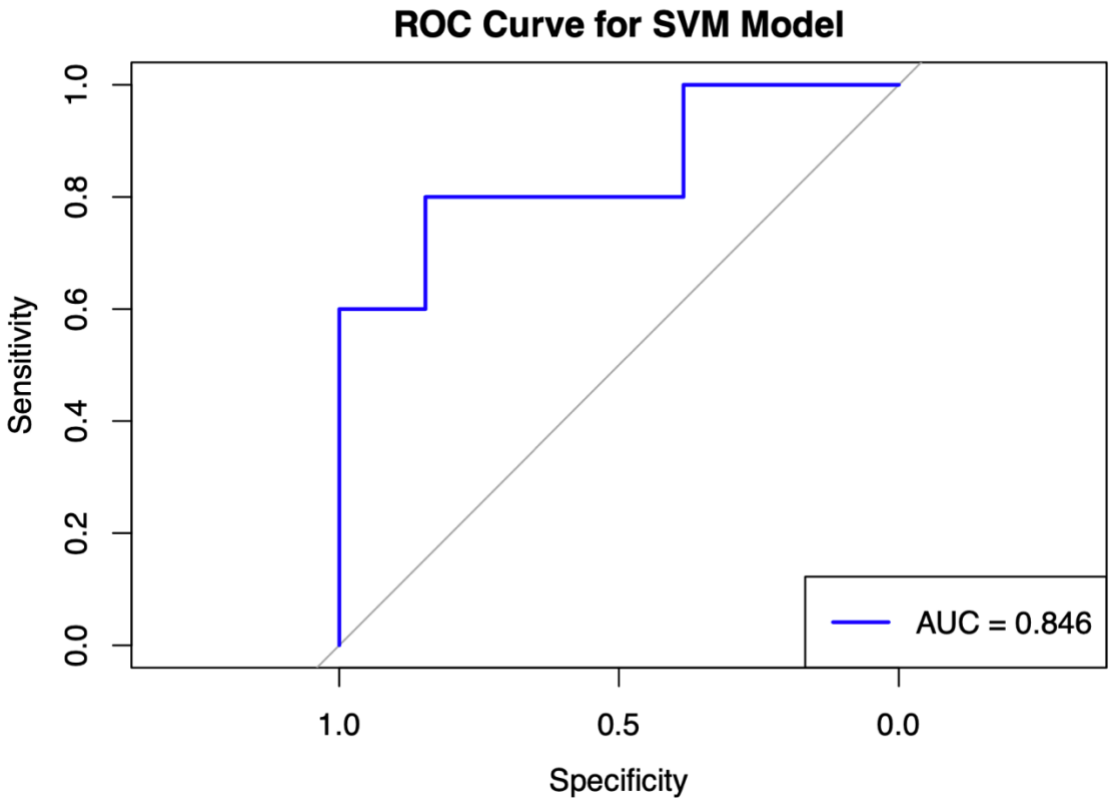


Figure 11: ROC curve and AUC value for predictive accuracy of SVM model

The Receiver Operating Characteristic (ROC) curve in Figure 11 was used to assess the predictive accuracy of the classification model. The ROC curve plots the true positive rate against the false positive rate at various threshold settings. The Area Under the Curve (AUC) value was calculated to be 0,846, indicating strong model performance. The model performance is calculated between 0-1, with an AUC value above 0,85 indicating strong classification performance (Narasimhan & Agarwal, 2017). A value of 0,846 is therefore, extremely satisfactory. This suggests that the model has a high ability to discriminate between contaminated and uncontaminated sites, supporting the reliability of the classification outputs.

5. DISCUSSION

5.1. INTERPRETATION OF RESULTS

It is important to note that a limitation of this study is the uneven spatial distribution of sampling sites relative to known pollution sources such as mines, tailings dams, and waste dumps. Due to constraints like access and safety, samples were more concentrated near visibly impacted or accessible areas. This may have introduced sampling bias, potentially skewing the model towards detecting contamination patterns associated with heavily affected zones while underrepresenting less-impacted areas. Without incorporating proximity to pollution sources as a variable, this bias could affect the generalisability of the machine learning predictions (A. Ramezan, A. Warner, & E. Maxwell, 2019). Future studies should consider stratified sampling and spatial cross-validation to reduce such bias and improve model robustness.

Based on the results obtained, there is no indication that severely contaminated water is entering the Hartbeespoort Dam. This conclusion is drawn from the presence of heavy metals in the surface soil that is in direct contact with the river. The observed decline in heavy metal concentrations downstream can be attributed to several factors. As the river progresses downstream, the flow rate decreases due to the gentle gradient of the topography and the absence of external hydrological inputs such as streams and tributaries feeding into the Crocodile River. Additionally, data collection was conducted during the dry season, during which the absence of rainfall prevents flooding and the subsequent transport of contaminants further downstream. Consequently, acid mine drainage (AMD) is detected only within the first four kilometres downstream.

In terms of spatial context, variations in land use and soil type also appear to influence contaminant levels. For instance, the spike in metal concentrations at site 4, despite the general downstream decline, likely reflects agricultural runoff. Similarly, anomalies observed at site 6 may be explained by differences in soil porosity and infiltration rate, leading to vertical migration of contaminants below the surface layer sampled (Malone, Minasny, Odgers, & McBratney, 2014). Elements such as copper, cobalt, and manganese, known for their mobility in acidic and low-clay soils, may have leached into subsoil layers, accounting for their unexpectedly low surface concentrations at this point (Delgado & Gomez, 2017).

Seasonal flooding has a large impact on the rate and distance at which AMD moves downstream because it is able to move over large distances before being diluted (Luo, et al., 2020). Flooding also causes an increase in erosion of soils that releases more soil acids into the water and contributes to the movement of other mine, agricultural, or industrial waste flowing into river systems that would not typically be present over dry months (Luo, et al., 2020). This impacts the results of the chemical concentrations being detected in AMD affected rivers.

Beyond hydrology, geochemical processes are central to explaining the observed patterns for elements such as sulphur, cobalt, and copper. For example, sulphur levels drop beyond site 2 likely due to a combination of sulphate reduction through waterlogged conditions, and precipitation of metal sulphides (Aguilar-Carrillo, Velázquez, Gutiérrez, & Reyes-Domínguez, 2022). Cobalt and copper, being more mobile in acidic environments, may be adsorbed onto clay particles or precipitated as hydroxides or carbonates as pH increases downstream (Delgado & Gomez, 2017). This also impacts communities that rely on the downstream rivers for water consumption.

From the results of this study, communities are not extremely affected by the contaminated water. The physical attributes such as the high levels of turbidity and discolouration of the water indicates that the water is not safe for consumption (Parbhakar-Fox & Lottermoser, 2015). The river also had a strong, unpleasant smell of sulphur which deters local residents from consuming or interacting with water from the river. There was also a high degree of education relating to the contamination levels across all sample sites as there were warnings to avoid physical interaction with the river at sample site 2. There is some degree of concern relating to the fact that no verbal warnings were issued at sample site 1. This site is the most contaminated and is easily accessible by the public as it is a part of a sculpture park. Sample points 3-6 are of little concern as AMD effects have decreased significantly by these points. Therefore, these points are regarded as safe by the standards of this study.

Currently there are no reports of health impacts within the communities around the study area, however, the physical interaction of people with the river at sample points 1 and 2 over extended periods of time could lead to the degradation of the aquatic environment due to death of species. The elevated amounts of cobalt and chromium in the river could affect the reproduction of species as well. The high amounts of copper also affects the respiratory function of organisms that live within the aquatic environment (Atibu, et al., 2018; Barrio-Parra, et al., 2018). The lower pH of the river at these sample points also makes the environment

unliveable for species residing in points 1 and 2 of the study area. As stated by Liao, et al. (2016) and Baloyi et al. (2023), other species could potentially consume fish from the river which could lead to harmful chemicals being passed up the food chain.

Cobalt has a major impact on communities. If there is any long term physical exposure to humans from points 1 and 2, they may experience cardiovascular, thyroid, and lung damage (Barrio-Parra, et al., 2018). Furthermore, the problematic amounts of chromium within the water at these two points may lead to an increased risk of cancer, skin rashes, respiratory issues and kidney issues (Atibu, et al., 2018; Karim, 2018). The elevated copper values at these points also leads to kidney issues among other problems such as liver and gastrointestinal complications (Atibu, et al., 2018; Karim, 2018). Copper as well as manganese have the potential to cause neurological damage that affects motor skills within people that consume contaminated water (Atibu, et al., 2018; Karim, 2018; Barrio-Parra, et al., 2018). The sulphur levels at the first two sample points also affect communities in a major way. This is through the introduction of headaches, nausea, and an irritated nervous system.

The levels of sulphur across the sample points are seen to continuously decrease along the study area. This is due to sulphuric acid being introduced to the river because of water and oxygen interacting with pyrite material (Skousen, Ziemkiewicz, & McDonald, 2019). The fact that this amount decreases over the sample points is due to the gentle gradient of the area. This gradient does not encourage the downstream movement of contaminants. The rest of the contaminants decrease steadily along all the sample points apart from sample points 4 and 6 where there are certain anomalies.

The elements studied in sample point 4 have extreme variability in concentrations compared to other study sites this is likely due to the difference in land use and land cover surrounding that region. Whereas sites 1-3 and 5-6 are primarily surrounded with grassland, site 4 is surrounded with a higher capacity of agricultural activity and built-up land. This means that waste from these activities largely influence the contamination that is present over site 4 (Moodley, Sheridan, Kappelmeyer, & Akcil, 2018). The concentrations of cobalt, chromium, and copper seem to have abnormally increased over point six whereas the general pattern shows these amounts decreasing over the sample points. This may be due to the introduction of chemicals from the small piece of agricultural land located close to sample point 6 or the difference in soil types for that area (Moodley, Sheridan, Kappelmeyer, & Akcil, 2018).

The extreme difference in contamination suggests that the soil type over that point could be more porous and have a high rate of infiltration. Therefore, elements with lower concentrations at this site may have moved further down within the soil. These elements would include copper, cobalt and manganese. This would mean that they are not extremely present on the surface at which sampling is taking place (Ben-Noah & Friedman, 2018; Xin, Tang, Liu, Zhang, & Tian, 2019). Because this study only accounts for surface soil, the decline of heavy metals over the region of interest may be due to the fact that the contaminants have infiltrated and migrated further below the surface.

Beyond these physical and chemical processes, land cover and slope further modulate contaminant behaviour. Steeper slopes upstream promote surface runoff and erosion. This facilitates the initial mobilisation of contaminants, whereas flatter areas downstream promote pooling, infiltration, and sedimentation (Yanites, 2018). This reduces contaminant transport distances. The spatial location of sample point 1 reinforces the idea that distance from the pollution source plays a critical role in determining contaminant loads, a pattern well established in AMD literature.

5.2. COMPARISON WITH LITERATURE

The study area is impacted by AMD as seen by the high amounts of sulphur being discharged from the Kromdraai mine. As detailed by Lemos, et al. (2020), sulphuric acid exits a mine due to the reaction that takes place between oxygen, water, and sulphur bearing rocks. This discharge is quantified by table 3, where it is shown that the measured sulphur amount surpasses the acceptable amount by almost 3000 mg/kg. This value exponentially decreases in the samples taken further downstream. As detailed by Nordstrom et al. (2015), AMD decreases in a exponential pattern when moving downstream due to the dilution of the contaminants. As contaminated water moves downstream, it interacts with a larger volumes of water bodies (Nordstrom, Blowes, & Ptacek, 2015). The types of contaminants also attribute to how water becomes neutralised over a distance. This is because heavy metals often settle out of the water as it flows. Heavy metals are prone to infiltrate soil, which reduces the contaminant concentrations further downstream (Nordstrom, Blowes, & Ptacek, 2015). The presence of vegetation around water sources may also reduce contaminant concentrations through the process of absorption. These contaminants are stored within plants through the process of bio-remediation, as previously detailed by Rambabu, et al. (2020). This is a likely process along the study area because access to the river is difficult through publicly available land. This

would mean that the environments are left undisturbed. Furthermore, the banks are populated with many large and uncontrolled plant species.

It is unlikely that the presence of these plants are actively apart of any remediation strategies as they are not well maintained. Seeing as Naidu, et al. (2019) has concluded that the cost of remediation would be extremely high based on remediation amounts in the US and Canada, it is extremely unlikely that a focus on remediation has taken place at this study area due to the isolated environment from large urban cities and other current mining intensive activities. Furthermore, mining operations were last active in this area over 100 years ago, in 1914, and there has been no reporting of AMD remediation efforts as of 2011 (The Department of Water Affairs, South Africa, 2011).

Since there has been no mining activity taking place in this region since 1914, there are no active mitigation strategies that are in place to prevent AMD in any capacity. This promotes the idea that abandoned mines are more responsible for contamination than active mines. This is because current mining operations try to actively monitor and account for any waste that is produced. This idea is presented by Baloyi et al. (2023) whereby it is identified that water quality monitoring is barely present within the Tweelopiespruit that flows into the Krugersdorp Game Reserve. On average, 20 and 60 ML of AMD contaminated water flows into the river stream a day in dry months and wet months respectively. This contamination originates from an abandoned gold mine. Currently only 12 ML is treated daily, which is less than half of the 27 ML of treatment needed to manage and maintain this river basin.

The findings of this study are particularly relevant within the context of the Bloubankspruit and Crocodile River systems, which have long been identified as environmentally vulnerable due to their proximity to legacy gold mining operations on the West Rand. As documented by Mvandaba et al. (2019), both rivers are affected by acid mine drainage (AMD), with the Bloubankspruit receiving acidic discharges from the Western Basin. This basin is a recognised AMD hotspot. The Crocodile River, similarly, exhibits elevated levels of sulphates and heavy metals due to AMD-contaminated tributaries. These systems represent the kinds of ecologically sensitive and poorly remediated sites that the current modelling approach tries to better understand and address.

By integrating SVM classification with XRF-derived geochemical data and spatial features, this study offers a spatially explicit method for contamination risk mapping, providing a

broader view than traditional point-based water sampling methods. While conventional sampling remains more precise in measuring dissolved contaminants, particularly in aquatic environments, the spatial model developed here adds value by highlighting hotspots of surface contamination and identifying areas for further investigation (Skousen, Ziemkiewicz, & McDonald, 2019). This complements existing water quality monitoring by offering a surface perspective on contamination and supports more efficient allocation of field resources. In AMD-impacted areas like Bloubankspruit and the Crocodile River, such models can enhance monitoring strategies by pinpointing where more detailed analysis is most needed.

The closure of mining activities are known to increase unemployment, crime, and substance abuse rates within areas that have previously relied on the mining industry for economic growth and employment (Siyongwana & Shabalala, 2019). This is not the case in the study area as this mine has not been operational for 111 years. Therefore, these social impacts are not currently present within the study area.

As per the previously mentioned study by Li et al. (2015), remediation of AMD areas takes an extremely long time. This is evident through the remediation efforts that have been ongoing since the 1980s in China, which is the world's largest producer of gold and coal. This means that the implementation of remediation strategies within the region of interest in this study is urgent as the process required intensive intervention and continuous monitoring and management over time. The act of deferring this process could lead to the further deterioration of the surrounding environment and a more exhaustive process of remediation in the future (Li, Jiao, Xiao, Chen, & Chang, 2015).

If an extensive amount of time passes before remediation can be implemented in the region, more severe impacts could affect the surrounding environment. For example, climate change could eventually cause parts of the river to run dry during winter months. According to Botezan, et al. (2020), remediation efforts would no longer be effective, especially wetland centred approaches, as the ecosystem would not be able to inhabit a dry environment. Although water may no longer transport contaminants downstream, AMD may still be present and permeate through the soil in close proximity to the mine in question. This could be detrimental as the transportation of AMD is not able to be tracked as effectively because it would infiltrate surface water (Botezan, et al., 2020).

5.3. ENVIRONMENTAL AND SPATIAL INFLUENCES

The environmental impacts of this study are mainly attributed to the land use and land cover patterns of the study area. Apart from AMD contaminants originating from the Kromdraai gold mine, agricultural waste is a factor that influences element concentrations particularly along points 4 and 6 of the sample sites. The other sites are all surrounded by grassland. It is therefore, inconclusive as to why agricultural waste or any other external waste has not affected sample point 5 due to the downstream transportation of contaminated material. The study area does not contain any wetlands as there are no facilities along the study area that maintain any type of wetland species. Furthermore, a study from The Department of Water Affairs, South Africa (2011) states that no remediation has yet taken place close to the Hartbeespoort Dam as of 2011. Because of this, it may be concluded that once contaminants have reached sample point 4, they infiltrate the soil so that contaminants are not present at the surface level of the next sample point.

Another consideration to account for would be the presence of anthropogenic activity that takes place over sample point 4. This may lead to municipal waste being disposed of within the river. Other small-scale pollution such as litter could be contributing to the contaminated state of the water as well. It is important to note that these amounts would not be extreme as it does not travel downstream as evident through the element readings at sample site 5.

5.4. MACHINE LEARNING ANALYSIS

The 60 sample points have gone through the SVM model 5 times to provide a robust cross classification model. The output of these results is extremely similar which may be attributed to the effectiveness of the model. The small amount of data that is available to train and test the model also attributes to this. The results from the cross-classification show that AMD contamination is present within the first 4-5 km of the mine before exponentially decreasing. This result is further validated by figure 7 in which the pre-classified points are weighted and plotted to indicate the rate at which AMD values decrease over all 6 sample points.

The SVM has a relatively high accuracy rate with an average of 0,85 across all models. The rest of the metrics such as precision, recall, and the F1 score all have satisfactory indicators at averages of 0,77, 0,76 and 0,76 respectively. These amounts are indicative of a well performing model, however, much more confidence could be instilled in this model if these values were greater. Although these metrics are satisfactory, they should ideally be above 0,85 (Chandra &

Bedi, 2021). These values could increase through the availability of more training and testing data. Adjusting thresholds for hyperplanes or mitigating the possibility of false positives and negatives could also improve these metrics (Machiwal, Cloutier, Güler, & Kazakis, 2018).

The process of incorrectly classifying positive and negative data points lies within the SVM model being over or under fitted (Lameski, Zdravevski, Mingov, & Kulakov, 2015). Overfitting the model means that the model is too complex, and too many parameters are being considered. Therefore, the model contains many points that are too close to the hyperplane that become incorrectly positively classified (Lameski, Zdravevski, Mingov, & Kulakov, 2015). The process of underfitting occurs when the model is too simple and is not able to capture the underlying pattern of the data. This leads to incorrect negative classifications. The model in this study does not have a high rate of false positives or negatives as it is generally advised that both should be kept under a 10-15% threshold in an environmental classification context (Lameski, Zdravevski, Mingov, & Kulakov, 2015). The 5 models average a 0,057% rate for false negatives and a 0,053% rate for false positives. This means that these values are of little to no concern.

While the SVM model applied in this study demonstrates satisfactory performance based on standard classification metrics, it is important to position these results within the broader context of environmental modelling research. Numerous studies have applied SVMs in geospatial contamination mapping and environmental monitoring which often highlights the model's strength in handling small datasets and high-dimensional spaces. However, these studies also emphasise key challenges, such as overfitting in low-sample contexts and reduced generalisability when external validation is absent (Mogashane, Maree, Mokoena, & Tshilongo, 2025). The relatively small dataset used in this study limits the complexity of the model and increases the risk of overfitting. Although cross validation was employed to mitigate this, the absence of an external dataset for validation may still restrict the model's predictive robustness in new or unobserved areas (Wei, et al., 2019; Mogashane, Maree, Mokoena, & Tshilongo, 2025). Furthermore, the variables taken into account may not fully capture all variables influencing AMD dispersion, such as hydrological fluctuations or seasonal changes. Incorporating additional environmental predictors or testing model performance against independent datasets from other AMD affected areas, would further strengthen the model's reliability and applicability in future studies.

5.5. IMPLICATIONS FOR REMEDIATION

Undocumented remediation methods within the study area may be responsible for the spatial decline of heavy metals. This is unlikely as a 2011 article by the Department of Water Affairs, South Africa mentions the importance of remediation at the Hartbeespoort dam. There is no mention of mitigating contamination by remediating rivers upstream from the dam. The department also acknowledges how many communities rely on the Crocodile River and has suggested that a pipeline from the Makolo Dam in Limpopo should be constructed to provide the communities around the Crocodile River with adequate access to water (The Department of Water Affairs, South Africa, 2011). This suggestion is not efficient for long term use. This is proven through the fact that after 13 years of proposing this idea, AMD is still present and severe within the first 4 km of the Kromdraai mine.

Proposed remediation strategies for sample points 1 and 2 would be a passive treatment such as constructed wetlands. This is because it is the most sustainable and cost-effective method. There is also evidence that this strategy has not been attempted yet due to the fact that there are no reported nature reserves nearby to manage and maintain the ecosystem (The Department of Water Affairs, South Africa, 2011). The only disadvantage of this method is that it has to be properly managed over time. Another passive system that can be implemented at sample points 1 and 2 would be the introduction of vertical flow systems. As previously mentioned by RoyChowdhury et al. (2015), layers of limestone, microbial populations and organic material neutralise water through chemical reactions and biological treatment. This treatment is also cost effective and sustainable, however, proper management needs to be implemented if this method is to be a long term solution (RoyChowdhury, Sarkar, & Datta, 2015; Naidu, et al., 2019).

If remediation occurs at sample points 1 and 2, there will automatically be less contaminants flowing further downstream. To extract the heavy metals that have most likely already settled within the soil and groundwater at points 3 to 6, another remediation process needs to be reviewed. This process could be the process of bioremediation. As previously mentioned in a study by Rambabu et al. (2020), this process is effective in the long term but is unsuitable for areas that need immediate intervention. This is not a concern for sample points 3-6 as the results of this study conclude that AMD is not severe along those points. This method works over both small and large areas but extensive research has to be done so that it is implemented around the environmental factors of the specific areas (Rambabu, et al., 2020). This should not be too

tedious to research over a 20 km study area. The process works by utilising plants and organisms to absorb heavy metals and contaminants from water and soil (Rambabu, et al., 2020).

These findings have clear implications for environmental management policy and intervention strategies in the Gauteng region. Firstly, the continued presence of AMD in the upper reaches of the Crocodile River highlights the urgent need to shift from conceptual planning to site-specific implementation of passive remediation systems. The spatial identification of hotspots, particularly sample points 1 and 2, allows policymakers to prioritise targeted intervention zones, thereby reducing resource wastage associated with large-scale blanket approaches. By incorporating low-cost sustainable methods like constructed wetlands and vertical flow systems, remediation becomes more feasible for government agencies with limited budgets (Pahl-Wostl, 2019). Moreover, the integration of machine learning models, as demonstrated in this study, provides a tool that can assist in locating contamination zones, forecasting pollutant spread, and evaluating remediation effectiveness over time. The findings also support the development of a spatially informed remediation policy that aligns with South Africa's National Water Act (Act No. 36 of 1998). This act mandates the protection of water resources and promotes public access to safe water (Pahl-Wostl, 2019).

5.6. STUDY LIMITATIONS AND FUTURE RESEARCH

This study has the potential to be improved by examining the depth extent of soil contamination. This could be done by extracting core samples at each sample point. Although this method is destructive to the environment, it is crucial in understanding how far heavy metals migrate within soils and groundwater. This provides a better understanding of the process of contamination and may help in concluding the best type of remediation for the area. Core sampling is also useful in identifying the differences in soil types over different areas. The benefit in doing this allows for a deeper understanding at what rate contaminants are able to infiltrate groundwater. This is done due to different levels of trafficability, aeration, and porosity of soils, which impacts the downward migration of contaminants (Delgado & Gomez, 2017)

Topography plays a central role in determining the spatial behaviour of AMD, particularly in controlling surface runoff, contaminant transport, and pooling of acidic water (Tomiyama & Igarashi, 2022). Elevation data allows researchers to predict where AMD-affected water may

accumulate or disperse, with steeper slopes typically promoting faster runoff and lower contamination retention, while flatter terrain encourages water pooling and contaminant accumulation (González, Cánovas, Olías, & Macías, 2020; Tomiyama & Igarashi, 2022). This makes terrain analysis using a DEM an essential component in AMD spatial modelling. In this study, a 30-meter resolution DEM was sourced from the United States Geological Survey (USGS) Earth Explorer, a publicly accessible platform providing Shuttle Radar Topography Mission (SRTM) data. While this DEM offers valuable regional coverage and ease of access, its 30 m spatial resolution imposes important limitations for detailed hydrological modelling in complex, anthropogenically altered landscapes such as the gold mining areas of Gauteng. Specifically, small topographic features like small drainage channels, mine tailings ridges, and constructed embankments may not be accurately captured at this scale (Modiegi, Rampedi, & Tesfamichael, 2020).

Furthermore, the vertical accuracy of SRTM data typically ranges around 10 meters, which can introduce uncertainty in derived slope and flow direction metrics (Malone, Minasny, Odgers, & McBratney, 2014). These resolution and accuracy constraints can lead to underestimation of contaminant retention zones or overgeneralisation of runoff patterns, thereby affecting the precision of machine learning models trained on these spatial predictors (Modiegi, Rampedi, & Tesfamichael, 2020). Therefore, while the DEM provides a sturdy foundation for AMD risk mapping, the results should be interpreted with caution. Ideally, future research would benefit from higher-resolution elevation data such as LiDAR or airborne photogrammetry (Modiegi, Rampedi, & Tesfamichael, 2020).

Another factor that could improve this research could be the migration and behaviour of contaminants during different seasons and climates. This could help in providing a better understanding of when and where AMD is most prevalent throughout the year. This may be due to flooding or the stagnant nature of river flow during summer and winter seasons. Smaller sample intervals could be helpful in this regard; however, the accessibility of the area provides a challenge in how small these intervals could potentially be. This is because a lot of the river is either only accessible through private land, or within a small valley. Large amounts of vegetation also cause data collection to be difficult and unsafe. In decreasing the interval sizes of sample sites, a larger dataset would naturally become available. This is useful in delivering more accurate and precise results of contamination levels from AMD sources.

The use of dry-season sampling affects the observed contaminant concentrations primarily for metals with solubility and transport dynamics sensitive to hydrological conditions. During the dry season, the absence of rainfall reduces surface runoff and limits the dilution of acidic mine water. This results in higher concentrations of certain metals in soil and water samples (Tomiyama & Igarashi, 2022). However, it also restricts the mobilisation of metals into aquatic systems, meaning the spatial extent of AMD during wetter seasons may be significantly underestimated.

This study has isolated the impacts of AMD over a small area so that the level of contamination from a singular decommissioned mine may be quantified. The province of Gauteng has an abundance of active and inactive gold mines that are located within extremely close proximities of each other all over the central and southeastern areas of the province. It also has many rivers running through these areas. Clark (2019) states that there are about 6000 abandoned mines within the whole of South Africa. Many of these are located around the golden arc, which is a series of mines that is located in Gauteng, the Free State, and the North West provinces. In order to fully grasp the extent of toxic waste within the river systems in Gauteng, the results of this research should be multiplied by about 6000 and take into account that many tailings flow into the same rivers. This increases the already toxic amounts of AMD within river and soil systems. The availability of healthcare data could improve this study as the impacts on local communities could be accurately reported. However, the healthcare data of this region is heavily restricted due to ethical concerns. It is also important to acknowledge that many cases are not reported due to the poverty-stricken nature of communities that surround and rely on water from contaminated sources. Healthcare data could drastically influence the urgency at which remediation strategies are implemented. This would effectively manage and mitigate damage to environments and communities.

Furthermore, the use of a relatively small dataset to train and validate the SVM model introduces spatial uncertainty in the resulting classification map. While the model shows high accuracy and low error margins in cross-validation, the limited number of sampling sites constrains the model's capacity to generalise beyond the sampled area. This can lead to an overestimation of certainty in predicted contamination zones and an underrepresentation of transitional areas between high and low contamination (Li, Zhang, Du, & Liu, 2017). In spatial modelling, especially when using machine learning, a lack of dense and diverse training data increases the risk of overfitting to localised patterns, reducing model transferability (Li, Zhang,

Du, & Liu, 2017). Future iterations of this study would benefit from incorporating spatial cross-validation techniques, uncertainty visualisation and higher-density sampling, particularly in underrepresented or geologically variable areas.

6. CONCLUSION

This study has found that AMD is extremely prevalent over the first 4 to 5 kilometres of decommissioned gold mines. This distance may increase based on how many mines contribute to heavy metals flowing into the rivers. Different sources of waste such as from agricultural and industrial activities also influence this. The XRF measurements indicate high levels of AMD contamination across points 1 and 2. This accounts for cobalt, chromium, copper, manganese, and sulphur values that exceed the acceptable amount within surface soil samples. Sample 4 contains anomaly values that do not conform to the pattern of contamination within element amounts. This increase in concentrations is likely attributed to the agricultural landscape alongside the sample point. All of the element measurements across all sample point do not exceed the health guidelines set by the WHO or SANS241 drinking water guidelines. This means that the physical environment is most at risk. Human health is not affected, and acute and chronic health issues are not of major concern at the current contamination levels.

The SVM model's performance is satisfactory in this study as it has an accuracy metric of 85%. The additional metrics such as precision, recall, and F1 score have values of 77%, 76%, and 76% respectively. This is satisfactory; however, higher percentages of these metrics would instil a higher level of confidence in the model (Chandra & Bedi, 2021). This could be improved with a larger dataset for training and testing. The SVM output a matrix per cross classification map in which it is seen that there are not many false positive or false negative values as these values each sit at around 0,05% of the dataset. Therefore, the robustness of cross classification as well as the performance of the classification demonstrates the benefits of using the SVM model for this study. Through using this model, it can be concluded that contamination is vigorously present over 4 km before dissipating into more reasonable values downstream.

The topography of a region can have major impacts on how AMD is able to move and spread downstream (Tomiyama & Igarashi, 2022). This study investigates a region with a constant gentle gradient, which means that stream flow does not occur at a high rate and contaminants are not readily transported downstream due to the topography making it difficult (Tomiyama & Igarashi, 2022). Flooding during wet seasons could, however, influence contaminants to be transported downstream in a rapid manner. Seasonal monitoring is important in this regard so that the transportation of contaminants due to the degree of flooding can be quantified (González, Cánovas, Olías, & Macías, 2020).

Seasons of heavy rainfall can contribute to the migration of surface contaminants downstream, however, the migration of contaminants deep within soil and water is inconclusive. Remediation methods are nevertheless recommended for areas downstream of rivers where surface contaminants are low because of the high likelihood that contamination is occurring further below the surface. Bioremediation is recommended for these areas, and passive systems are recommended for areas of high surface contamination. Bioremediation is the process in which harmful substances within soil is decreased through the absorption of plants and activity of micro-organisms (Morrissy, et al., 2021). A study by Majone, et al. (2015) outlines the process of bioremediation in Europe. They note that this process is highly promising, extremely cost effective, and sustainable for the surrounding environment. A high understanding of biological processes beneath the surface is however needed and there needs to be an active effort in maintaining the environment (Majone, et al., 2015). Passive systems are suitable for areas of high surface contamination because it is sustainable and cost effective, similarly to bioremediation. This process uses wetlands and the existing environment to reduce contamination through a natural approach. A study by Pat-Espadas et al. (2018) suggests that local vegetation should be considered when implementing passive treatment as it is the easiest to manage and has a higher survival rate in the local environment. Furthermore, the survival of wetland species is difficult to predict in changing climates, whereas local plants are already acclimatised to seasonal changes (Pat-Espadas, Loredo Portales, Amabilis-Sosa, Gómez, & Vidal, 2018).

The impact that AMD has on local communities and ecosystems include detriment to fish and aquatic organisms by affecting respiratory, reproductive, and neurological process. This is through elevated amounts of cobalt, chromium, copper, and manganese (Atibu, et al., 2018; Barrio-Parra, et al., 2018). The death of organisms is also sometimes present due to high amounts of sulphur. This causes the pH of the river to be unliveable for sensitive species (Likus-Cieślik, Smoliński, Pietrzykowski, & Bąk, 2019). A wide variety of these effects are also present in communities. A summation of elevated contaminant levels contains health issues within the cardiovascular, respiratory, neurological, kidney, liver, or gastrointestinal parts of human health. Another extreme would be the introduction of different types of cancers in people that are in constant contact with the contaminants over a long period of time (Atibu, et al., 2018; Barrio-Parra, et al., 2018). Minor effects could include headaches, nausea, and skin rashes, however these are more likely where there is only an elevated amount of sulphur and no other contaminants (Atibu, et al., 2018; Barrio-Parra, et al., 2018). There is no readily

available healthcare data for the region due to ethical reasons, therefore, the extent to which communities are already affected is unknown.

The main limitation of this study is the inability to extract core samples, which would provide a better understanding of the migration of heavy metals within soil. This is because permissions would most likely have to be granted as extracting these samples may cause damage to the environment. Furthermore, the sample points in question are located in tourist spots and privately owned lodges which need to be preserved. Another limitation to this study includes the small dataset in which the SVM model was trained. A more extensive dataset would have been ideal in outputting more accurate results and would improve the values of the performance metrics like precision, recall, and the F1 score. These suggestions may be integrated with future studies of a similar nature to improve the results.

This study set out to assess the spatial extent of AMD contamination from a decommissioned mine using XRF data and machine learning techniques, specifically SVM. In doing so, this study has successfully identified sample areas for AMD through a buffer analysis, screened the environment for hazardous chemicals. This was done through the use of XRF equipment. And lastly, identified wider contamination through the use of an SVM model. These objectives were successfully met. The integration of spatial data with SVM enabled the mapping of contamination zones, confirming that AMD is most severe within the first 4 to 5 kilometres downstream of the Kromdraai mine. The XRF data provided robust elemental readings, allowing for detailed analysis of heavy metal concentrations across sampling sites. By incorporating both spatial and geochemical insights, the study demonstrated the usefulness of machine learning in predicting contamination trends.

Additionally, the research highlighted environmental risks to aquatic ecosystems and potential latent risks to human health. This study not only reinforces the urgency of remediating abandoned mines but also supports the use of machine learning in environmental assessments to inform evidence-based policy and land management.

Beyond the general contamination patterns observed, this study identified outliers like the spike in arsenic concentrations at sample point 4, which likely reflect localised anthropogenic influences. These anomalies emphasise the spatial variety of contamination and challenge the viability of uniform remediation approaches across the region. Key quantitative findings show exceedances of sulphur by nearly 3000 mg/kg and significant concentrations of cobalt,

chromium, and copper within the first 4 km downstream, confirming that contamination is both intense and spatially constrained. The SVM model achieved a mean accuracy of 85%, with precision, recall, and F1 scores near 76–77%, and low misclassification rates indicating satisfactory but improvable performance (Sethy, Barpanda, Rath, & Behera, 2020). However, prediction uncertainty remains due to the limited dataset, potential overfitting, and coarse spatial resolution of DEM inputs. Despite these limitations, this study contributes to the growing body of research applying supervised machine learning to environmental spatial analysis, specifically demonstrating how SVM can effectively delineate AMD risk zones from limited soil sampling data.

These findings can inform targeted remediation planning and support scalable AMD risk assessments in other abandoned mining regions of South Africa. Future research should prioritise seasonal sampling to account for climatic variability in contaminant mobility, compare alternative classification algorithms such as Random Forest or ANN, and incorporate higher-resolution elevation data and hydrological modelling to better simulate contaminant transport pathways. Integrating these would not only refine prediction accuracy but also expand the policy relevance of machine learning for sustainable mine rehabilitation and water resource protection at national and regional scales.

7. REFERENCES

A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing, 11*(2), 185.

Acharya, B. S., & Kharel, G. (2020). Acid mine drainage from coal mining in the United States–An overview. *Journal of Hydrology, 588*, 125061.

Aguilar-Carrillo, J., Velázquez, A., Gutiérrez, E. J., & Reyes-Domínguez, I. A. (2022). Partitioning and mobility of arsenic (As) and lead (Pb) in an ancient Pb–Zn mine in central Mexico: role of amorphous ferric arsenate. *Applied Geochemistry, 136*, 105172.

Alengebawy, A., Abdelkhalek, S. T., Qureshi, S. R., & Wang, M. Q. (2021). Heavy metals and pesticides toxicity in agricultural soil and plants: Ecological risks and human health implications. *Toxics, 9*(3), 42.

Alvarenga, P., Guerreiro, N., Simões, I., Imaginário, M. J., & Palma, P. (2021). Assessment of the environmental impact of acid mine drainage on surface water, stream sediments, and macrophytes using a battery of chemical and ecotoxicological indicators. *Water, 13*(10), 1436.

Amponsah-Dacosta, F., & Mhlongo, S. E. (2016). A review of problems and solutions of abandoned mines in South Africa. *International Journal of Mining, Reclamation and Environment, 30*(4), 279-294.

Arabgol, R., Sartaj, M., & Asghari, K. (2016). Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model. *Environmental Modeling & Assessment, 21*(1), 71-82.

Assaad, H. I., Hou, Y., Zhou, L., Carroll, R. J., & Wu, G. (2015). Rapid publication-ready MS-Word tables for two-way ANOVA. *SpringerPlus, 4*(1), 33.

Atibu, E. K., Lacroix, P., Sivalingam, P., Ray, N., Giuliani, G., Mulaji, C. K., . . . Poté, J. (2018). High contamination in the areas surrounding abandoned mines and mining activities: An impact assessment of the Dilala, Luilu and Mpingiri Rivers, Democratic Republic of the Congo. *Chemosphere, 191*, 1008-1020.

Baloyi, J., Ramdhani, N., Mbhele, R., & Ramutshatsha- Makhwedzha, D. (2023). Recent Progress on Acid Mine Drainage Technological Trends in South Africa: Prevention, Treatment, and Resource Recovery. *Water, 15*(19), 3453.

Baloyi, J., Ramdhani, N., Mbhele, R., & Simate, G. S. (2024). Acid Mine Drainage from Gold Mining in South Africa: Remediation, Reuse, and Resource Recovery. *Mine Water and the Environment, 43*(3), 418-430.

Barrio-Parra, F., Elío, J., De Miguel, E., García-González, J. E., Izquierdo, M., & Álvarez, R. (2018). Environmental risk assessment of cobalt and manganese from industrial sources in an estuarine system. *Environmental geochemistry and health, 40*(2), 737-748.

Bartram, A. K., Jiang, X., Lynch, M. D., Masella, A. P., Nicol, G. W., Dushoff, J., & Neufeld, J. D. (2014). Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS microbiology ecology, 87*(2), 403-415.

Barzegar, R., Moghaddam, A. A., Deo, R., Fijani, E., & Tziritis, E. (2018). Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Science of the total environment, 621*, 697-712.

Ben-Noah, I., & Friedman, S. P. (2018). Review and evaluation of root respiration and of natural and agricultural processes of soil aeration. *Vadose Zone Journal, 17*(1), 1-47.

Bezabih, T. D., Glaety, M. G., Wako, D. A., & Worku, S. G. (2024). Geospatial data analysis: A comprehensive overview of python libraries and implications. *Ethics, Machine Learning, and Python in Geospatial Analysis*, 72-93.

Bischoff-Mattson, Z., Maree, G., Vogel, C., Lynch, A., Olivier, D., & Terblanche, D. (2020). Shape of a water crisis: Practitioner perspectives on urban water scarcity and 'Day Zero'in South Africa. *Water Policy, 22*(2), 193-210.

Bobbins, K. (2015). *Acid Mine Drainage and its Governance in the Gauteng City-Region.* Johannesburg: Gauteng City-Region Observatory.

Botezan, C., Constantin, V., Meltzer, M., Radovici, A., Pop, A., Alexandrescu, F., & Stefanescu, L. (2020). Is there sustainable development after mining? a case study of three mining areas in the apuseni region (Romania). *Sustainability, 12*(23), 9791.

Çakir, M., Yilmaz, M., Oral, M. A., Kazanci, H. Ö., & Oral, O. (2023). Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture. *Journal of King Saud University–Science, 35*.

Chalkley, R. (2024). Geospatial monitoring and modelling of acid mine drainage and mining footprints.

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. *International Journal of Information Technology, 13*(5), 1-11.

Che, N. S., Bett, S., Okpara, E. C., Olagbaju, P. O., Fayemi, O. E., & Mathuthu, M. (2021). An assessment of land use and land cover changes and its impact on the surface water

quality of the Crocodile River catchment, South Africa. *River deltas research-recent advances*, 29.

Chen, D., Feng, Q., & Liang, H. (2021). Effects of long-term discharge of acid mine drainage from abandoned coal mines on soil microorganisms: Microbial community structure, interaction patterns, and metabolic functions. *Environmental Science and Pollution Research, 28*(38), 53936-53952.

Chen, L. X., Huang, L. N., Méndez-García, C., Kuang, J. L., Hua, Z. S., Liu, J., & Shu, W. S. (2016). Microbial communities, processes and functions in acid mine drainage ecosystems. *Current opinion in biotechnology, 38*, 150-158.

Chen, W., Geng, Y., Hong, J., Dong, H., Cui, X., Sun, M., & Zhang, Q. (2018). Life cycle assessment of gold production in China. *Journal of Cleaner Production, 179*, 143-150.

Clark, C. (2019, October 24). *'There's a lot of money down there': the deadly cities of gold beneath Johannesburg*. Retrieved from The Guardian: https://www.theguardian.com/cities/2019/oct/24/theres-a-lot-of-money-down-there-the-deadly-cities-of-gold-beneath-johannesburg

Curtis, C. (2021). AMD. *Quest, 17*(2), 9.

Danades, A., Pratama, D., Anggraini, D., & Anggriani, D. (2016). Comparison of Accuracy Level K-Nearest Neighbor Algorithm and Support Vector Machine Algorithm in Classification Water Quality Status. *6th International conference on system engineering and technology (ICSET)*, 137-141.

Declercq, Y., Delbecque, N., De Grave, J., De Smedt, P., Finke, P., Mouazen, A. M., . . . Verdoodt, A. (2019). A comprehensive study of three different portable XRF scanners to assess the soil geochemistry of an extensive sample dataset. *Remote Sensing, 11*(21), 2490.

Delgado, A., & Gomez, J. A. (2017). The Soil. Physical, Chemical and Biological Properties. *Principles of Agronomy for Sustainable Agriculture*, 15-30.

Department of Environmental Affairs, Republic of South Africa. (2010). *Framework for the Management of Contaminated Land.* Johannesburg: WSP Environment & Energy.

Dold, B. (2014). Evolution of acid mine drainage formation in sulphidic mine tailings. *Minerals, 4*(3), 621-641.

Edmunds, W. M., Ahmed, K. M., & Whitehead, P. G. (2015). A review of arsenic and its impacts in groundwater of the Ganges–Brahmaputra–Meghna delta, Bangladesh. *Environmental Science: Processes & Impacts, 17*(6), 1032-1046.

Edwards, D. P., Sloan, S., Weng, L., Dirks, P., Sayer, J., & Laurance, W. F. (2014). Mining and the African environment. *Conservation Letters, 7*(3), 302-311.

Emerson, R. W. (2022). ANOVA assumptions. *Journal of Visual Impairment & Blindness, 116*(4), 585-586.

Ezeora, O. S., Heckenbergerova, J., & Musilek, P. (2015). Statistical analysis of environmental measurements for design of energy-efficient monitoring systems. *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, 1143-1148.

Fernández-Caliani, J. C., Giráldez, M. I., & Barba-Brioso, C. (2019). Oral bioaccessibility and human health risk assessment of trace elements in agricultural soils impacted by acid mine drainage. *Chemosphere, 237*, 124441.

Galhardi, J. A., & Bonotto, D. M. (2016). Hydrogeochemical features of surface water and groundwater contaminated with acid mine drainage (AMD) in coal mining areas: a case study in southern Brazil. *Environmental Science and Pollution Research, 23*(18), 18911-18927.

Galván, L., Olías, M., Cánovas, C. R., Sarmiento, A. M., & Nieto, J. M. (2016). Hydrological modeling of a watershed affected by acid mine drainage (Odiel River, SW Spain). Assessment of the pollutant contributing areas. *Journal of Hydrology, 540*, 196-206.

Gharakhanlou, N. M., & Perez, L. (2023). Flood susceptible prediction through the use of geospatial variables and machine learning methods. *Journal of hydrology, 617*, 129121.

González, R. M., Cánovas, C. R., Olías, M., & Macías, F. (2020). Seasonal variability of extremely metal rich acid mine drainages from the Tharsis mines (SW Spain). *Environmental Pollution, 259*, 113829.

Guo, L., & Cutright, T. J. (2017). Metal storage in reeds from an acid mine drainage contaminated field. *International Journal of Phytoremediation, 19*(3), 254-261.

Henson, R. N. (2015). Analysis of variance (ANOVA). *Brain mapping, 1*, 477-481.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma, 265*, 62-77.

Hobbs, P. J. (2017). TDS load contribution from acid mine drainage to Hartbeespoort Dam, South Africa. *Water sa, 43*(4), 626-637.

Hsieh, W. W. (2022). Evolution of machine learning in environmental science—A perspective. *Environmental Data Science, 1*, e3.

Hu, P., Yang, B., Dong, C., Chen, L., Cao, X., Zhao, J., . . . Christie, P. (2014). Assessment of EDTA heap leaching of an agricultural soil highly contaminated with heavy metals. *Chemosphere, 117*, 532-537.

Ighalo, J. O., Kurniawan, S. B., Iwuozor, K. O., Aniagor, C. O., Ajala, O. J., Oba, S. N., . . . Igwegbe, C. A. (2022). A review of treatment technologies for the mitigation of the toxic environmental effects of acid mine drainage (AMD). *Process Safety and Environmental Protection, 157*, 37-58.

Jensen, D. (2015). Rapid analysis of secondary manganese oxides using portable XRF: Applications to environmental geochemistry and mineral exploration. *Master's thesis, Montana Tech of The University of Montana*.

Jiang, J., Wang, Y. P., Yu, M., Cao, N., & Yan, J. (2018). Soil organic matter is important for acid buffering and reducing aluminum leaching from acidic forest soils. *Chemical Geology, 501*, 86-94.

Karaouzas, I. (2018). Agro-industrial wastewater pollution in Greek river ecosystems. *The Rivers of Greece: Evolution, Current Status and Perspectives*, 169-204.

Karim, N. (2018). Copper and human health-a review. *Journal of Bahria University Medical and Dental College, 8*(2), 117-122.

Kgotso, S. (2018). *The effect of heavy metals associated with Acid Mine Drainage (AMD) on select aquatic insects in the Tweelopiespruit, Rietspruit and Bloubankspruit river system on the West Rand, Gauteng.* Johannesburg: University of Johannesburg.

Khanyile, S. (2024). Renaturing Gauteng's mining wastescapes? . *Gauteng City-Region Observatory*.

Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean journal of anesthesiology, 70*(1), 22.

Komonweeraket, K., Cetin, B., Aydilek, A. H., Benson, C. H., & Edil, T. B. (2015). Effects of pH on the leaching mechanisms of elements from fly ash mixed soils. *Fuel, 140*, 788-802.

Lameski, P., Zdravevski, E., Mingov, R., & Kulakov, A. (2015). SVM parameter tuning with grid search and its impact on reduction of model over-fitting. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference*, 464-474.

Lemos, M., Valente, T., Reis, P. M., Fonseca, R., Delbem, I., Ventura, J., & Magalhães, M. (2020). Mineralogical and geochemical characterization of gold mining tailings and their potential to generate acid mine drainage. *Minerals, 11*(1), 39.

Li, J., Zhang, Y., Du, D., & Liu, Z. (2017). Improvements in the decision making for Cleaner Production by data mining: Case study of vanadium extraction industry using weak acid leaching process. *Journal of Cleaner Production, 143*, 582-597.

Li, X. N., Jiao, W. T., Xiao, R. B., Chen, W. P., & Chang, A. C. (2015). Soil pollution and site remediation policies in China: A review. *Environmental Reviews, 23*(3), 263-274.

Liao, J., Wen, Z., Ru, X., Chen, J., Wu, H., & Wei, C. (2016). Distribution and migration of heavy metals in soil and crops affected by acid mine drainage: Public health implications in Guangdong Province, China. *Ecotoxicology and environmental safety, 124*, 460-469.

Likus-Cieślik, J., Smoliński, A., Pietrzykowski, M., & Bąk, A. (2019). Sulphur contamination impact on seasonal and surface water chemistry on a reforested area of a former sulphur mine. *Land Degradation & Development, 30*(2), 212-225.

Liu, X., Bai, Z., Shi, H., Zhou, W., & Liu, X. (2019). Heavy metal pollution of soils from coal mines in China. *Natural Hazards, 99*(2), 1163-1177.

Luo, C., Routh, J., Dario, M., Sarkar, S., Wei, L., Luo, D., & Liu, Y. (2020). Distribution and mobilization of heavy metals at an acid mine drainage affected region in South China, a post-remediation study. *Science of the Total Environment, 724*, 138122.

Machiwal, D., Cloutier, V., Güler, C., & Kazakis, N. (2018). A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environmental Earth Sciences, 77*(19), 681.

Majone, M., Verdini, R., Aulenta, F., Rossetti, S., Tandoi, V., Kalogerakis, N., . . . Fava, F. (2015). In situ groundwater and sediment bioremediation: barriers and perspectives at European contaminated sites. *New biotechnology, 32*(1), 133-146.

Malone, B. P., Minasny, B., Odgers, N. P., & McBratney, A. B. (2014). Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma, 232*, 34-44.

Martín-Crespo, T., Gómez-Ortiz, D., Martín-Velázquez, S., Martinez-Pagan, P., De Ignacio, C., Lillo, J., & Faz, A. (2018). Geoenvironmental characterization of unstable abandoned mine tailings combining geophysical and geochemical methods (Cartagena-La Union district, Spain). *Engineering Geology, 232*, 135-146.

McIntosh, K. G., Guimarães, D., Cusack, M. J., Vershinin, A., Chen, Z. W., Yang, K., & Parsons, P. J. (2016). Evaluation of portable XRF instrumentation for assessing potential environmental exposure to toxic elements. *International journal of environmental analytical chemistry, 96*(1), 15-37.

Mejía-Piña, K. G., Huerta-Diaz, M. A., & González-Yajimovich, O. (2016). Calibration of handheld X-ray fluorescence (XRF) equipment for optimum determination of elemental concentrations in sediment samples. *Talanta, 161*, 359-367.

Mhlongo, S. E. (2023). Physical hazards of abandoned mines: A review of cases from South Africa. *The Extractive Industries and Society, 15*, 101285.

Mim, M. A., & Shawkat Zamil, K. A. (2018). GIS-Based Analysis of Changing Surface Water in Rajshahi City Corporation Area Using Support Vector Machine (SVM), Decision Tree & Random Forest Technique. *Machine Learning Research, 3*(2), 11-17.

Minerals Council South Africa. (2024). *Facts and Figures Pocketbook 2024.* Minerals Council South Africa.

Minnaar, A. (2020). Water pollution and contamination from gold mines: acid mine drainage in Gauteng Province, South Africa. In Water, governance, and crime issues. *Springer International Publishing*, 193-219.

Modiegi, M., Rampedi, I. T., & Tesfamichael, S. G. (2020). Comparison of multi-source satellite data for quantifying water quality parameters in a mining environment. *Journal of Hydrology, 591*, 125322.

Mogashane, T. M., Maree, J. P., Mokoena, L., & Tshilongo, J. (2025). Research Activities on Acid Mine Drainage Treatment in South Africa (1998–2025): Trends, Challenges, Bibliometric Analysis and Future Directions. *Water, 17*(15), 2286.

Moodley, I., Sheridan, C. M., Kappelmeyer, U., & Akcil, A. (2018). Environmentally sustainable acid mine drainage remediation: Research developments with a focus on waste/by-products. *Minerals Engineering, 126*, 207-220.

Morrissy, J. G., Currell, M. J., Reichman, S. M., Surapaneni, A., Megharaj, M., Crosbie, N. D., . . . Ball, A. S. (2021). Nitrogen contamination and bioremediation in groundwater and the environment: A review. *Earth-Science Reviews, 222*, 103816.

Mosley, L. M., Fitzpatrick, R. W., Palmer, D., Leyden, E., & Shand, P. (2014). Changes in acidity and metal geochemistry in soils, groundwater, drain and river water in the Lower Murray River after a severe drought. *Science of the Total Environment, 485*, 281-291.

Mujuru, M. (2014). Management and Mitigation of Acid Mine Drainage (AMD) in South Africa: Lessons for mineral Africa. *Africa Institute of South Africa*, 1-6.

Müller, A., Österlund, H., Marsalek, J., & Viklander, M. (2020). The pollution conveyed by urban runoff: A review of sources. *Science of the Total Environment, 709*, 136125.

Mvandaba, V., Kapangaziwiri, E., & Shadung, J. (2019). SURFACE WATER AND GROUNDWATER RESOURCES MONITORING, CRADLE OF HUMANKIND WORLD HERITAGE SITE, GAUTENG PROVINCE, SOUTH AFRICA.

Myers, T. (2016). Acid mine drainage risks–A modeling approach to siting mine facilities in Northern Minnesota USA. *Journal of Hydrology, 533*, 277-290.

Nachana'a Timothy, E. T. (2019). Environmental pollution by heavy metal: an overview. *Chemistry, 3*(2), 72-82.

Naidoo, S. (2015). An assessment of the impacts of acid mine drainage on socio-economic development in the Witwatersrand: South Africa. *Environment, development and sustainability, 17*(5), 1045-1063.

Naidu, G., Ryu, S., Thiruvenkatachari, R., Choi, Y., Jeong, S., & Vigneswaran, S. (2019). A critical review on remediation, reuse, and resource recovery from acid mine drainage. *Environmental pollution, 247*, 1110-1124.

Narasimhan, H., & Agarwal, S. (2017). Support vector algorithms for optimizing the partial area under the ROC curve. *Neural computation, 29*(7), 1919-1963.

Ngole-Jeme, V. M., & Ndava, J. (2023). The implications of AMD induced acidity, high metal concentrations and ochre precipitation on aquatic organisms. *Polish Journal of Environmental Studies, 32*(4), 2959.

Nguyen, T. T., Soda, S., Kanayama, A., & Hamai, T. (2021). Effects of cattails and hydraulic loading on heavy metal removal from closed mine drainage by pilot-scale constructed wetlands. *Water, 13*(14), 1937.

Nguyen, T. T., Zhang, W., Li, Z., Li, J., Ge, C., Liu, J., & Yu, L. (2016). Assessment of heavy metal pollution in Red River surface sediments, Vietnam. *Marine pollution bulletin, 113*(1-2), 513-519.

Nordstrom, D. K., Blowes, D. W., & Ptacek, C. J. (2015). Hydrogeochemistry and microbiology of mine drainage: An update. *Applied Geochemistry, 57*, 3-16.

Oberholster, P. J., Schoeman, Y., Truter, J. C., & Botha, A. M. (2022). Using periphyton assemblage and water quality variables to assess the ecological recovery of an ecologically engineered wetland affected by acid mine drainage after a dry spell. *Processes, 10*(5), 877.

Obi, F. O., Ugwuishiwu, B. O., & Nwakaire, J. N. (2016). Agricultural waste concept, generation, utilization and management. *Nigerian Journal of Technology, 35*(4), 957-964.

Ott, W. R. (2018). Environmental statistics and data analysis. *Routledge*.

Oyedotun, T. D. (2018). X-ray fluorescence (XRF) in the investigation of the composition of earth materials: a review and an overview. *Geology, Ecology, and Landscapes, 2*(2), 148-154.

Pahl-Wostl, C. (2019). The role of governance modes and meta-governance in the transformation towards sustainable water governance. *Environmental science & policy, 91*, 6-16.

Parbhakar-Fox, A., & Lottermoser, B. G. (2015). A critical review of acid rock drainage prediction methods and practices. *Minerals Engineering, 82*, 107-124.

Park, F. (2014). *Heritage Impact Assessment Report.* Johannesburg: South African Heritage Resources Agency.

Pat-Espadas, A. M., Loredo Portales, R., Amabilis-Sosa, L. E., Gómez, G., & Vidal, G. (2018). Review of constructed wetlands for acid mine drainage treatment. *Water, 10*(11), 1685.

Patel, D. D., Mehta, D. J., Azamathulla, H. M., Shaikh, M. M., Jha, S., & Rathnayake, U. (2023). Application of the weighted arithmetic water quality index in assessing groundwater quality: A case study of the South Gujarat region. *Water, 15*(19), 1-10.

Potts, P. J., & Sargent, M. (2022). In situ measurements using hand-held XRF spectrometers: a tutorial review. *Journal of Analytical Atomic Spectrometry, 37*(10), 1928-1947.

Pytlakowska, K. (2016). Speciation of inorganic chromium in water samples by energy dispersive X-ray fluorescence spectrometry. *Journal of Analytical Atomic Spectrometry, 31*(4), 968-974.

Rahimoon, S. U., Yamamoto, K., Ghafar, M. A., Anwar, M. M., Majeed, M., Khattak, W. A., & Muhammad, M. (2025). Exploration of science of remote sensing and GIS with GEE. *Google Earth Engine and Artificial Intelligence for Earth Observation*, 49-76.

Raji, I. B., Hoffmann, E., Ngie, A., & Winde, F. (2021). Assessing Uranium Pollution Levels in the Rietspruit River, Far West Rand Goldfield, South Africa. *International Journal of Environmental Research and Public Health, 18*(16), 8466.

Rambabu, K., Banat, F., Pham, Q. M., Ho, S. H., Ren, N. Q., & Show, P. L. (2020). Biological remediation of acid mine drainage: Review of past trends and current outlook. *Environmental Science and Ecotechnology, 2*, 100024.

Rimayi, C., Odusanya, D., Weiss, J. M., de Boer, J., & Chimuka, L. (2018). Contaminants of emerging concern in the Hartbeespoort Dam catchment and the uMngeni River estuary 2016 pollution incident, South Africa. *Science of the Total Environment, 627*, 1008-1017.

RoyChowdhury, A., Sarkar, D., & Datta, R. (2015). Remediation of acid mine drainage-impacted water. *Current Pollution Reports, 1*(3), 131-141.

Sakala, E., Fourie, F., Gomo, M., & Coetzee, H. (2019). Groundwater vulnerability mapping of Witbank coalfield in South Africa using deep learning artificial neural networks. *South African Journal of Geomatics, 8*(2), 282-293.

Sankhla, M. S., Kumari, M., Nandan, M., Kumar, R., & Agrawal, P. (2016). Heavy metals contamination in water and their hazardous effect on human health-a review. *Int. J. Curr. Microbiol. App. Sci, 5*(10), 759-766.

Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture, 175*, 105527.

Shakoor, M. B., Nawaz, R., Hussain, F., Raza, M., Ali, S., Rizwan, M., . . . Ahmad, S. (2017). Human health implications, risk assessment and remediation of As-contaminated water: a critical review. *Science of the Total Environment, 601*, 756-769.

Sharma, N., Saharia, M., & Ramana, G. V. (2024). High resolution landslide susceptibility mapping using ensemble machine learning and geospatial big data. *Catena, 235*, 107653.

Sharma, S., & Sharma, V. (2016). Performance of various machine learning classifiers on small datasets with varying dimensionalities: A study. *Circulation in Computer Science, 1*(1), 30-35.

Siebecker, M. G., Chaney, R. L., & Sparks, D. L. (2018). Natural speciation of nickel at the micrometer scale in serpentine (ultramafic) topsoils using microfocused X-ray fluorescence, diffraction, and absorption. *Geochemical Transactions, 19*(1), 14.

Simate, G. S., & Ndlovu, S. (2014). Acid mine drainage: Challenges and opportunities. *Journal of Environmental Chemical Engineering, 2*(3), 1785-1803.

Siyongwana, P. Q., & Shabalala, A. (2019). The socio-economic impacts of mine closure on local communities: evidence from Mpumalanga Province in South Africa. *GeoJournal, 84*(2), 367-380.

Skousen, J. G., Ziemkiewicz, P. F., & McDonald, L. M. (2019). Acid mine drainage formation, control and treatment: Approaches and strategies. *The Extractive Industries and Society, 6*(1), 241-249.

Solanki, Y. S., Agarwal, M., Gupta, A. B., Gupta, S., & Shukla, P. (2022). Fluoride occurrences, health problems, detection, and remediation methods for drinking water: A comprehensive review. *Science of the Total Environment, 807*, 150601.

South African Bureau of Standards. (2022, April 28). SANS 241 Drinking Water. Pretoria, Gauteng, South Africa.

Steyn, M., Oberholster, P. J., Botha, A. M., Genthe, B., Van den Heever-Kriek, P. E., & Weyers, C. (2019). Treated acid mine drainage and stream recovery: Downstream impacts on T benthic macroinvertebrate communities in relation to multispecies toxicity bioassays. *Journal of environmental management, 235*, 377-388.

Stoker, P., Tian, G., & Kim, J. Y. (2020). Analysis of variance (ANOVA). *Basic quantitative research methods for urban planners*, 197-219.

Szramowiat-Sala, K. (2023). Artificial intelligence in environmental monitoring: application of artificial neural networks and machine learning for pollution prevention and toxicity measurements.

Talukdar, B., Kalita, H. K., Basumatary, S., Saikia, D. J., & Sarma, D. (2017). Cytotoxic and genotoxic affects of acid mine drainage on fish Channa punctata (Bloch). *Ecotoxicology and environmental safety, 144*, 72-78.

The Department of Environmental Affairs, Republic of South Africa. (2010). *Framework for the Management of Contaminated Land.* WSP Environment & Energy.

The Department of Water Affairs, South Africa. (2011, July 14). *Water supply in the Crocodile West River system is well planned*. Retrieved from South African Government: https://www.gov.za/news/media-statements/water-supply-crocodile-west-river-system-well-planned-14-jul-2011

Thompson, L. (2014). A history of South Africa. *Yale University Press*.

Tighe, M., Rogan, G., Wilson, S. C., Grave, P., Kealhofer, L., & Yukongdi, P. (2018). The potential for portable X-ray fluorescence determination of soil copper at ancient metallurgy sites, and considerations beyond measurements of total concentrations. *Journal of environmental management, 206*, 373-382.

Tomiyama, S., & Igarashi, T. (2022). The potential threat of mine drainage to groundwater resources. *Current Opinion in Environmental Science & Health, 27*, 100347.

Turcotte, M. A., Jamieson, H. E., & Parsons, M. B. (2025). Mobility of cobalt in mine waste: Evidence from a historic silver mining district in Canada. *Applied Geochemistry, 181*, 106296.

Van Groeningen, N., Christl, I., & Kretzschmar, R. (2021). The effect of aeration on Mn (II) sorbed to clay minerals and its impact on Cd retention. *Environmental Science & Technology, 55*(3), 1650-1658.

Veerasingam, S., Vethamony, P., Murali, R. M., & Fernandes, B. (2015). Depositional record of trace metals and degree of contamination in core sediments from the Mandovi estuarine mangrove ecosystem, west coast of India. *Marine pollution bulletin, 91*(1), 362-367.

Vicente-Beckett, V. A., Taylor McCauley, G. J., & Duivenvoorden, L. J. (2016). Metal speciation in sediments and soils associated with acid-mine drainage in Mount Morgan (Queensland, Australia). *Journal of Environmental Science and Health, 51*(2), 121-134.

Vöröš, D., Geršlová, E., Díaz-Somoano, M., Sýkorová, I., Suárez-Ruiz, I., Havelcová, M., & Kuta, J. (2018). Distribution and mobility potential of trace elements in the main seam of the Most Coal Basin. *International Journal of Coal Geology, 196*, 139-147.

Wei, L., Yuan, Z., Zhong, Y., Yang, L., Hu, X., & Zhang, Y. (2019). An improved gradient boosting regression tree estimation model for soil heavy metal (Arsenic) pollution monitoring using hyperspectral remote sensing. *Applied Sciences, 9*(9), 1943.

Wieder, W. L., & Shoop, S. A. (2018). State of the knowledge of vegetation impact on soil strength and trafficability. *Journal of terramechanics, 78*, 1-14.

Williams, K. M., & Turner, A. M. (2015). Acid mine drainage and stream recovery: Effects of restoration on water quality, macroinvertebrates, and fish. *Knowledge and Management of Aquatic Ecosystems*(416), 18.

Worlanyo, A. S., & Jiangfeng, L. (2021). Evaluating the environmental and economic impact of mining for post-mined land restoration and land-use: A review. *Journal of Environmental Management, 279*, 111623.

World Health Organisation. (2011). *Guidelines for Drinking-water Quality.* Geneva: World Health Organisation.

Wu, C. (2021). Acid mine drainage prediction techniques and geochemical modelling: case study on gold tailing dams, West Rand, Witwatersrand basin area, South Africa.

Xin, J., Tang, J., Liu, Y., Zhang, Y., & Tian, R. (2019). Pre-aeration of the rhizosphere offers potential for phytoremediation of heavy metal-contaminated wetlands. *Journal of Hazardous Materials, 374*, 437-446.

Yang, Y., Christakos, G., Guo, M., Xiao, L., & Huang, W. (2017). Space-time quantitative source apportionment of soil heavy metal concentration increments. *Environmental Pollution, 223*, 560-566.

Yanites, B. J. (2018). The dynamics of channel slope, width, and sediment in actively eroding bedrock river systems. *Journal of Geophysical Research: Earth Surface, 123*(7), 1504-1527.

Young, K. E., Evans, C. A., Hodges, K. V., Bleacher, J. E., & Graff, T. G. (2016). A review of the handheld X-ray fluorescence spectrometer as a tool for field geologic investigations on Earth and in planetary surface exploration. *Applied Geochemistry, 72*, 77-87.

Zhan, Y., Yang, M., Zhang, S., Zhao, D., Duan, J., Wang, W., & Yan, L. (2019). Iron and sulfur oxidation pathways of Acidithiobacillus ferrooxidans. *World Journal of Microbiology and Biotechnology, 35*(4), 60.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., . . . Shi, W. (2021). Machine learning: new ideas and tools in environmental science and engineering. *Environmental science & technology, 55*(19), 12741-12754.

8. APPENDIX

8.1 ETHICS WAIVER

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
Registration number: REC-101114-044

26 September 2024

Re: Ms Nikita Bullah (2303383)

Waiver letter number: HRECNMW24/10/06

To whom it may concern,

Ms. Bullah is a registered masters' student in the School of Geography, Archaeology, and Environmental Studies at the University of the Witwatersrand, Johannesburg. This letter is to confirm that, at the time of writing, Ms. Bullah does not need ethical clearance for the study entitled *'Using machine learning to identify the spatial extent of acid mine drainage from gold mining activities in Gauteng, South Africa'*. This decision has been reached based upon a description of the project supplied by Ms. Bullah to the University Human Research Ethics Committee (Non-Medical), which has been evaluated by the Chairs and Deputy Chairs. If, however, Ms. Bullah changes the methods of data collection and analysis for this study, this decision may no longer be valid. If such changes take place, this should be communicated to the University Human Research Ethics Committee (Non-Medical) as soon as possible. This waiver letter is valid until 25 September 2027.

Please feel free to contact me should you require any further information.

Thank you.

Yours sincerely,
S Schoeman

Shaun Schoeman (Administrative Officer)
Solomon Mahlangu House, 10th Floor, Room 10004, Jorissen Street, Braamfontein, Johannesburg
Private Bag 3, Wits 2050
T + 27(0)11 717 1408 | **E** Shaun.Schoeman@wits.ac.za | hrec-medical.researchoffice@wits.ac.za
www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

8.2 RAW XRF COLLECTED DATA

Sample pt.	X latitude	Y longitude	As	Co	Cr	Cu	Mn	Ni	Pb	S	Zn
1	**27.77892136**	**-25.9868753**	0,00	0,00	0,00	17,00	95,00	0,00	8,00	9690,00	60,00
1	**27.77871633**	**-25.9868753**	5,00	126,00	199,00	38,00	972,00	68,00	15,00	6220,00	150,00
1	**27.77865368**	**-25.9869778**	0,00	0,00	0,00	40,00	980,00	65,00	12,00	6300,32	140,00
1	**27.77867077**	**-25.9870803**	7,00	140,00	0,00	35,00	950,00	70,00	10,00	6400,60	145,00
1	**27.77878467**	**-25.9869437**	0,00	0,00	0,00	30,00	960,00	0,00	14,00	6500,00	155,00
1	**27.77885301**	**-25.9868013**	0,00	0,00	0,00	32,00	970,00	0,00	11,00	6600,59	150,00
1	**27.778984**	**-25.9867785**	5,50	0,00	0,00	33,00	900,00	67,00	9,00	6700,00	160,00
1	**27.77885301**	**-25.9869323**	0,00	0,00	0,00	34,00	930,00	0,00	13,00	6800,00	165,00
1	**27.77871633**	**-25.9870006**	4,50	0,00	207,00	31,00	910,00	69,00	10,00	6900,16	170,00
1	**27.77871063**	**-25.9869551**	6,50	135,00	0,00	36,00	940,00	71,00	11,50	7000,00	175,00
2	**27.82182452**	**-25.9718269**	0,00	0,00	0,00	21,00	462,00	0,00	7,00	4060,00	38,00
2	**27.82183643**	**-25.9718137**	0,00	0,00	0,00	36,00	1300,00	0,00	11,00	629,50	117,00
2	**27.82184304**	**-25.9718362**	0,00	0,00	0,00	15,00	148,00	0,00	5,00	4820,00	39,00
2	**27.82186421**	**-25.9718415**	0,00	0,00	0,00	22,00	572,00	0,00	8,00	8630,00	35,00
2	**27.82185627**	**-25.9718388**	0,00	0,00	0,00	44,00	373,00	0,00	10,00	4984,00	40,00
2	**27.8218404**	**-25.971823**	0,00	0,00	0,00	19,00	630,00	0,00	9,00	3984,00	34,00
2	**27.82185627**	**-25.9718282**	0,00	0,00	0,00	26,17	580,83	0,00	8,33	4517,92	50,50
2	**27.82182187**	**-25.9718097**	0,00	0,00	0,00	25,00	590,00	0,00	8,00	4600,00	51,00
2	**27.82183246**	**-25.9718309**	0,00	0,00	0,00	30,00	600,00	0,00	9,50	4800,00	49,00
2	**27.8218523**	**-25.9718468**	0,00	0,00	0,00	24,00	570,00	0,00	7,50	4700,00	52,00
3	**27.85444149**	**-25.9589041**	0,00	45,00	0,00	22,00	1322,00	0,00	0,00	4830,00	55,00
3	**27.8543426**	**-25.9589453**	0,00	0,00	0,00	7,00	116,00	0,00	0,00	1083,00	20,00
3	**27.85430139**	**-25.9589741**	0,00	33,00	0,00	6,00	191,00	0,00	0,00	3580,00	24,00
3	**27.85429933**	**-25.9589494**	0,00	0,00	0,00	10,00	352,00	0,00	0,00	207,46	17,00
3	**27.85434878**	**-25.9589206**	0,00	85,00	0,00	17,00	876,00	11,00	4,00	3320,00	40,00
3	**27.8543941**	**-25.9589061**	0,00	0,00	0,00	14,00	573,00	0,00	0,00	3836,47	30,00
3	**27.85428079**	**-25.9589968**	0,00	0,00	0,00	12,00	600,00	0,00	0,00	3700,00	25,00
3	**27.85428697**	**-25.958968**	0,00	0,00	0,00	15,00	500,00	0,00	0,00	3600,00	22,00
3	**27.85441265**	**-25.9589123**	0,00	0,00	0,00	13,00	550,00	0,00	0,00	4000,00	28,00
3	**27.85433023**	**-25.9589267**	0,00	0,00	0,00	16,00	650,00	0,00	0,00	3900,00	26,00
4	**27.88242167**	**-25.9490509**	85,00	0,00	62,00	13,00	1472,00	0,00	0,00	403,06	209,00
4	**27.88240533**	**-25.9490465**	0,00	56,00	0,00	33,00	831,00	20,00	7,00	551,38	302,00
4	**27.88240087**	**-25.9490346**	90,00	0,00	65,00	30,00	800,00	25,00	6,00	600,56	300,00
4	**27.88238602**	**-25.9490287**	0,00	0,00	0,00	31,00	997,00	30,00	5,00	550,00	290,00
4	**27.88239642**	**-25.949042**	0,00	0,00	0,00	32,00	820,00	0,00	8,00	620,34	295,00
4	**27.88241424**	**-25.949042**	78,00	0,00	0,00	29,00	780,00	21,00	4,00	478,44	280,00
4	**27.88241424**	**-25.9490569**	87,00	0,00	0,00	28,00	1390,00	0,00	9,00	640,00	285,00
4	**27.88240979**	**-25.9490376**	0,00	61,00	0,00	27,00	810,00	23,00	10,00	610,93	275,00
4	**27.8823979**	**-25.9490287**	0,00	0,00	0,00	26,00	1429,00	0,00	11,00	558,72	270,00

4	**27.88239196**	**-25.9490346**	0,00	0,00	0,00	34,00	830,00	18,00	12,00	615,63	268,00
5	**27.8987031**	**-25.9097893**	0,00	63,00	0,00	28,00	364,00	0,00	0,00	751,74	43,00
5	**27.89868971**	**-25.9097836**	0,00	0,00	0,00	9,00	195,00	0,00	0,00	728,16	17,00
5	**27.89870692**	**-25.9097721**	0,00	0,00	0,00	0,00	93,00	0,00	0,00	2648,47	12,00
5	**27.89871839**	**-25.9097989**	0,00	34,00	0,00	0,00	284,00	0,00	0,00	525,81	20,00
5	**27.89871265**	**-25.9097836**	0,00	0,00	0,00	12,00	286,00	0,00	0,00	836,00	22,00
5	**27.89873177**	**-25.9097683**	0,00	0,00	0,00	0,00	220,00	0,00	0,00	636,94	18,00
5	**27.89872221**	**-25.9097606**	0,00	0,00	0,00	26,17	580,83	0,00	8,33	4517,92	50,50
5	**27.89874897**	**-25.9097568**	0,00	0,00	0,00	25,00	590,00	0,00	8,00	4600,00	51,00
5	**27.89876236**	**-25.909753**	0,00	0,00	0,00	30,00	600,00	0,00	9,50	4800,55	49,00
5	**27.89874515**	**-25.9097645**	0,00	0,00	0,00	24,00	570,00	0,00	7,50	4700,00	52,00
6	**27.92900282**	**-25.8738125**	0,00	42,00	0,00	19,00	503,00	0,00	0,00	2760,00	45,00
6	**27.92900005**	**-25.8738061**	0,00	0,00	0,00	16,00	291,00	0,00	0,00	381,94	30,00
6	**27.92900558**	**-25.8738015**	0,00	0,00	0,00	8,00	102,00	0,00	0,00	1500,00	15,00
6	**27.92899637**	**-25.8738199**	0,00	340,00	180,00	70,00	1000,00	25,00	15,00	1049,86	81,00
6	**27.92899637**	**-25.8738125**	0,00	0,00	150,00	28,00	1268,00	0,00	0,00	1424,00	71,00
6	**27.9290019**	**-25.873818**	0,00	0,00	0,00	17,00	364,00	28,00	0,00	1064,35	50,00
6	**27.9290065**	**-25.8738079**	0,00	45,00	0,00	20,00	500,00	0,00	0,00	1100,00	55,00
6	**27.92900282**	**-25.8737978**	0,00	30,00	0,00	15,00	550,00	22,00	0,00	1200,01	60,00
6	**27.92901295**	**-25.8737969**	0,00	35,00	0,00	22,00	480,00	0,00	0,00	1300,25	65,00
6	**27.92900926**	**-25.8737978**	0,00	0,00	0,00	18,00	600,00	0,00	0,00	1150,00	57,00