UNIVERSITY OF CAPE COAST

PREDICTING GROUNDWATER QUALITY USING SUPERVISED MACHINE LEARNING AND GEOCHEMICAL MODELS

MICHAEL ABOAH

2023

UNIVERSITY OF CAPE COAST

PREDICTING GROUNDWATER QUALITY PARAMETERS USING SUPERVISED MACHINE LEARNING

BY

MICHAEL ABOAH

Thesis submitted to the Department of Environmental Science, School of Biological Sciences, College of Agriculture and Natural Sciences and the University of Cape Coast, in partial fulfilment of the requirements for the award of Master of Philosophy Degree in Environmental Science.

APRIL 2023

DECLARATION

Declaration of Candidate

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree at this University or elsewhere.

Candidate's Signature …………………… Date…………………………

Name: Michael Aboah

Supervisors' Declaration

I hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on the supervision of the thesis laid down by the University of Cape Coast.

Supervisor's Signature ……………………… Date …………...…..

Name: Prof. Michael Miyittah

ABSTRACT

Studies relating to groundwater have asserted that groundwater quality assessment is difficult, time-consuming and costly. An easy, vigorous, cost and time-effective tool is needed to predict water quality. The study employed supervised learning algorithms (decision tree regression and polynomial regression techniques) to build a model for assessing and predicting groundwater quality using easily measured parameters. The study employed experimental design (factorial design) and random sampling technique (multistage sampling technique) for the data collection. Model performance determinants such as R^2, RMSE and d-statistics were used to compare the performance of the model with aqueous geochemical models such as Visual Minteq, Phreeq C and Wateq4F. ANOVA was used to determine the significance mean differences in the predicted groundwater chemical parameters of the study regions. The estimated R-square, RMSE and d-statistics for Visual Minteq (0.997, 16.97 and .987), Phreeq C (0.999, 33.16 and 0.960), Wateq4F (0.972, 15.33 and 0.988) and machine learning model (0.999, 1.690 and 1.00) indicated that the model developed has high predicting ability over the aqueous geochemical models. Model validating tools like accuracy (0.96), RMSE (1.690) and R^2 (> 95%) demonstrated that the model could be used to forecast groundwater quality with high accuracy using easily measured parameters. ANOVA test demonstrated significant mean differences in the predicted groundwater chemical parameters of the study regions (P-value = 0.00 < 0.05). It is recommended that artificial intelligence tools, such as supervised learning as an easy, time and cost-effective way of predicting water quality should be encouraged in groundwater assessment.

KEYWORDS

Geochemical models

Groundwater Quality

Modelling

Prediction

Supervised Learning Algorithm

ACKNOWLEDGEMENTS

I acknowledged my supervisor, Prof. Michael Miyittah and Prof. Hugh Komla Akotoye for their encouragement towards the success of this thesis.

DEDICATION

I dedicate this thesis to all researchers in this area of study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AI	:	Artificial Intelligence
ANOVA	:	Analysis of Variance
DTR	:	Decision Tree Regression
GWI	:	Groundwater Quality Index
MDGs	:	Millennium Development Goals
NTU	:	Nephelometric Turbidity unit
PR	:	Polynomial Regression
PCA	:	Principal Component Analysis
SLA	:	Supervised Learning Algorithm
SDG	:	Sustainable Development Goal
WHO	:	World Health Organisation

CHAPTER ONE

INTRODUCTION

Excessive water pollution has compromised water quality and increased water distribution costs and associated health-related concerns (Adjei & Adjokatse, 2022). Globally, water samples are usually collected from the fields and analysed using laboratory standards. According to McCaig (2020), water quality analyses are complicated, expensive and require much time for the laboratory and in-depth analyses of the many water quality indicators.

Mohapatra et al. (2021) utilised the traditional methods in groundwater quality prediction and confirmed that the traditional methods cannot predict water quality parameters with ease due to slow processing times and errors from manual data entry. The traditional methods, though valuable, often fail to provide real-time and accurate predictions using easily measured parameters. They heavily depend on manual sampling, laboratory analysis, periodic monitoring and both chemical and physical parameters of water samples. A nonphysical method, which is an easy, vigorous, cost and time-effective tool and combines the operations of the traditional in its decisions is needed to predict water quality using easily measured parameters.

Nordstrom and Campbell (2014) and Bahlol et al. (2023) employed geochemical models to predict groundwater quality. Findings of the studies showed that geochemical models such as Visual Minteq, Phreeq C and Wateq4F could be used to predict groundwater quality with ease, but it only occurs at the mineral phase. Artificial Intelligence (AI) has emerged as a complex computer system in water quality prediction and modelling. El Bilali et al. (2021) and several other authors employed AI in water quality prediction and modelling. They concluded that AI could predict water quality parameters in an easy, inexpensive and time-effective way.

As reported by Kortei (2020), Ghana's underground water bodies (particularly, in the Central, Greater Accra and Western Regions) are receiving progressive pollution every day, but the extent to which the pollution has occurred is unknown. This study employed an AI approach, particularly decision tree regression (DTR) and polynomial regression approaches to build a model to predict groundwater chemical parameters using easily measured water quality parameters such as pH, total dissolved substances, temperature, dissolved oxygen, total hardness, turbidity and salinity. The performance of the AI model would be validated using the model-predicting indices such as accuracy, root means square errors (RMSEs), d-statistics and R-square. Moreover, the performance of the AI model would also be compared with existing aqueous geochemical models such as Visual Minteq, Phreeq C and Wateq4F to validate its ability in predicting groundwater quality.

Background to the Study

The United Nations (UN) report stated that 1.8 million people die from drinking contaminated water every year (Denise et al., 2022; Rubinstein, 2022). According to Forest et al. (2013), about 2.5 billion sicknesses and five million losses of lives occur every year due to the lack of access to clean and safe drinking water. These losses are estimated to be greater than mortalities caused by accidents, terrorist attacks and crimes. Nwankwo et al. (2020) expressed that water pollution contributes to about 80 percent of health-related complications in impoverished states like Ghana.

Adeyemi et al. (2022) and Unigwe and Egbueri (2022) asserted that several factors, including global human population growth and climate change, affect water quality. Many concerns have been raised globally about the massive destruction of water resources, predominantly in Africa and the subsequent health and environmental effects. In response to this health and ecosystem-threatening issue, governments have been mandated to protect and sustain available water sources (Denise et al., 2022), thus improving or discovering new water sources to prevent global water shortage, disease outbreaks, hunger and malnutrition.

Corno et al. (2022) suggested that the connection between human health and water quality is dynamic and complicated, therefore, a more complicated and dynamic system must be used to mimic and estimate the association between water and health. Geochemical models have been used to predict groundwater quality (Barzegar et al., 2018; Dey et al., 2023). They explained that geochemical models use water chemistry data, such as the concentrations of major ions and trace elements at the mineral phases to estimate the concentrations of groundwater.

Artificial intelligence (AI) has also been developed as a computer system to mimic human brain and solve complex problems. According to Foppen et al. (2020) and Minh et al. (2022), AI can model and predict the relationships between water and human health. Aldhyani et al. (2020); Calvert (2020); Mettu and Latifi (2020) employed artificial intelligence to model and predict water quality. The authors indicated that assessing groundwater quality of a given aquifer in the laboratory is expensive, time-consuming and laborious. Traditional methods pose challenges due to inaccuracies arising from complex laboratory analysis, manual data entry and sluggish processing speeds. For example, much time spent on data entry and processing affects the ability to draw better conclusions about the relationships among change water quality parameters. According to Van der Veer Martens et al. (2017), the failure to draw better conclusions about the variabilities in groundwater might lead to inaccurate representations of the real system and lower groundwater model performance and forecasting accuracy.

Akhtar et al. (2021) recognised water quality index (WQI) as a crucial method for estimating underground water quality and its acceptability for drinking. This method provides mechanisms for presenting cumulatively a generated quantitative expression to indicate water quality level. WQI is utilised extensively throughout the world due to its capability of fully expressing information on water quality and its acceptability (Banda & Kumarasamy, 2020). Mokarram et al. (2022) added that in predicting water quality level of any water source, WQI is the most common determinant. Many researchers have adopted various water quality indices to estimate groundwater drinking suitability and river water quality in other countries (Gautam et al., 2022; Mukherjee & Singh, 2022; Sarkar et al., 2022).

Becher et al. (2022) asserted that WQI is essential in water quality assessment. WQI provides a comprehensive and standardised way to evaluate the overall condition of water based on various parameters of quality (Şener et al., 2017). The hidden contaminants could persist in the groundwater system for decades or even hundreds of years due to the relatively slow movement of the underground water and pollutants in the subsurface. There is an urgent need to employ suitable and cost-effective tools in assessing and modelling groundwater quality using easily measured parameters.

Agbasi and Egbueri (2022) found that using traditional techniques in assessing the quality of groundwater sometimes contributes to losses in the economic parts of the assessment, impacting the ability to make decisions regarding groundwater quality management processes. Moreover, they cannot predict groundwater quality using easily measured parameters. Therefore, it has become crucial to employ a potential and affordable technique to ensure quick and accurate water quality assessment and analysis. In such a case, the supervised learning algorithm (SML) is an alternative to forecast the general groundwater quality level based on findings of analyses that do not require costly reactors or highly advanced measuring techniques but the use of easily measured parameters such as pH and electrical conductivity.

Statement of the Problem

In Ghana, studies assessed the physicochemical parameters of the surface water bodies (rivers, streams and lakes) and found that some of the surface water bodies are polluted (Yeleliere et al., 2018; Kortei et al., 2020; Ampim et al., 2021; Amuah et al., 2022). Abanyie et al. (2020) argued that water pollution comprises surface and underground water pollution and if surface water bodies have been assessed, then there is a need to assess the state of the underground water bodies in Ghana. Kortei (2020) and Amuah et al. (2022) conducted a study in Southern Ghana and asserted that underground water bodies are receiving progressive pollution every day, but the extent to which the pollution has occurred is unknown. Studies are, therefore, needed to estimate the extent of groundwater pollution in Southern Ghana.

Studies such as McCaig et al. (2020) and Agrawal et al. (2021) showed that water quality assessment is laborious, complicated and expensive and much time is needed for the complex laboratory and in-depth analyses of the many water quality parameters. To find a better way to make water quality experiments less expensive and laborious, Mohapatra et al. (2021) and Singha et al. (2022) employed traditional methods such as ANOVA in the prediction and modelling of water quality. They asserted that the traditional methods are associated with difficulty due to errors from manual data entry and slow processing times. This hinders prediction efficiency and scalability and the drawing of better conclusions about the relationships among change water quality parameters (Mohapatra et al., 2021). There is a need for an easier and less laborious method for water quality assessment.

Prommer et al. (2019) assessed the effectiveness of geochemical models such as Phreeq C and Visual Minteq in groundwater quality prediction. Singh et al. (2013) employed geochemical models such as Wateq4F to predict groundwater quality using the mineral phase ions. Singh et al. (2013) and Prommer et al. (2019) noted that geochemical models although effective tools in predicting groundwater quality could only predict water quality at the mineral phases. Geochemical models cannot analyse datasets and detect patterns that are not easily identifiable (Zuo et al., 2021). In addition, they cannot save time, money and resources and extract concrete information from raw data to solve complex problems. An easy, efficient, risk-free, cost-effective and vigorous tool is needed to model and predict groundwater quality beyond the mineral phases and extract concrete information from raw data to solve complex groundwater problems. A study by El Bilali et al. (2021), Foppen et al. (2020) and Minh et al. (2022) found that using AI (specifically, supervised machine learning using regression approaches) in water quality prediction and modelling can help predict water quality parameters in an easy, inexpensive and time-effective way.

From the literature review, there has not been any work done in predicting and modelling groundwater chemical parameters and classifying groundwater pollution levels using DTR and polynomial regression approaches and geochemical models such as Visual Minteq, Phreeq C and Wateq4F. Moreover, no study has been done on the use of supervised learning (with DTR and polynomial regression approaches) in modelling chemical parameters of groundwater quality using easily measured parameters. This current study employs the supervised learning algorithm (particularly the decision tree regression and polynomial regression algorithms) in AI to build a model for predicting and modelling the chemical parameters of the groundwater in the study regions (Central, Greater Accra and Western) using easily measured parameters such as pH, TDS, temperature, salinity, dissolved oxygen, total hardness and total dissolved substances. The predictions of the AI model would be compared with the geochemical models to establish the accuracy of the model in predicting groundwater quality.

General Objective of the Study

The main objective of the study is to develop a model to predict groundwater chemical parameters using easily measured parameters. The study also seeks to determine factors that influence groundwater pollution in the study regions (Western, Central and Greater Accra Regions) and estimate groundwater quality indices for the study regions using supervised learning regression models.

Specific Objectives

1. To develop a model and determine its performance in predicting groundwater chemical parameters based on easily measured parameters.
2. To compare the performance of the model with aqueous geochemical models (such as Visual Minteq, Phreeq C and Wateq4F) in predicting groundwater quality parameters.
3. To establish if there are statistically significant mean differences between and among the predicted groundwater chemical parameters of the study regions (Western, Central and Greater Accra Regions).

Hypotheses

H_{A1}: The model could predict chemical parameters of groundwater using easily measured parameters with high accuracy.

H_{A2}: The model could predict groundwater quality chemical parameters with high accuracy compared to Visual Minteq, Phreeq C and Wateq4F.

H_{A3} There would be statistically significant mean differences between and among predicted groundwater chemical parameters of the study regions.

Significance of the Study

The findings of the study would help determine environmental parameters that influence the groundwater quality of the study regions. The use of supervised learning in predicting groundwater quality at a larger scale serves as an effective tool for improving, managing and developing vital groundwater resources. The study would help understand the present groundwater quality and pollution levels which are critical for predicting future safe drinking water sources. The outcomes of the study would instigate further studies using data based on regions in Ghana and data availability to strengthen the obtained conclusions. It would help increase attention on water quality necessary to realise the SDGs, regarding water security (SDG 6), health (SDG 3) and food production (SDG 2).

The findings of this study would help contribute to the achievement of the African Union Agenda 2063. By accurately predicting groundwater quality, the study would help promote sustainable water resource management, ensuring access to clean and safe water for all Africans. The findings of the study would contribute to the development of effective policies and strategies for water resource management, which are crucial for achieving the vision of a prosperous and peaceful Africa with sustainable development outlined in Agenda 2063. The findings of the study would also provide decision-makers with information on how to allocate resources effectively for groundwater quality monitoring and management, particularly in areas where water quality is likely to be poor.

The findings that would be obtained from geochemical models would help understand the interactions between groundwater and geologic materials and predict the composition and concentration of contaminants in the groundwater. Also, using SLA as a new technology would help generalise the data from the regions to areas with less information, promoting an understanding of groundwater's possible menaces and susceptibilities. The evaluation of groundwater change would serve as a foundation for regional policymakers to determine the amount of water stress and provide efficient withdrawals and supplies of groundwater to alleviate water scarcity. Artificial intelligence (AI) tools such as supervised learning help reduce the time during data collection processes.

Delimitation

The study was limited to three regions in Ghana (Central, Greater Accra and Western Regions), where there are more industrial, agricultural and domestic discharges into the environment. Replicates of three of the groundwater (from each well) were taken for the data analysis. The groundwater parameters were analysed on-site to control temperature variation and the influence of weather differences among the three selected regions in Ghana. The lack of significant and overall weather similarity across the sampling period ensured uniform hydrologic conditions across the sampling period, allowing for more accurate water-quality data comparisons.

All meter recordings were first calibrated at each sampling excursion to facilitate accurate measurements. The data that would be collected would be in a standardised format to ensure consistency and accuracy. It would also be organised in a systematic way that is easy to search and retrieve. Data would be stored securely in google drive to protect it from unauthorised access, loss, or theft. Passwords would be placed on databases or encrypted sensitive data.

Limitation

Natural and anthropogenic processes influence groundwater quality; the study could not classify those factors into either anthropogenic or natural processes. Furthermore, weather pattern differences and temperature among the three regions in Ghana might influence the study's findings. The study was limited to groundwater quality parameters that are easily measured, which might not allow full representation of the processes associated with groundwater quality changes in Ghana.

Temperature variation among the study regions might affect the results of the study. The internal resistance or error by each of the instruments used might likewise affect the findings of the study. Although the use of artificial intelligence tools such as supervised learning helps provide a better understanding of the hydrological system of underground water, a large data size is needed to represent and model the underground water processes.

Scope of the Study

This study set out to develop a model which could predict groundwater quality based on easily measured parameters. It also assessed, estimated and compared the groundwater quality of the study. It also focused on three regions in Ghana: groundwater quality indexes of Central, Greater Accra and Western Regions and activities that contribute to groundwater pollution in the regions. Finally, the study focused on modelling, protecting, managing, and safeguarding groundwater quality in the selected regions.

Definition of Terms

Artificial intelligence (AI) is a computer system that needs human assistance to perform tasks normally. It imitates and behaves like humans in its operations.

Decision tree regression (DTR) is a type of regression analysis that uses a decision tree as a predictive model to map features (input) to a target variable (output). In decision tree regression, the algorithm breaks down a dataset into smaller and smaller subsets while at the same time, an associated decision tree is incrementally developed.

Geochemical models are models for predicting chemical species in solution or solid phases in equilibrium with a solution. It helps explain the chemical compositions of materials and understand the distribution and behaviour of different elements at different depths and temperatures.

Groundwater Quality is the state of the groundwater, thus, whether or not it is good for use. It consists of information on underground water's physicochemical, biological and radiological properties.

Supervised Learning (SL) is a subset of machine learning techniques that utilise labelled datasets to train the model and forecast the output. The system is trained using a learning process to produce the target values from the input.

Organisation of the Study

This study consists of seven chapters. Chapter One comprises the introduction, background of the study, the statement of the problem, research objectives, significance of the study, delimitation, limitation, scope of the study, definition of terms and organisation of the study. Chapter Two encompasses the review of relevant literature on the subject matter under consideration. The research design and methods of data collection are spelt out in Chapter Three. Chapter Four contains the research findings, discussions and interpretation of the results based on field data analysis. Chapter Five presents the research findings, discussions and interpretation of the results based on machine learning prediction analysis. Chapter Six summarises the major findings of the study, its conclusion and recommendations, policy implication, practical impractical implication and suggestions for further studies.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This section provides a review of relevant literature surrounding the objectives of the study and the research topic. It covers the theoretical review, theoretical framework and conceptual frameworks based on groundwater sustainability and supervised learning algorithms. It also reviews concepts such as groundwater quality and pollution, AI or supervised learning. The chapter also covers areas like SML techniques and model evaluation metrics as well as the comparison between ML and geochemical models, conceptual framework and empirical review.

Theoretical Review

This part of the study reviews the Theory of Predictive Modelling. It discusses the strengths and weaknesses and why the theory has been selected for this study.

Predictive modelling theory (by Francis Galton and Norbert Wiener 1990s)

This current study employed the Predictive Modelling Theory (PMT) by Francis Galton and Norbert Wiener developed in the 1990s with Galton's regression analysis (Kerby, 2015). The theory focuses on utilising historical data to train models that can make accurate predictions about future outcomes (Machta et al., 2013). The theory states that when a model is trained and tested, it stands the chance to predict future events or outcomes with a high accuracy by analysing patterns in a given set of input data. In addition, when selected variables are carefully executed, it enhances the accuracy and interpretability of the model (Machta et al., 2013).

According to the theory, a vital component of every model is the training. Training a model ensures the model captures patterns within the available data (Kerby et al., 2015). Evaluation metrics, such as accuracy and RMSE, are essential strengths in quantifying model performance (Verhagen et al., 2018). The theory holds that relying solely on a single metric may pose a weakness, as it might not capture all dimensions of the model's behaviour. The theory emphasises the importance of understanding the relationship between predictors and predictands, providing a strength in guiding model design (Frees et al., 2014). Model validation serves as a strength by assessing the model's performance on independent datasets (Verhagen, 2018). Interpretability is a strength when models are designed to be interpretable, facilitating understanding and trust (Frees et al., 2014).

A potential weakness of the theory lies in incomplete variable selection, which may lead to issues such as overfitting or underfitting. Overfitting hinders the model's generalisation of new and unseen data. The theory has been selected for the study because its emphasis on variable selection, model training, validation and evaluation metrics aligns well with the objectives of performance evaluation, predictability assessment, statistical significance and model comparison (Kerby et al., 2015). The theory emphasises the development of a model that can predict future outcomes, aligning with the goal of assessing the predictability of groundwater chemical parameters (Kerby et al., 2015). The PMT plays a crucial role in comparing the developed model with traditional aqueous geochemical models

(Varoquaux & Poldrack, 2019). By applying the theory, the study ensures that the developed model's predictive capabilities are assessed and compared with established geochemical models.

Conceptual Review

Groundwater quality

Groundwater Quality (GI) is the condition of groundwater's chemical, biological and physical properties regarding its domestic, industrial and agricultural uses (Najafzadeh et al., 2022). GI is determined by bacteria levels, dissolved oxygen concentration, salinity or turbidity. GI depends on rock and soil properties, groundwater velocity, catchment area, recharge water quality, urbanisation and exploitation of water resources, anthropogenic activities, and atmospheric inputs (Wang & Li, 2022). Groundwater quality is inherently associated with human health, poverty reduction, livelihoods, food security, ecosystem preservation and our societies' economic growth (Zhou et al., 2022). Wang et al. (2022) opined that poor underground water quality poses health risks and impacts ecosystems. Groundwater quality determines the water's usefulness for drinking, irrigation of food, hygiene, fodder and feed crops, aquaculture, food animals, manufacturing and industrial production (Guan et al., 2022).

El-Aziz et al. (2017) and Teixeira et al. (2021) grouped groundwater quality based on its chemical composition, bacterial (coliform) content, salt content, colour, taste, odour and turbidity of the water. The main chemicals that define water quality comprise alkaline elements such as sodium (Na) and potassium (K). Also, alkaline earth elements like magnesium (Mg), calcium (Ca), sulphate, Chloride and the anionic complexes bicarbonate and heavy metals (e.g., Cd, As, Fe, Pb, Hg), and organic compounds (Shamsuddin et al., 2019; Gaur et al., 2022). Teixeira et al. (2021) added that these physicochemical and biological parameters could be used to determine drinking water safety based on some standards.

Groundwater pollution

Groundwater pollution refers to the contamination of underground water sources, known as aquifers, by harmful substances introduced into the ground (Sahoo et al., 2022). Groundwater contaminants are substances that, when dissolved in groundwater, will travel with the water and end up in drinking water wells (Li et al., 2022). Contaminants include heavy metals, pesticides, fertilizers, petroleum hydrocarbons, industrial chemicals, and pathogens (Balasubramanian et al., 2022).

Groundwater pollution originates from various sources, including industrial activities, agricultural practices, urban runoff, improper waste disposal, and natural processes (Akhtar et al., 2021). These can be grouped into natural and anthropogenic sources (Krok et al., 2022; Dinesh et al., 2022). Natural sources include mineral dissolution, geological processes, and leaching from soil and rock formations. Anthropogenic sources involve activities such as industrial discharge, agricultural runoff, improper waste disposal, and leakage from underground storage tanks (Khan et al., 2014).

Artificial intelligence

Artificial intelligence (AI) is the construction of machines, programmes and robots which show intelligent behaviour as humans (Ibrahim et al., 2022). AI is the machines' ability to analyse, learn and comprehend information independently, just as humans do. AI is in the form of a robot, machine, software or computer programme. AI includes simulating human intelligence procedures by machines such as computer systems (Aggarwal et al., 2022). AI is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (Alldritt et al., 2022). It is applied to projects of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, generalise, discover meaning or learn from experience. For example, AI can be found in computer search engines, natural language processing, medical diagnosis and voice or handwriting recognition in this limited meaning (Islas-Cota et al., 2022; Oluyisola et al., 2022).

Speech translation, machine sensing, problem-solving, robots, and gaming are the five main applications of AI (De Keyser et al., 2022). Someone with expertise in robotics, natural language processing, and knowledge acquisition supports these implementation areas as well. Artificial intelligence (AI) needs a foundation of specialised hardware and software for building and training machine learning algorithms. While no computer language is exclusively associated with AI, a few stand out, including R, Python and Java (Rana et al., 2021).

AI systems typically operate by ingesting massive amounts of labelled training data, analysing the data for correlations and patterns, and then projecting future states using these patterns. For example, an image recognition programme can learn to recognise and describe items in photographs (Bekhit, 2022). Moreover, learning, self-correction and reasoning are the three cognition processes that artificial intelligence programming concentrates on at every command (Namasivayam et al., 2022). Learning processes consist of data gathering and rules development for converting that data into valuable information. Algorithms are rules for providing computing equipment to execute a task (Bekhit et al., 2022). Self-correction techniques are used in AI programming to optimise programmes and guarantee the most precise and workable outcomes (Austin et al., 2022).

AI has a wide application within computer science, medical science, biological science, and other fields of business (Boyd & Wilson, 2017). AI is essential because it provides businesses access to existing hidden operational information and has the potential to perform better than humans under certain conditions (van Herck, 2022). For instance, it assists with meticulous activities like analysing voluminous legal documents to ensure that crucial fields are appropriately filled. AI systems can often complete tasks quickly and with few errors (Kelly et al., 2022).

Due to AI's ability to analyse massive amounts of data faster and produce predictions that are more accurate than humans sometimes, AI techniques like Artificial Neural Networks and deep learning are rapidly evolving (Boyd et al., 2017). A human researcher would be drowned under the daily flood of enormous amounts of data; AI solutions use machine learning to quickly transform data into useful information. It, therefore, excels at activities requiring attention to detail, saves time on data-intensive tasks, generates trustworthy results, and is continually accessible thanks to AI-powered virtual agents (Namasivayam et al., 2022). The primary drawback of AI is the expensive expense of processing the enormous amounts of data that are required for AI programming (Kelly et al., 2022). Although AI only understands what has been demonstrated to generalise information from one activity to another, it demands technical instruction and few skilled employees (Scharre, 2019), reducing the labour force or human resources in the working fields.

Types of artificial intelligence

There are three main types of artificial intelligence, Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI) (Kuusi & Heinonen, 2022). Artificial Narrow Intelligence (ANI) or weak AI is a type of AI designed for a specific task or a narrow range of tasks (Zawacki-Richter et al., 2019). ANI systems are designed to excel in specific tasks or functions, such as image recognition, language translation or game playing (Jiang et al., 2022). They are engineered with algorithms optimised for performing tasks, enabling efficient and accurate performance within their designated domain.

ANI relies on task-specific algorithms trained on large datasets to perform its functions (Holzinger et al., 2019). For example, deep learning algorithms, including convolutional neural networks (CNNs) for image recognition and recurrent neural networks (RNNs) for natural language processing, are commonly used in ANI systems (Sultan et al., 2020). ANI cannot generalise beyond its training data or adapt to new situations autonomously. ANI systems are rigid and may struggle to cope with novel scenarios or tasks outside their predefined scope. ANI is used for medical image analysis, disease diagnosis, and personalized treatment recommendations (Arabahmadi et al., 2022). ANI powers algorithmic trading systems for automated stock trading and fraud detection algorithms for identifying suspicious transactions (Harris, 2022).

Artificial General Intelligence (AGI) refers to AI systems that can understand, learn and apply knowledge across a wide range of tasks, similar to human intelligence (Goertzel, 2014). AGI aim to replicate the broad spectrum of cognitive abilities exhibited by humans, including perception, reasoning, learning, problem-solving and creativity (Korteling et al., 2021). Unlike narrow AI systems, which are designed for specific tasks, AGI aims to possess the flexibility and adaptability to perform a diverse range of tasks with human-like proficiency (Aithal, 2023). AGI encompasses humanoid robots, self-driving cars, virtual assistants, medical diagnosis systems, educational tutoring platforms, creative AI artworks and scientific research tools. These types of AI aim to emulate human-like cognitive abilities and adaptability across diverse tasks and domains (Korteling et al., 2021).

AGI systems are equipped with learning mechanisms that enable them to acquire knowledge from diverse sources, including data, experience, and interaction with the environment (Aithal et al., 2023). These systems can continuously learn and improve their performance over time, exhibiting adaptive behaviour akin to human learning processes (Zawacki-Richter et al., 2019). AGI emphasises the ability to generalise knowledge across different domains and transfer learning from one task to another (Zhu et al., 2023). AGI systems can apply learned knowledge to novel situations and tasks, demonstrating versatility and robustness. AGIs are used in healthcare, education, finance and robotics to solve complex problems (Sharma et al., 2019).

Artificial Super Intelligence (ASI) is a type of AI that possesses cognitive abilities far superior to those of humans and could potentially outperform humans in all intellectual tasks (Azamat, 2021). ASI transcends human capabilities, exhibiting levels of comprehension, creativity, and problem-solving prowess beyond human comprehension (Jiang et al., 2022). ASI includes HAL 9000 from "2001: A Space Odyssey," Skynet from the "Terminator" franchise, and Ultron from Marvel Comics (Newby, 2001). ASI is envisioned to possess cognitive abilities that far exceed the combined intellect of all human minds (Aithal et al., 2023). With access to vast repositories of knowledge, data, and computational resources, ASI could analyze complex phenomena, devise novel solutions, and predict future outcomes with unparalleled accuracy (Jiang et al., 2022).

Applications of artificial intelligence

Artificial intelligence (AI) is used in various technologies and helps solve complex problems (Lu, 2019). According to Goncalves et al. (2022), automation tools help companies accomplish their work within a limited time. In addition, it can be used to process language, thus, a computer programmes interprets human language. For example, robots perform complex or inconsistent activities for people to complete. They are being employed in car manufacturing. In healthcare, AI applications are utilised to make complex tools for detecting and identifying diseases as well as cancer cells (Chan-Olmsted, 2019). Artificial intelligence (AI) aids in the analysis of medical problems using laboratory and other health information to ensure early detection (Goncalves et al., 2022).

AI is being utilised to discover new pharmaceuticals by merging historical data with medical understanding. It can also exploit weaknesses and nutritional deficiencies in soils (Assunço et al., 2022). AI can analyse where weeds are growing to utilise computer vision, robotics, and machine learning techniques (Hanoon et al., 2021). AI robots can assist in harvesting crops at a bigger volume and faster rate than human labour (Verma, 2022).

Machine learning

Machine learning (ML) is a subfield of AI in which computers utilise algorithms to sift through large amounts of data in search of hidden patterns that may be used to make predictions (Hashimoto et al., 2018). ML helps use statistical methods and algorithms to train data and classify or predict variables (Schölkopf et al., 2022). Deep learning (DL) is a subset of machine learning considered as automation of predictive analytics.

DL repeats a task and modifies it slightly each time to better the outcome. According to Tulbure et al. (2022), it is called ‘deep learning’ since the neural networks have multiple (deep) layers to permit learning. ML involves getting a computer to do something without programming it (Schölkopf, 2022). It is where machines learn and acquire skills without human involvement (Gundersen et al., 2022).

MLAs are computational techniques that enable computers to learn patterns and make predictions from data without being explicitly programmed (Galloway, 2022). These algorithms utilise statistical methods to analyse large datasets, identify patterns, and make decisions or predictions based on the patterns they’ve learned (Thompson et al., 2022). MLAs are divided into unsupervised, reinforcement learning and supervised (Schölkopf et al., 2022). Unsupervised machine learning is an algorithm that is trained on unlabelled data and is sorted by similarities and differences (Larsen-Greiner et al., 2022). Reinforcement machine learning is a type of learning that uses positive reinforcement (Alavizadeh et al., 2020). Again, the data sets are not labelled, but the AI system receives feedback after completing an action or a series of activities (Larsen-Greiner et al., 2022).

MLA is trained on labelled data, meaning each input data point is associated with a corresponding target output (Wang et al., 2022). Common MLAs include linear regression, decision trees, support vector machines, and neural networks (Hanoon et al., 2021). MLAs work by iteratively adjusting their internal parameters based on the error or difference between the predicted outputs and the actual labels in the training data (Ciccozzi et al., 2022). This helps the model to minimise error through optimization techniques like gradient descent or backpropagation, ultimately learning the underlying patterns in the data (Thompson et al., 2022).

Supervised machine learning techniques

Linear regression is a supervised learning algorithm used for modelling the relationship between a dependent variable and one or more independent variables (Maulud & Abdulazeez, 2020). It assumes a linear relationship between the independent variables (predictors) and the dependent variable (response) (Panigrahi et al., 2023). The primary goal of linear regression is to fit a straight line or hyperplane to the data points in such a way that it minimises the difference between the observed and predicted values (Apley & Zhu, 2020). Linear regression can be expressed mathematically as:

$$Y = a + bX + \in \quad [2.1]$$

Where:

Y is the dependent variable

X is the independent variable

b is the coefficient, representing the slope of the linear relationship between each independent variable and the dependent variable.

ϵ represents the error term, which accounts for the variability in the data that is not explained by the linear model.

Polynomial regression is a type of regression analysis that models the relationship between the independent variable and the dependent variable as an nth-degree polynomial function (Adesanya et al., 2018). Unlike linear regression, which assumes a linear relationship between variables, polynomial regression can capture nonlinear relationships between variables by fitting a curve to the data (Archontoulis & Miguez, 2015). The general form of a polynomial regression model is:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_n X^n + \varepsilon \qquad [2.2]$$

where:

Y is the dependent variable (response),

X is the independent variable (predictor),

B_0, β_1, β_2 …, β_n are the coefficients of the polynomial terms,

and ε is the error term representing the difference between the observed and predicted values.

Polynomial regression allows for more flexibility in modelling complex relationships between variables compared to linear regression (Adesanya et al., 2018). By increasing the degree of the polynomial (i.e., the value of *n*), the model can fit more complex curves to the data (Dalal & Zickar, 2012). However, higher-degree polynomials can lead to overfitting, where the model captures noise in the data rather than the underlying pattern, resulting in poor generalization to new data (Bilbao et al., 2017). To fit a polynomial regression model, the degree of the polynomial must be specified. This can be determined based on domain knowledge, experimentation, or through techniques such as cross-validation (Archontoulis et al., 2015). Polynomial regression is commonly used in various fields, including engineering, physics, economics, and biology, where relationships between variables are nonlinear (Brunton et al., 2016). According to Archontoulis et al. (2015), it can be used for exploratory data analysis and visualising trends in data when a simple linear model is insufficient to capture the underlying pattern.

Random forest is an ensemble learning technique that combines multiple decision trees to make predictions (Joshi & Srivastava, 2014). It operates by creating numerous decision trees during training and outputs the mode (for classification) or mean prediction (for regression) of the individual trees (Khatibi et al., 2014). This method derives its strength from the combination of multiple models, with each decision tree constructed independently and subsequently aggregated to yield a final prediction (Joshi et al., 2014). Each decision tree in a random forest is trained on a distinct subset of the training data and operates by making predictions through a series of binary splits within the feature space (Sani et al., 2020). Random forest offers insights into feature importance, allowing practitioners to discern the contributions of different features to the overall predictive power of the model (Khatibi et al., 2014).

Support Vector Machines (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. In regression, SVM seeks to find the hyperplane that best separates the data points while maximizing the margin between different classes. SVM can effectively handle high-dimensional data and is capable of capturing non-linear relationships through the use of kernel functions (Amarappa et al., 2014).

Support Vector Machines (SVM) is a robust and versatile supervised learning algorithm used extensively for classification and regression tasks (Amarappa et al., 2014). One of its notable features is its capability to perform linear and non-linear classification by mapping input data into a higher-dimensional feature space using kernel functions (Singla & Shukla, 2020). SVM can identify an optimal hyperplane that effectively separates different classes of data points, aiming to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class (Amarappa et al., 2014). An essential concept in SVM is that of support vectors, which are the data points lying closest to the decision boundary or hyperplane (Singla et al., 2020).

SVM includes a regularization parameter (C) that allows for balancing between maximizing the margin and minimizing classification errors (Boateng et al., 2020). SVM's versatility extends beyond binary classification tasks, as it can be adapted to handle multi-class classification through strategies like one-vs-one or one-vs-rest (Boateng et al., 2020). SVM's sensitivity to outliers is worth noting, as outliers can significantly impact the position of the decision boundary and the margin (Kasula, 2019). SVM's ability to handle high-dimensional data and capture non-linear relationships makes it particularly well-suited for complex classification tasks (Rodríguez-Pérez & Bajorath, 2022). Proper handling of outliers, either through parameter tuning or outlier detection techniques, is essential to ensure the robustness of the SVM model (Osisanwo et al., 2017). SVM finds utility across diverse domains such as image classification, text categorization, bioinformatics, finance, and healthcare (Singla et al., 2020).

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of the human brain's neural networks (Eluyode & Akomolafe, 2013). ANN consists of interconnected nodes organised into layers, including input, hidden, and output layers (Okwu et al., 2021). ANNs consist of interconnected nodes, or neurons, organised into layers: an input layer, one or more hidden layers, and an output layer (Eluyode et al., 2013). ANNs are structured as a network of interconnected nodes, or neurons, arranged in layers (Ahamed & Akthar, 2016). Neurons are the fundamental units of ANNs and each neuron receives input signals, processes them through a weighted sum, applies an activation function, and produces an output signal (Ahamed et al., 2016). Neurons are organised into layers: input layer, hidden layers, and output layer (Okwu et al., 2021).

Weights represent the strength of connections between neurons. Each connection between neurons is assigned a weight that determines the importance of the input signal. Through a process called training, ANNs learn to adjust the weights of connections between neurons to minimise the difference between predicted and actual outputs (Ibrahim et al., 2021). This process involves feeding the network with labelled training data and using optimization algorithms such as gradient descent to update the weights iteratively (Kufel et al., 2023). In ANN, activation functions introduce non-linearity into the network, enabling ANNs to learn complex patterns and relationships in data (Singh et al., 2023). Common activation functions include sigmoid, tanh, ReLU (Rectified Linear Unit), and softmax (Ibrahim et al., 2021). ANNs consist of multiple layers of neurons (Kufel et al., 2023). The input layer receives input data, the hidden layers process the data through weighted connections and activation functions, and the output layer produces the final predictions or outputs (Mayowa & Olajide, 2024).

Bias units in ANN are additional neurons in each layer that provide flexibility and help the network learn more complex patterns (Mayowa et al., 2024). Bias units have a constant input value and their associated weights. Connections represent the pathways through which signals are transmitted between neurons (Hasnaoui et al., 2024). Each connection has an associated weight that determines the strength of the signal transmitted (Okwu et al., 2021). In ANN, the loss function measures the difference between predicted and actual outputs, quantifying the model's performance (Eluyode et al., 2013). During training, the goal is to minimise the loss function by adjusting the weights of connections (Okwu et al., 2021). Optimizers are algorithms used to update the weights of connections during training (Ahamed et al., 2016). Common optimizers include gradient descent, stochastic gradient descent, and Adam (Eluyode et al., 2013).

ANNs are capable of learning complex patterns and relationships in data, making them powerful tools for tasks such as classification, regression, pattern recognition, and decision-making (Ibrahim et al., 2021). They are widely used in various fields, including finance, healthcare, image and speech recognition, natural language processing, and autonomous vehicles, among others (Amiri et al., 2023). ANNs offer flexibility, scalability, and adaptability, allowing them to handle large datasets and perform parallel processing tasks efficiently (Serey et al, 2023). Despite their effectiveness, ANNs require careful design, tuning, and training to achieve optimal performance, and their black-box nature may pose challenges for interpretability and understanding of underlying decision-making processes (Amiri et al., 2023).

Evaluation metrics

Evaluation metrics serve as critical tools in assessing the performance of predictive models (Corona et al., 2019). Effective evaluation metrics provide meaningful insights into the model's behaviour, strengths, and weaknesses, allowing stakeholders to make informed decisions (Coronado et al., 2022). Evaluation metrics include accuracy, coefficient of determination, RMSE and D-statistics. Accuracy in the context of machine learning refers to the proportion of correctly classified instances out of the total instances evaluated by a model (Raschka, 2018). Accuracy assumes that each class is represented fairly and equally, meaning that the model's predictions are evaluated without bias towards any particular class (Kotsiantis et al., 2007). It is mostly used to assess the overall performance of a classification model (Raschka et al., 2018). Mathematically, accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions made by the model (Ebrahim et al., 2023). For example, if a model correctly classifies 90 out of 100 instances, its accuracy would be 90%.

R-squared or coefficient of determination is a statistical measure used to assess the goodness of fit of a regression model to the observed data (Onyutha, 2020). R-squared quantifies the proportion of the variance in the dependent variable (groundwater quality) that is predictable from the independent variables (features) in the model (Onyutha et al., 2020). R-squared values range from 0 to 1, where 0 indicates that the model does not explain any of the variability in the dependent variable, and 1 indicates that the model perfectly explains all of the variability (Onyutha et al., 2020). R-squared measures how well the regression model fits the observed data points, with higher values indicating a better fit (Onyutha et al., 2020). R-squared alone does not provide information about the appropriateness of the model or the significance of individual predictors (Rights & Sterba, 2023). Adjusted R-squared is commonly used to account for the number of predictors in the model, providing a more accurate representation of the model's goodness of fit when comparing models with different numbers of predictors (Chicco et al., 2021; Rights et al., 2023).

RMSE measures the average deviation between the predicted values and the actual values (Hodson, 2022). RMSE is used for evaluating the performance of regression models (Karunasingha et al., 2022). It measures the average deviation between the predicted values and the actual values in a dataset. RMSE is calculated by taking the square root of the average of the squared differences between the predicted and actual values. Like R-squared, lower RMSE values indicate better model performance (Hodson et al., 2022). RMSE is sensitive to outliers in the data, as large errors can disproportionately influence the overall metric (Karunasingha et al., 2022).

D-statistics or Durbin-Watson statistics measures the presence of autocorrelation in the residuals of a regression model (Guerard et al., 2022). D-statistics range from 0 to 4, with values close to 2 indicating no autocorrelation, values below 2 indicating positive autocorrelation, and values above 2 indicating negative autocorrelation (Sah & Pandey, 2023). D-statistics are particularly relevant in time series analysis and regression models where observations may be correlated over time (Daniyal et al., 2023).

Geochemical models (Visual Minteq, Phreeq C and Wateq4F)

Geochemical models or theoretical geochemistry is the practice of using chemical thermodynamics, chemical kinetics or both, to analyse the chemical reactions that affect geologic systems, commonly with the aid of a computer (Dippong et al., 2019). Numerous hydrogeochemical models and investigations of ionic ratios in underground water. Dippong et al. (2019) analysed the hydrogeological features of two hydrological systems in Turkey using a forward hydrogeochemical model. Results demonstrated that natural processes (water-rock interactions) rather than human activities governed the mechanism affecting water and soil in the aquifer.

Naderi Peikam and Jalali (2016) investigated saturation indices (SI) and mass transfers in the water-rock system using the geochemical programme PhreeqC. As input data, temperature, pH, and the principal ions described in the software's solution keywords are used. The minerals chosen were biotite, quartz, muscovite, feldspar, andalusite, mica, calcite, sylvite, dolomite, kaolinite, Ca-montmorillonite, chlorite, sillimanite, and carbon dioxide (CO2). The findings of the investigation demonstrated that PhreeqC can forecast water quality ions.

Wang et al. (2019) investigated the multi-water quality parameters influencing iron release in polyethene pipes. They used model water (Visual MINTEQ) with varying concentrations of chloride, sulphate, bicarbonate and pH levels to create polyethene pipes using the response surface methodology. They did, however, assess the change in iron concentration. Regression models were also employed to describe the link between the five water quality measures and iron release. The coefficients of determination for the fitting equations of total and soluble iron concentrations in water were 0.890 and 0.870, respectively. The concentration of iron in water increased faster in the presence of humic acid (HA) than in the absence of the other four variables (chloride, sulphate, bicarbonate, and pH). Visual MINTEQ results indicate that a decreased HA concentration increased the degree of saturation of iron particles. This leads to a significant increase in the iron concentration in water.

Geochemical models and machine learning

Geochemical models are models for predicting concentrations of chemical species in equilibrium and saturation indices (SI's) of solid phases in equilibrium with a solution. Geochemical models such as Visual Minteq and Wateq4F are used to model chemical ions in water (Orina, 2015), while Phreeq C is used to simulate chemical reactions and transport processes in natural or contaminated water (Parshotam, 2014). Khalid et al. (2023) and Davand et al. (2022) also added that these aqueous models have a high ability to predict concentrations of chemical substances in equilibrium.

Geochemical models are utilised in a range of sectors, including ecological sustainability and clean-up, the oil industry, and economic geology (Orina et al., 2015). The models aid in understanding the nature of natural waters; the movement and breakdown of pollutants in flowing groundwater or surface water; the ion speciation of plant nutrients in the soil and of regulated metals in stored solid wastes; the formation and dissolution of rocks and minerals in geologic formations in response to injection of industrial wastes, steam, or carbon dioxide; and the dissolution of carbon dioxide in seawater and its effect on ocean acidification (Davand et al., 2022). According to Dippong et al. (2019), the greatest disadvantage of geochemical models is that they cannot analyse datasets and detect patterns that are not easily identifiable otherwise. Automation is not possible; thus, they cannot save time, money, and resources, and cannot extract concrete information from raw data to solve complex problems. Most water quality measures may be monitored using the methodologies outlined in the applicable standards. Results of water quality evaluation vary greatly depending on the parameters used (Wang et al., 2022). Because of the high cost, technological complexity, and inability to account for water quality variations, including all water quality metrics is impossible (Dippong et al., 2019). More and more scientists are optimistic that massive volumes of data can be collected and evaluated to satisfy the complicated and extensive needs for gauging water quality thanks to recent developments in machine learning techniques.

Machine learning (ML) has already been rapidly adopted as a new approach for processing and analysing data due to its high accuracy, adaptability, and ease of use in a variety of contexts (Hashimoto et al., 2018). ML is well-suited to tackling complex nonlinear relational data, which in turn makes it easier to uncover the underlying processes at play (Messaoud et al., 2020). ML can collect useful raw data knowledge and offers detailed tests to understand dynamic and data-rich issues. MLAs can learn and process data from the input. ML can extract concrete lessons from raw data to solve complex and data-rich business problems quickly. Unlike geochemical models, machine learning can be used to predict groundwater quality to save time, money, and resources, and extract concrete information from raw data to solve complex problems.

The areas of environmental science and technology have found a great use for machine learning in past years due to their flexibility. This means that despite the difficulty, employing ML for groundwater monitoring and assessment may lead to more reliable outcomes (Wang et al, 2019). ML is a strong data analysis method that is often used to find patterns or predict outcomes from large amounts of data. Data collection, method selection, model construction and validation are all necessary steps before putting machine learning into practise. The algorithm one decides to use is an important part of each of these steps. Both supervised and unsupervised learning are important types of machine learning technology (Wang et al, 2019). Labelling in the datasets is the primary delineator among these two types. Predictive functions are derived from the labelled training datasets using supervised learning. Inputs and predicted and output values are included in each training case. It can be used to create a predictive model that can accurately predict an output value given a set of input data by analysing the data and attempting to establish correlations between both the inputs and output values.

Conceptual Framework

The conceptual framework was developed based on the predictive model theory and the research objectives. Predictive model theory holds that a model that can predict future outcomes depends on historical data and patterns (Machta et al., 2013). Selecting variables such as (pH and turbidity) relevant to groundwater quality can be used to train a model using historical data and validate and evaluate the prediction performance of a model against established metrics. The theory proposes that groundwater monitoring is expensive, therefore, there should be an effective and efficient way to assess, model and predict groundwater quality and quantity. This is the goal of the current study.

Performance evaluation refers to the process of assessing how well the predictive model performs in predicting groundwater chemical parameters (Singha et al., 2021). It involves using various metrics, such as accuracy, R^2 and RMSE, to measure the model's effectiveness in providing accurate predictions (Singha et al., 2021). Predictability assessment involves determining the predictability of groundwater chemical parameters using easily measured water quality parameters (Mahapatra et al., 2012). Statistical significance refers to examining whether there are meaningful and statistically significant differences between the observed (actual) groundwater chemical parameters and the ones predicted by the model (Raiber et al., 2012). Model comparison involves comparing the performance of the developed supervised machine learning model with traditional aqueous geochemical models (Visual Minteq, Phreeq C, and Wateq4F).

The conceptual framework demonstrates that evaluating and assessing the model with metrics (such as accuracy, R-square, and RMSE) helps determine its performance in predicting the groundwater chemical parameters (using easily measured parameters). Furthermore, when the model performance is compared with models like geochemical models (such as Visual Minteq, Phreeq C and Wateq4F), it would help validate its predicting ability in groundwater chemical parameters. This helps establish the statistical significance of the model in groundwater quality assessment and management.

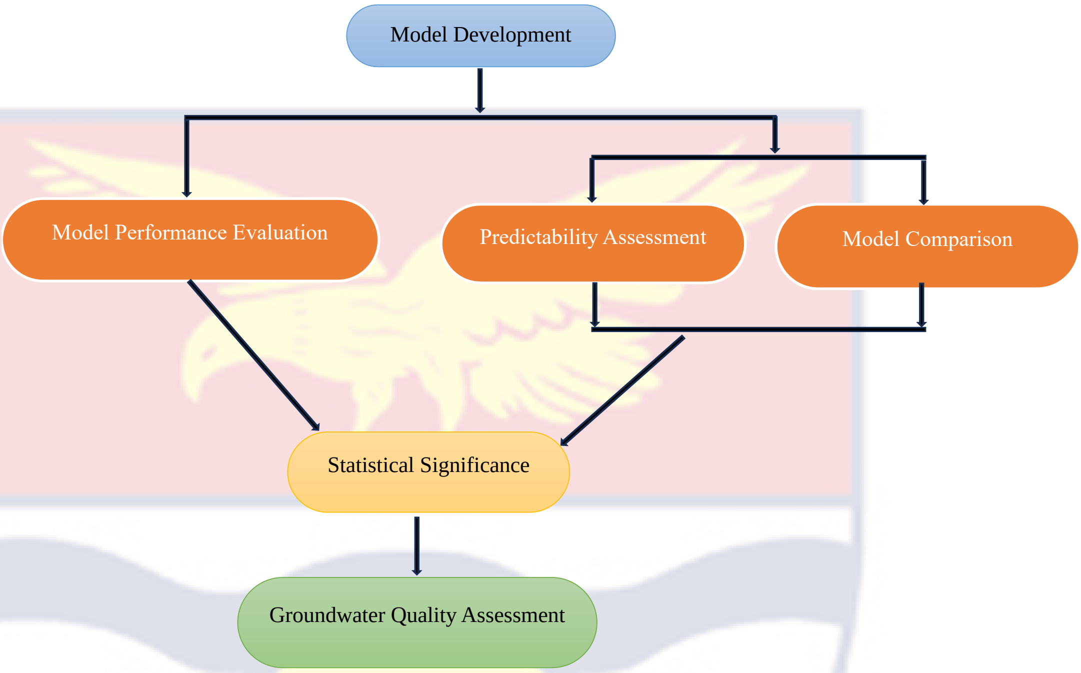

Figure 2.1: Conceptual Framework

Empirical Review

Groundwater quality

Ibrahim (2019) evaluated the fitness of groundwater for drinking in Jordan's major groundwater basins. All physical and chemical elements were nearly below the maximum permitted level from the findings. The microbiological parameter (i.e., E. coli count) surpassed the maximum permissible limit in all tested locations. E. coli was the most influential parameter determining WQI. The total hardness (TH) of groundwater samples measured in $CaCO_3^-$ ranges from 23 to 861 mg/l. The turbidity value was less than 1 NTU. The sulphate concentration ranged from 9 to 605 mg/l. The chloride ion (Cl) value ranged from 24 to 249 mg/l. The maximum permitted amount of Chloride in drinking water was 200 mg/l, with a higher limit of 500 mg/l. Nitrate ion (NO_3^-) concentration ranges from 1 to 61 mg/l. Ibrahim et al. (2019) asserted that nitrate levels above the maximum allowable limit of 50 mg/l could cause infants health-related issues, such as methemoglobinemia. In addition, the author recommended a low Fluoride (F) concentration intake since it helps protect children and adults against dental caries. The presence of heavy metals in drinking water, likewise, could pose health risks at low concentrations.

Sakizadeh et al. (2016) used ANN with regression analysis to establish the nonlinear associations among the groundwater parameters using time series analysis. Sakizadeh et al. (2016) aimed to predict WQI using ANN for the concentrations of 16 groundwater quality variables collected from 47 wells and springs in Andimeshk during 2006–2013 by Iran's Ministry of Energy. These included TDS, EC, pH, turbidity, calcium, total hardness, Mg, PO_4^{3-}, SO_4^{2-}, F^-, NO_3^-, Fe^{2+}, Cu^{2+}, Mn and Cr^{4+}. They believed such a prediction could reduce the computation time and effort and the possibility of error in the calculations. To calculate the water quality index, Sakizadeh et al. (2016) assigned each groundwater quality parameter a weight. Sakizadeh et al. (2016) calculated the groundwater index using the formula

$$QWI = \frac{\sum qiwi}{\sum WI} \qquad [2.3]$$

where qi is the rating scale,

wi is the unit weight of each water quality parameter.

The study's findings showed that techniques like Bayesian regularisation and Ensemble approaches could help solve this problem. Water quality index prediction was effectively employed using Bayesian regularisation and Ensemble averaging approaches. On the other hand, the performance of

Bayesian regularisation was roughly better, with a minimum test error showing that these methods have a strong generalisation capacity in this field. The authors did well and produced quality work covering 16 water indicators. Notwithstanding, they did not include temperature, salinity and ions like Cl, HCO_3 and Na, which sometimes influence groundwater quality.

Roy et al. (2021) conducted a study using the water quality index (WQI) and multivariate statistical methods to assess the water quality and connected environmental impact of 38 groundwater samples. Techniques such as entropy weight WQI and weighted arithmetic, PCA, cluster analysis (CA), correlation analysis (R) and spatial mapping were used in the study. According to estimated WQI values, iron (Fe), dissolved oxygen (DO), and turbidity were the most important criteria for the non-potability of groundwater samples after evaluating the physicochemical characteristics of each sample. Thirty-four (34) samples had Fe contamination levels above the WHO recommended range (0.3-1 mg/l), with the highest value being 15.23 mg/l. The correlation matrix (R) indicates that there is a common source and geochemical processes for all ions (Roy et al., 2021). The PCA data and screen plot diagram that followed them showed three crucial components with a total variance of 84.5 percent, illuminating the reasons behind the deterioration of water quality. The methodologies for identification and management, along with data ranges for all parameters, were used to create the spatial distribution map. To examine three crucial categories, cluster analysis (CA) used Ward’s method, which involved sampling locations, parameter analysis, and dendrogram plotting. In light of the aforementioned, it is advised to periodically evaluate physicochemical parameters in order to protect water resources and emphasise management strategies for maintaining water quality.

Machine learning

Yadav et al. (2020) predicted and forecasted the groundwater levels of various ML techniques in India. The authors took their dataset from a trusted resource and used it to analyse groundwater levels in the study area. They trained KNN and removed the clustering method that was unwanted and also removed irrelevant datasets. They analysed the pre-processed dataset with the random forest algorithm and used charts and plots to predict the present and future groundwater levels of the study area. The bootstrap sample from S for each tree in the forest is selected. Before altering the procedure and picking the subset of features f F, S I denote the ith bootstrap (where F is the set of features). Instead of splitting on F, the node splits on the best feature in f. F is substantially smaller in practice than f. The random forest algorithm method analysed all the factors and whole attributes from the dataset. The findings showed that the random forest algorithm predicts the groundwater level very well, and data-driven modelling approaches perform sufficiently well in predicting groundwater level changes. The study of Mohammad et al. (2020) was limited to arid and semi-arid regions where the groundwater resource is highly utilised for agriculture, industry, and municipal purposes; thus, it could not explain groundwater variability in tropical groundwater and rainforest areas like Ghana.

Shiri et al. (2021) developed an alternative approach called Integrative firefly machine learning models (ANN-FA and SVM-FA). They also used metaheuristic algorithms like the Genetic Algorithm (GA) to train machine learning models like the ANN and SVM. The Firefly Algorithm (FA) was inserted in the MLP and SVM models to build the integrative models MLP-FA and SVM-FA. The FA algorithm was inspired by the natural behaviour of fireflies, which attract each other through flashing. As a result, each firefly indicates a different answer. The objective function is introduced by the intensity of each firefly's light—lower-intensity fireflies (Xj) trail higher-intensity fireflies (xi). For each well, statistical indicators were calculated. Hence, the global indicators were calculated by averaging the values from the wells investigated. Shiri et al. (2021) then input data into the model and standardised the dataset between 1 and 0. Results showed that AI models could be used to model groundwater quality characteristics. The study was well conducted, but the authors did not include groundwater parameters such as pH, turbidity, temperature, sulphate ions, etc.

Singha et al. (2021) conducted a study to model and predict the state of safe drinking water sources in the future and understand the quality and pollution levels existing in groundwater. In the Raipur district of Chhattisgarh, India, groundwater samples totalling 226 were taken from a region with a high concentration of agricultural land. An entropy weight-based groundwater quality index was created by evaluating various physicochemical properties (EWQI). A deep learning (DL) model was employed by Singha et al. (2021) to forecast the quality of groundwater. The authors compared it to three different machine learning (ML) models: random forest (RF), eXtreme gradient boosting (XGBoost), and artificial neural network (ANN) (ANN). Five error measures were used to assess the effectiveness of models in terms of prediction. Results revealed that the DL model has the highest accuracy in terms of R^2, with an R^2 of 0996, compared to the RF (R-square = 0.886), ANN (R^2 = 0.917) and XGBoost (R^2 = 0.0.927). The proposed technique is cross-verified for the uncertainty of the DL model output by running it ten times using a newly randomised dataset, where slight differences in the mean value of performance metrics are noticed. Furthermore, the relevance of input variables computed by prediction models reveals that the DL model is the most realistic and accurate approach to groundwater quality prediction. As a result, they proposed using artificial intelligence to predict water quality because it has high accuracy in controlling pollution and improving water quality.

Elbeltagi et al. (2021) employed four independent approaches to predict WQI using a variable elimination methodology: M5P tree model (M5P), additive regression (AR), support vector machine (SVM) and random subspace (RSS). The datasets were separated into two classes, in a ratio of 80:20, for model building the training dataset and verifying the dataset, using a fivefold cross-validation approach (testing dataset). The models were assessed using numerical and graphical assessment methods. Traditional WQI computation takes longer when calculating sub-indices and frequently results in huge inaccuracies. The optimum input was varied during the training and validation stages, with the ideal input parameters (TDS, pH, EC, Ca, Cl and Mg) being the most common. According to the results (MAE = 0.5243, R-square = 0.9993, percent RAE = 3.8449, RMSE = 0.6356 and RRSE percent = 3.9925), AR outperformed the other data-driven models (MAE = 0.5243, R-square = 0.9993, percent RAE = 3.8449, RMSE = 0.6356 and RRSE percent = 3.9925). The authors established that AR yielded a better result than data-driven models in assessing and modelling groundwater quality parameters like EC, pH, Cl, Mg, TDS and Ca which was also confirmed by Egbueri and Agbasi (2022) in Nigeria.

According to Egbueri et al. (2022), it is very necessary to simulate and model water quality parameters. Furthermore, numerous modelling methodologies yield more robust and dependable insights than a single model. For the modelling of groundwater quality parameters such as EC, pH, TH and TDS, computational methods such as modified heavy metal index (MHMI), synthetic pollution index (SPI) and pollution load index (PLI) were compared. The groundwater resources were physiochemically analysed using standard procedures. Both anthropogenic and non-anthropogenic activities were shown to impact the concentrations of the water quality measurements. From the polynomial regression, the findings of the water quality measurements are significantly connected. The principal component analysis and varimax-rotated factor analyses of lead, zinc and nickel affect the perception of water quality metrics. The pH, TDS, EC and TH values were used to group the water samples using Q-mode hierarchical and K-means clustering methods. Supervised learning algorithm and Multiple linear regression (MLR) techniques simulated and predicted TDS, pH, TH and EC. Regarding forecasting pH, the ANN model outperformed the MLR model.

Kouadri et al. (2021) used eight artificial intelligence algorithms to predict WQI in the Illizi region of Southeast Algeria: random forest (RF), multilinear regression (MLR), random subspace (RSS), M5P tree (M5P), artificial neural network (ANN), additive regression (AR), locally weighted linear regression (LWLR), and support vector regression (SVR). The work technique was concentrated on two scenarios and twelve possible input combinations were produced using best subset regression. In the first scenario, all parameters were utilised as inputs to shorten the time to calculate WQI. In the second scenario, water quality fluctuated in critical situations when relevant assessments were unavailable, and all inputs were reduced based on sensitivity analysis. The models were assessed utilising correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE) and relative absolute error (RAE). TDS and TH were the key factors influencing WQI in the research region. The MLR model outperforms other models by 3.1708×10^{-8} percent, 1.2573×10^{-10} percent, 2.1418×10^{-8}, 1.4572×10^{-8}, and 1 for RRSE, RAE, RMSE, MAE, and R, respectively. The RF model was utilised in the second scenario, and the error rates for RRSE, RAE, RMSE, MAE, and R were 5.9642, 4.693, 3.2488, 1.9942 and 0.9984, respectively. They concluded that the study's findings would help water managers improve sustainable groundwater resource management plans in terms of WQI.

AI was employed by Aldhyani et al. (2020) to measure water quality (WQ). According to the researchers, AI can forecast water quality patterns and track seasonal variations in WQ. On the other hand, using multiple models to predict the WQ produces better results than just one. Aldhyani et al. (2020) developed advanced AI algorithms to forecast the water quality index (WQI). They also used artificial intelligence models such as the long short-term memory (LSTM) and supervised learning algorithms like support vector machine (SVM), Naive Bayes (NARNET) and K-nearest neighbour (K-NN). They used the models they had constructed to examine the dataset. They demonstrated that the generated models could accurately classify water quality and predict WQI due to their greater resilience. The anticipated findings showed that the LSTM did not perform well in predicting the WQI compared to the NARNET model. For the WQC prediction, the SVM method and the WQI value had the lowest accuracy (94.01%). Furthermore, the LSTM and NARNET models obtained equivalent accuracy to check the point which is a little different from the regression coefficient.

Summary of Chapter

This chapter reviewed literature on groundwater quality and pollution, including substances that contribute to groundwater contamination. It highlights groundwater pollution as a global issue affecting human health and ecological well-being worldwide. Consequently, there is a need to assess the chemical, physical and biological attributes of groundwater to gauge its quality accurately. To address this, both artificial intelligence and traditional approaches have been employed. AI methods such as MLA, incorporating decision tree regression and polynomial regressions, offer high accuracy, cost-effectiveness, and the ability to model nonlinear relationships among groundwater parameters. However, existing studies often employ complex and time-consuming methods that overlook easily measurable parameters for predicting and modelling groundwater quality. No work has compared the performance of SML with geochemical models, nor has attempted to predict groundwater quality using easily measured parameters. Addressing this gap, the study would employ SLA with decision tree regression and polynomial regression techniques to model groundwater quality using easily measured parameters.

CHAPTER THREE

MATERIALS AND METHODS

Introduction

This chapter consists of the study area, research design, sampling technique, data types for the study, sampling procedure, data processing and analysis and summary of the chapter.

Study Area

Central Region can be found in the south-western centre of Ghana. It shares a boundary with the Ashanti Region on the north, Eastern Region on the north-east, Greater Accra Region on the south-east, Western Region on the west, and bounded by the Gulf of Guinea on the south. The coastline stretches to about 150km (Acheampong et al., 2017). Central Region is situated at approximately 5.8980° N latitude and 1.0408° W longitude. Central Region has primary patterns of rainfall, that is, main and minor rainy seasons. The main rainy season ranges from April to July and September to November makes the minor rainy season. The relative humidity of the region is between 50 and 85 percent (Mohammed et al., 2022).

Greater Accra Region is the capital region of Ghana. Greater Accra Region lies at around 5.6037° N latitude and 0.1870° E longitude. The region occupies a land area of 245km of the total land of Ghana (Ampim et al., 2021). It is one of the most populated and developed cities in Ghana, having 87.4 percent of its inhabitants staying in the city. Many of the economic activities are owned by private individuals. Primary occupations found in the region include farming as well as office jobs.

Western Region is located in the southern part of Ghana and shares a boundary with Ivory Coast on the west and Central Region on the east, with the capital city being Sekondi-Takoradi. Western Region is located at approximately 5.3487° N latitude and 1.9822° W longitude. Western Region comprises an area of 13,842 sq. km, has a population of 2,060,585, and has the highest rainfall in Ghana, with lush green hills and fertile soils (Boakye et al., 2023). Vegetation cover of the region occupies about 75 percent of its total land. Some of the occupations that can easily be found in the region consist of farming, animal husbandry and fishing.

Central, Greater Accra and Western Regions are noted for high population, water pollution, illegal mining, and domestic and agricultural activities, escalating the water pollution issues (Yeleliere et al., 2018; Amuah et al., 2022). In these regions, domestic water supply accounts for around 95 percent of groundwater use, especially in rural and small towns. Approximately 41 percent of families rely on groundwater as their primary water source (Livingston, 2021; Amuah et al., 2022). This means that water sustenance is key to ensuring healthy lives and economic growth in the regions.

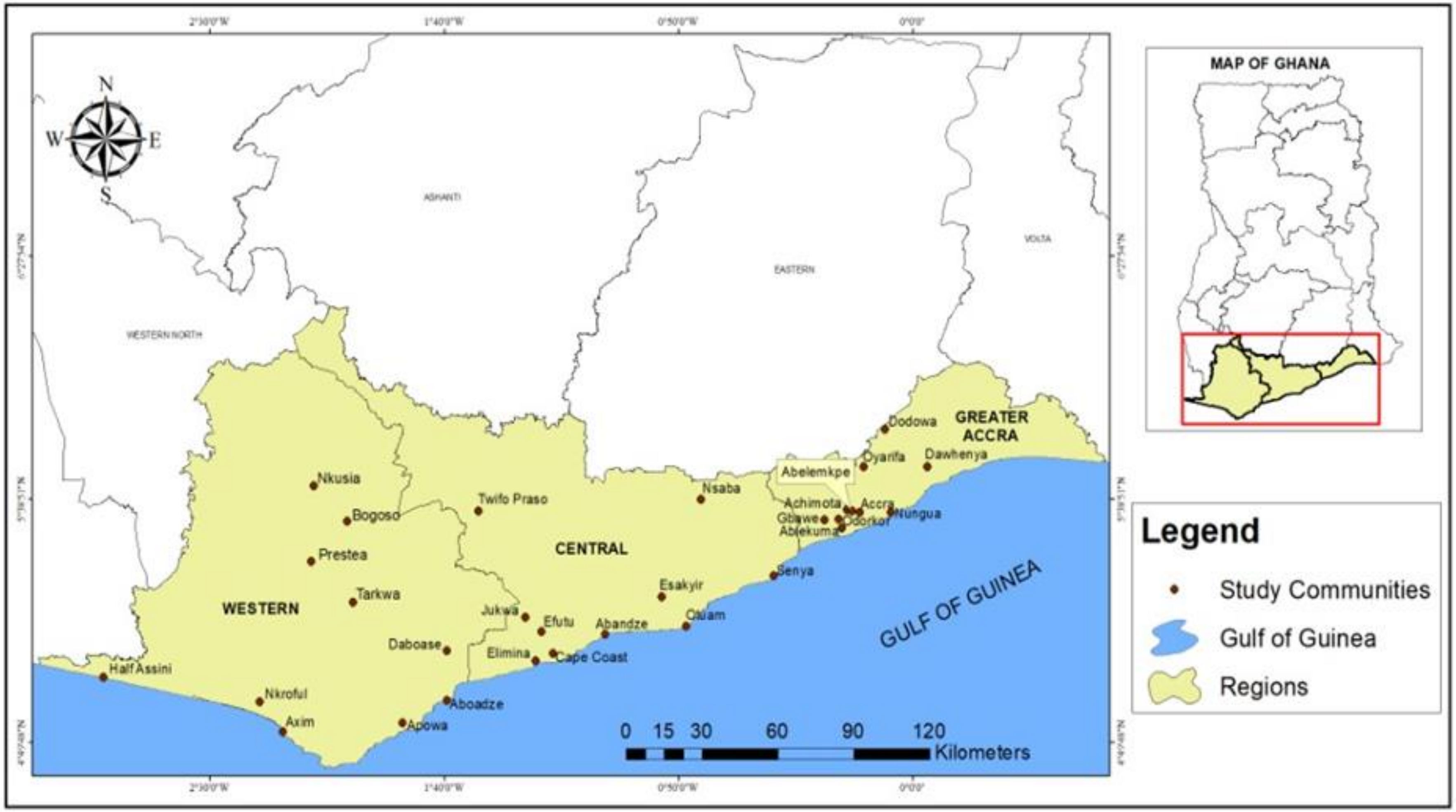

Figure 3.1: Map of the Site (Regions)

Research Design

This study employed an experimental research design, particularly factorial design, for the study. Factorial design is an experimental approach used for assessing the influence of multiple independent variables on a dependent variable (Kritikos et al., 2019). This design allows the use of quantitative data to conclude (predict groundwater quality), validate the hypothesis (analysis testing) and establish causality (what contributes to groundwater quality pollution) (Randall et al., 2013). According to Watkins and Newbold (2020), though factorial design helps examine multiple factors and their interactions, a larger sample size is required to make the design. Chong et al. (2021) explained that adequate sample sizes help detect small or subtle effects, particularly when investigating interactions between variables.

SML uses large data to predict new output values, evaluate performance and create, train and quantitatively test models (Shouval et al., 2021). SML uses labelled datasets to train algorithms to classify data or predict outcomes accurately under supervision (Meng et al., 2020) to achieve accurate and reliable results (Meng et al., 2020). The design allows the use of analytical techniques, such as factorial ANOVA or regression analysis for accurate interpretation of results. It enhances the generalisability of findings by capturing the multifaceted nature of real-world phenomena (Shouval et al., 2021), yielding results that are more robust and ecologically valid.

This design aligns with the study because it would help develop a model to predict groundwater quality chemical parameters using easily measured parameters. It would help compare the performance of the model with aqueous geochemical models (such as Visual Minteq, Phreeq C and

Wateq4F) in predicting groundwater quality parameters. This design would help establish if there were statistically significant mean differences between and among the predicted groundwater chemical parameters of the study regions.

Sampling Technique

This study employed a random sampling technique, particularly a stratified random sampling technique. This sampling technique involved dividing the study area into strata based on specific characteristics and then randomly selecting wells within each stratum (Ganesha & Aithal, 2022). This means that the towns in the study regions had an equal chance of being sampled. The stratified random sampling technique was chosen to ensure high internal validity, as randomisation minimises the potential influence of confounding groundwater parameters.

The study required a large dataset, therefore, this sampling technique had high external validity to mimic and represent the characteristics of overall groundwater quality in each of the selected regions. Moreover, this method was selected because it required minimal advanced knowledge about groundwater quality but the findings of the study reflect the overall features of the groundwater quality in each of the study regions.

Data Types for the Study

The dataset used in this study included secondary and primary data. These data types are in the form of quantitative data. The secondary data had already been gathered over time from the study areas. The primary data were gathered through field data collection.

Secondary data

The secondary data consisted of a dataset of 3,134 samples and 53,278 variables on groundwater from the three regions (from online data on the selected regions to train the model). The dataset included variables such as dissolved oxygen (DO/mg/l), pH, electrical conductivity (EC/mS/cm), total dissolved substances (TDS/mg/l), potassium (K^-/mg/l), nitrate (NO_3^-/mg/l), hydrogen carbonate (HCO_3^-/mg/l), sodium (Na^+/mg/l), magnesium (Mg^{2+}/mg/l), total hardness (TH/mg/l), calcium (Ca^{2+}/mg/l), temperature (^oC), silicate (SiO_3^-/mg/l), salinity (mg/l), phosphate (PO_4^{3-}/mg/l), sulphate (SO_4^{2-}/mg/l), chlorine (Cl^-/mg/l) and turbidity (NTU).

Primary Data Collection

Nine hundred bottles were purchased (three bottles for each well) to collect water samples for the study. The bottles were washed with hot water and dried under the sun to remove any chemicals or agents that might contaminate the water quality. Each bottle was rinsed with groundwater samples collected at each sample location. Three regions were considered; however, one hundred wells each were sampled from each region. Ten towns were considered in each town as shown in Table 3.1.

Table 3.1: Towns Sampled from Each Region

Western	Central	Greater Accra
Aboadze	Abandze	Ablekuma
Apowa	Cape Coast	Abelemkpe
Axim	Efutu	Accra
Bogoso	Elmina	Achimota
Daboase	Esaba	Dawhenya
Half Assini	Esakyir	Dodowa
Nkroful	Jukwa	Gbawe
Nkusia	Otuam	Nungua
Tarkwa	Senya	Odorkor
Prestea	Twifo Praso	Oyarifa

Out of these towns, 10 wells were sampled from each town, summing up to 300 (10 x 10 x 3 = 300)]. At each sampling site, three containers were used to draw the water into each bottle (for replicates) to a depth of 1.5 litres. Easily measured parameters such as pH, temperature, turbidity, electrical conductivity, salinity and total dissolved substances were measured in situ, using a pH meter, thermometer, turbidity meter, digital conductivity meter, hydrometer and digital TDS meter, respectively. Readings were taken three consecutive times for the average value of each parameter to be recorded. This process was repeated for 3 different water samples for each of the wells sampled. Water samples drawn from the wells were transported to the Ghana Water Company Limited, Greater Accra, for the chemical parameters (dissolved oxygen, nitrate, phosphate, sulphate, sodium, calcium, silicate, hydrogen bicarbonate, chlorine, potassium and magnesium).

Laboratory Analysis of Samples

Hach HQ2200 Portable pH/EC/TDS/DO Meter was used in measuring concentrations of pH, EC, TDS and DO of the water samples. The electrode was rinsed, zeroed and immersed in each of the water samples for reading and recording after stabilisation. The concentration of EC was measured in uS/cm, while TDS and DO were measured in mg/l. The concentration of turbidity was determined using Hanna Instruments HI-93102 metre. The probe was immersed in each of the water samples and readings after stabilisation were taken and recorded with duplicates of three. The concentrations of salinity of the groundwater samples were measured using AZ-8372 Salinity Meter. The water samples were poured into a bottle and placed in the turbidity meter.

After allowing the samples to stabilise, readings in NTU were taken and recorded with three repeated replicates.

Total hardness was analysed through titration method. Buffer solution (such as NH_4Cl) and hardness indicator were added to each measured water sample. Subsequent titration such as EDTA was conducted to determine hardness concentration. Sulphate, nitrate, phosphate and chloride were similarly analysed through titration methods. Reagents such as barium chloride were used for sulphate, potassium dichromate for nitrate, ammonium molybdate for phosphate and silver nitrate for chloride and each concentration was read and recorded in mg/l.

Sodium concentrations were determined using a flame photometer method. Each water sample was treated with reagents (such as Sodium chloride) and readings were taken using a spectrophotometer. Potassium concentrations were measured using an atomic absorption spectrometer. A portion of each water sample was pipetted into volumetric flasks and each concentration was recorded in mg/l. Calcium and hydrogen bicarbonate were analysed through EDTA titrimetric methods. Magnesium concentrations were determined using a titrimetric method with an Eriochrome Black T indicator. Silicate concentrations were analysed using the Spectroquant Silicate Test.

Machine Learning Algorithm

Decision tree regression

This current study employed decision tree regression (DTR) in the groundwater quality prediction. DTR is a supervised learning algorithm used for regression tasks, where the goal is to predict a continuous target variable (Rathore & Kumar, 2016). DTR works by recursively partitioning the feature space into smaller regions and fitting a simple model (usually a constant value) in each region (Pathak et al., 2018). DTR was utilised due to its capability to capture intricate connections between characteristics and target variables.

The decision tree regression algorithm builds a tree structure where each internal node represents a decision based on a feature, each branch represents the outcome of the decision, and each leaf node represents the prediction (De Caigny et al., 2018). The prediction at each leaf node in a decision tree regression model is simply the average (or another measure of central tendency) of the target variable for the training samples that fall into the leaf (Sharma & Kumar, 2016). This means that the prediction "*y*" for a given input "*X*" is determined by traversing the tree from the root node to a leaf node based on the values of the input features. At the leaf node, an average value of the target variable values associated with the training samples in the leaf node is used for the prediction. Mathematically, the prediction "*y*" for a given input "*X*" can be expressed as:

$$y = \frac{1}{N}\sum_{i=1}^{N} yi \qquad [3.1]$$

where:

y is the predicted target variable value.

N is the number of training samples that fall into the leaf node.

yi is the target variable value for the training samples that fall into the leaf node.

DTR was applied to predict the concentrations of chemical parameters (sodium, potassium, calcium, chlorine, hydrogen bicarbonate, sulphate, phosphate, silicate and nitrate ions) in the groundwater based on easily measured parameters (turbidity, dissolved oxygen, electrical conductivity, salinity, total hardness, temperature and pH). The process began with data preparation and thus datasets containing all parameters of groundwater (secondary dataset) and easily measured parameters (primary dataset) were prepared. Data exploration and visualisation techniques such as stacked lines, boxplots and correlation heatmaps were employed to understand the data's distribution and relationships between variables.

The dataset was split into training and testing sets. DTR was initialised and trained on the training data to learn the relationships between the input and the target variables. After training, the model was evaluated on the testing data using metrics like RMSE and R^2 to assess its performance in predicting chemical parameters of the groundwater sampled. Once evaluated, the trained model was used to predict the chemical parameters of the groundwater on the primary dataset, wherein DTR predicted the chemical parameters of the groundwater based on easily measured parameters.

Polynomial regressions

A polynomial regression model (PRM) is used in ML to model the relationship between independent and dependent variables (Dalal & Zickar, 2012). It models the relationship as an nth-degree polynomial (Bera et al., 2021). Polynomial regression can be expressed as:

$$Y = \beta_0 + \beta_1 X_1^2 + \beta_2 X_2^3 + \ldots + \beta_n X_n^n + \epsilon \qquad [3.2]$$

Y is the dependent variable/groundwater chemical parameters to be predicted,

β_0 – is the intercept,

$\beta_1, \beta_2, \ldots, \beta_n$ are the respective slopes of the dependent variables,

∈ is the residual (error)/difference between the observed and predicted values of Y

$X_1, X_2, \ldots, X^n$ are the independent variables or groundwater physical parameters

PRM was employed to capture potential nonlinear relationships among the change groundwater parameters using easily measured parameters. For instance, squared terms, cubic terms or even higher-order polynomial terms of the original features were added to each variable (pH, EC, Temperature, DO, turbidity, total hardness and salinity). In the model, easily measured parameters such as pH, EC, Temperature, DO, turbidity, total hardness and salinity served as the input variables (independent variables) when predicting groundwater chemical parameters like sodium, potassium, calcium, chlorine, hydrogen bicarbonate, sulphate, phosphate, silicate and nitrate ions.

Figure 3.2: Model Data Processing Stages

Training and Testing of Model

The model development was divided into two stages: training and testing phases. In the training process, training data was input into the model for the model to learn the patterns in the dataset. The learning model used the execution engine to predict the test variables. From the above, SL algorithm involved data input, machine processing and prediction, as shown in Figure 3.2.

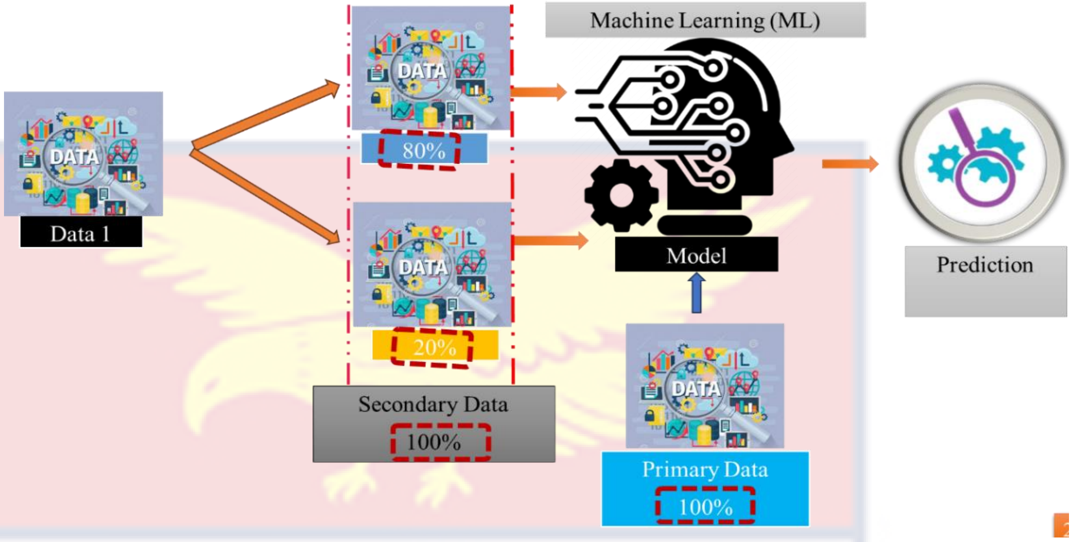

Figure 3.3: Training and Testing of Model

The functions were created for the model to work. The data was imported into the model and asked the system to pick 80 percent of the secondary data for the training and 20 percent together with 100 percent of the primary for testing the model. The model was made to learn the data science kits in order to ensure accurate prediction. Other functions were created for the testing or the prediction of the data. At this stage, the algorithms displayed the accuracy of the predicted ions and graphs or plots were generated. This was where the test scores were output from the model for the calculation of the groundwater quality index.

Independent variables were turbidity, total hardness, dissolved oxygen (DO), pH and electrical conductivity (EC), salinity and total dissolved substances (TDS), while dependent variables were potassium (K^-/mg/l), nitrate (NO_3^-/mg/l), hydrogen carbonate (HCO_3^-/mg/l), sodium (Na^+/mg/l), magnesium (Mg^{2+}/mg/l), calcium (Ca^{2+}/mg/l), silicate (SiO_2/mg/l), salinity (mg/l), phosphate (PO_4^{3-}/mg/l), sulphate (SO_4^{2-}/mg/l) and chlorine (Cl^-/mg/l).

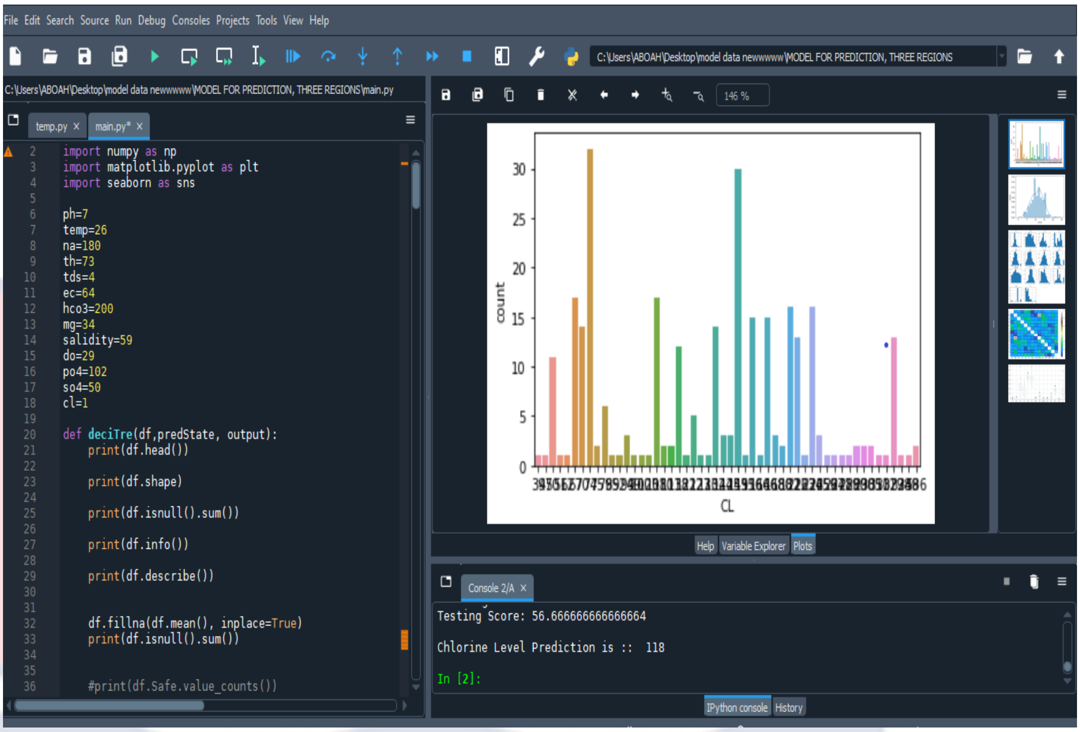

Figure 3.4: Graphical view of running the model in Spyder (Anaconda) version 3.9.7

Performance Evaluation

ML models have many performance evaluation tools. The performance of the current model was enhanced by optimising parameter combinations through genetic algorithms. RMSE, R^2 and De-statistics were used to determine the performance of the model.

Root mean squared error (RMSE)

The Root Mean Squared Error was used to evaluate the model prediction error for comparing regression models. It showed the average difference between the observed known outcome values and the model's anticipated value (Ćalasan, 2020; Karunasingha, 2022). RMSE is calculated using the formula:

$$\text{RMSE} = \sqrt{\frac{\sum(P-O)^2}{n}} \qquad [3.5]$$

Where RMSE = room means square error

P = predicted means of the chemical ions

O = observed means of the field data

n = the number of groundwater parameters

The smaller the RMSE, the higher the model's ability to describe the dependent variables. The higher RMSE means that the model has a low predicting ability, thus, independent variables cannot perfectly explain the dependent variables.

Index of agreement (D-Statistics)

D-statistics was used to evaluate the degree to which the model can predict without error. It is a ratio of the mean square error and the potential error (Malinsky et al., 2021). The index of agreement can be calculated from the formula:

$$\text{D-Statistics} = \sqrt{\frac{\sum(P-O)^2}{\sum(|P|-|O|)^2}} \qquad [3.6]$$

Where P = predicted means of the chemical ions

O = observed means of the field data

n = the number of groundwater parameters

D-statistics value varies from zero to one. D-statistics of 1 implies that the model can be used to predict groundwater quality to perfection and 0 connotes that there is no agreement between the predicted and observed values, therefore, the model cannot be used to predict groundwater quality.

Data Analysis

The data analysis was divided into field data analysis and model prediction analyses.

Field data analysis

Descriptive statistics such as mean, minimum, maximum and standard deviation were used to understand the physicochemical parameters sampled from the three regions. Mean represented the average value, minimum and maximum denoted the range of values and standard deviation indicated the variability from the mean. The mean values of the groundwater assessed were compared with WHO limits for safe drinking water (as shown in Table 3.2).

Table 3.2: Groundwater Quality Standards of World Health Organisation (2022, 2002 and 2004)

Parameters	WHO
pH	6.5-8.5
Temperature/^{0}C	25-33
Na ions/mg/l	200
Cl ions/mg/l	250
TH/mg	< 60
TDS/mg/l	< 300
K ions/mg/l	300
Mg ions/mg/l	50
SO_4 ions/mg/l	250
EC/mg/l	400
Ca ions/mg/l	100
SiO_3/mg/l	100
Turbidity/NTU	< 5
DO/mg/l	6.5 – 8
PO_4 ions/mg/l	2
NO_3 /mg/l	50
HCO_3 /mg/l	10

Where NTU = Nephelometric turbidity unit, uS/cm = micro simen per centimetre, ^{o}C = degree Celsius, and mg/l = milligram per litre. DO = dissolved oxygen, TH = total hardness, EC = electrical conductivity, TDS = total dissolved substances.

Wilcox diagram was employed to identify the main environmental factors (precipitation, rock and evaporation) that influence the wells sampled. Gibbs's diagram was used to determine the suitability of the wells for drinking. PCA was used to account for the factors that contributed to the variations in the groundwater dataset (the loading effects of the groundwater parameters). The eigenvalues of the principal components showed the decomposition value of the groundwater data matrix. The eigenvectors (principal components) explained the directions of the new feature space and the eigenvalues determined their magnitude.

Model prediction analyses

Objective 1: To develop a model and determine its performance in predicting groundwater chemical parameters. Research Objective 1 was analysed using RMSE and regression analyses (r-square). High accuracy indicated better model performance, thus, when the accuracy is above 79% or 0.79. A lower RMSE indicated better model performance or smaller prediction errors. A higher R^2 value (0.8 and above) signified better model fit, indicating the percentage variability in groundwater chemical parameters predicted using easily measured parameters.

Objective 2: "To compare the performance of the model with aqueous geochemical models (such as Visual Minteq, Phreeq C and Wateq4F) in predicting groundwater quality parameters. Research Objective 2 was analysed using RMSE, R^2 and d-statistics. High accuracy indicates better model performance, thus, when the accuracy is above 79% or 0.79. A lower RMSE suggests fewer prediction errors and improved model performance. A higher R-square value (0.8 and above) signified a better model fit, indicating the percentage variability in groundwater chemical parameters explained by the easily measured parameters. It offered insights into the goodness-of-fit of the model. When d-statistics value is close to 1, the performance of the model is high.

Objective 3: "To establish if there are statistically significant mean differences between and among the predicted groundwater chemical parameters of the study regions." ANOVA test was used to analyse Research Objective 3. The ANOVA test was conducted to determine whether there were significant mean differences between and among the predicted groundwater parameters of the study regions. An alpha-value of 0.05 was used, meaning there is 95 percent certain that the dataset represents the variations in the groundwater data. Furthermore, p-value less than the alpha value (p-value < 0.05) indicated statistically significant mean differences between the observed and the predicted groundwater quality chemical parameters.

Calculation of Groundwater Quality Index

The water quality index (WQI) was calculated to determine the water quality. The calculation was based on the weighted arithmetic index method (Brown et al.,1972). The unit weight factors (U) of each variable were calculated using

$$\text{Weight factors (U)} = \frac{u}{M_S}, \qquad [3.1]$$

$$\text{where U} = \frac{1}{\frac{1}{Ms}} \qquad [3.2]$$

where MS is the WHO standard accepted value for the parameters as shown in Table 3.2. The sub-index value for each parameter (A) was calculated using

$$\left(\frac{H1-Ho}{Ms-Ho}\right) x\ 100\%, \qquad [3.3]$$

where Ms is the standardly accepted value of the n^{th} parameters

H1 is the mean concentration of the n^{th} parameters

Ho is the actual value of the parameter in pure water (usually Ho = 0 g/l, except pH).

$$WQI = \frac{\sum UA}{\sum U}. \quad [3.4]$$

'U' and 'A' were multiplied and divided by the total outcome and then divided by the summation of 'U' to get the groundwater quality index. Finally, the calculated groundwater quality index was compared with WHO standard values as shown in Table 3.2.

Table 3.3: Groundwater Quality Index (WQI) Classification

WQI range	Water Types	Interpretations
<50	Excellent water	Water quality is almost unaltered, and the situation is nearly perfect.
50–100	Good water	Water quality is maintained with only minor degradation; conditions seldom depart from tolerable levels.
100.1–200	Poor water	Water quality is almost always reduced, and it is barely good.
200.1–300	Very poor water	Parameters are not near to being clean.
>300	Water unsuitable for drinking	Water quality is often compromised, and the characteristics are far from optimum.

Source: Agrawal et al. (2021).

Ethical Considerations

Permission to conduct fieldwork was obtained from the Institutional Review Board of the University of Cape Coast. To gather data, the researcher first submitted a copy of the sampling procedure to the University of Cape Coast and Institutional Review Board (IRB) for review. The study strategy of the investigator was in line with the Research Ethics Policy of the University of Cape Coast.

Summary of Chapter

The chapter presents information on the study area, data acquisition and analysis procedures and calculation of the groundwater quality index. The section reflected the root mean square, ANOVA, PCA and WQI calculation. Standard values such as WHO standards values were presented, as well as, the Gibbs and Wilcox diagrams for determining environmental factors that influence groundwater ions in the study regions.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This chapter presents and discusses results from the field data analysis. The chapter uses and displays analyses like PCA and ANOVA, as well as tables and charts of the findings.

Results from Field Data Analysis

This section presents the descriptive statistics of the physicochemical parameters of the groundwater sampled from the study regions. It includes PCA and Wilcox and Gibbs's diagrams, showing the main environmental factors that influence groundwater quality.

Descriptive Statistics of the Physicochemical Parameters in Groundwater

The physicochemical parameters of water are usually used in determining the quality, type and nature of groundwater. One hundred wells each from Central, Greater Accra and Western Regions were sampled and analysed. From each well, 3 replicates were taken for the physicochemical parameters as shown in Table 3. It presents the means, standard deviations and skewness as well as minimum and maximum values of the physicochemical parameters.

Table 4.1: Descriptive Statistics of Physicochemical Parameters of Groundwater Sampled (Central, Greater Accra and Western Regions) (n = 300)

Parameters	Mean	WHO Standards	Std.	Skewness	Minimum	Maximum
DO	7.20	6.5 – 8	1.57	-0.09	1.84	11.76
pH	6.82	6.5-8.5	1.09	0.95	5.00	12.40
Temperature	28.40	25-33	2.52	0.05	23.60	33.10
Total Hardness	159.58	< 300	94.84	-0.19	13.85	372.00
TDS	453.77	< 300	30.52	0.60	376.47	555.67
EC	426.51	<400	82.78	0.23	245.86	672.56
Turbidity	8.11	< 5	3.02	0.02	2.43	13.87
Na^+	242.44	<200	113.28	0.51	33.00	675.00
K^+	124.92	<300	35.37	0.15	21.00	239.00
Ca^{2+}	171.76	<100	48.06	0.71	42.00	421.00
Mg^{2+}	52.27	<50	28.19	1.42	3.00	127.00
SO_4^{2-}	82.14	<250	17.94	0.69	60.00	261.00
PO_4^{3-}	3.97	<2	1.02	0.30	1.00	5.12
Cl^-	147.72	<250	80.76	1.65	39.00	586.00
HCO_3^-	10.91	<10	3.30	-0.02	3.00	22.00
SiO_3^{2-}	48.69	<100	18.58	0.45	8.00	133.00
NO_3^-	42.38	<50	8.12	0.90	7.00	76.00

Table 4.1 shows descriptive statistics of the physicochemical parameters of groundwater sampled from the study regions. It compares the means of the physicochemical parameters with WHO water quality standards for safe drinking water. From the Table, the mean dissolved oxygen (DO) was 7.20 ± 1.57 gm/l and ranged from 1.84 mg/l to 11.76 mg/l, with skewness of -0.09 mg/l. The pH value ranged from 5 to 12.40, with a mean value of 6.82 ± 1.09 and skewness of 0.95. The mean pH value of the groundwater sampled from all the regions met the WHO standard pH value. The range, the mean

values, standard deviation and skewness of the temperature (^oC) were 23.6 ^oC to 33.10 ^oC and 26.91 ± 2.52 ^oC, 2.52 ^oC and 0.05 ^oC, respectively.

The mean total hardness (TH) of 159.58 ± 2.52 94.84 mg/l ranged from 13.85 mg/l to 372.00 mg/l and had skewness of -0.193 mg/l. The mean total dissolved substance (TDS) was 453.77 ± 30.52 mg/l and ranged from 376.47 mg/l to 555.67 mg/l. Total dissolved substances had skewness of 0.60 mg/l. The mean electrical conductivity of 426.51 ± 82.78 mS/cm ranged from 245.86 mS/cm to 675.56 mS/cm and skewness of 0.23 mS/cm. Turbidity ranged from 2.43 NTU to 13.85 NTU, with a mean value of 8.11 ± 3.02 NTU and skewness of 3.02 NTU and 0.02 NTU, respectively.

The mean sodium (Na) ion concentration was 242.44 ± 113.28 mg/l, ranging from 33.00 mg/l to 675.00 mg/l and skewness of 0.15 mg/l. The potassium (K) concentration ranged from 21.00 mg/l to 239.00 mg/l, with a mean value of 124.92 ± 35.37 mg/l and skewness of 0.15 mg/l. Mean calcium (Ca) of 171.7 ± 48.06 mg/l ranged from 42.00 mg/l to 421.00 mg/l, leaving and skewness of 0.71 mg/l. The mean magnesium ion (Mg) was 52.27 ± 28.19 mg/l and ranged from 3.00 mg/l to 127.00 mg/l, with skewness of 1.42 mg/l.

The mean sulphate ions of 82.14 ± 17.94 mg/l ranged from 60.94 mg/l to 126.00 mg/l and had skewness of 0.69 mg/l. The phosphate (PO_4^{3-}) concentration ranged from 1.00 mg/l to 5.12 mg/l, with a mean value of 3.17 ± 1.02 mg/l. The skewness of the phosphate ions was 1.61 mg/l and 0.30 mg/l, respectively. The mean chlorine concentration was 147.72 ± 80.76 mg/l and ranged from 39.00 mg/l to 586.00 mg/l, leaving skewness of 1.65 mg/l.

The mean hydrogen bicarbonate (HCO_3^-) ion concentration of 10.91 ± 3.30 mg/l ranged from 3.00 mg/l to 22.00 mg/l and had skewness of -0.02

mg/l. The mean silicate ion concentration of 48.69 ± 18.58 mg/l ranged from 8.00 mg/l to 133.00 mg/l, with skewness of 0.45 mg/l. The mean nitrate ion (NO_3^-) concentration was 42.38 ± 8.12 mg/l and ranged from 7.00 mg/l to 76.00 mg/l, with skewness of 0.90 mg/l.

Wilcox and Gibb's Diagrams Display the Classification of Groundwater

The study explored the factors that influence the concentration of ions in the groundwater of the study regions. Gibbs (on the left) and Wilcox (on the right) diagrams with three distinct areas (precipitation, rock and evaporation dominance) were employed to demonstrate the source of ions in the groundwater. Gibbs ratio $Na^+/(Na^++Ca^{2+})$ for cations and $Cl^-/(Cl^-+HCO_3^-)$ and Wilcox ratio (sodium %) for anions of water samples were plotted separately against the respective TDS. These show the types of dominance controlling the groundwater quality. Gibbs diagram shows the chemical interaction between rock-forming minerals of the aquifer and the groundwater is the main mechanism in contributing ions to the groundwater. Wilcox diagram helps evaluate water quality. The following graphs represent the distribution and factors contributing to groundwater pollution.

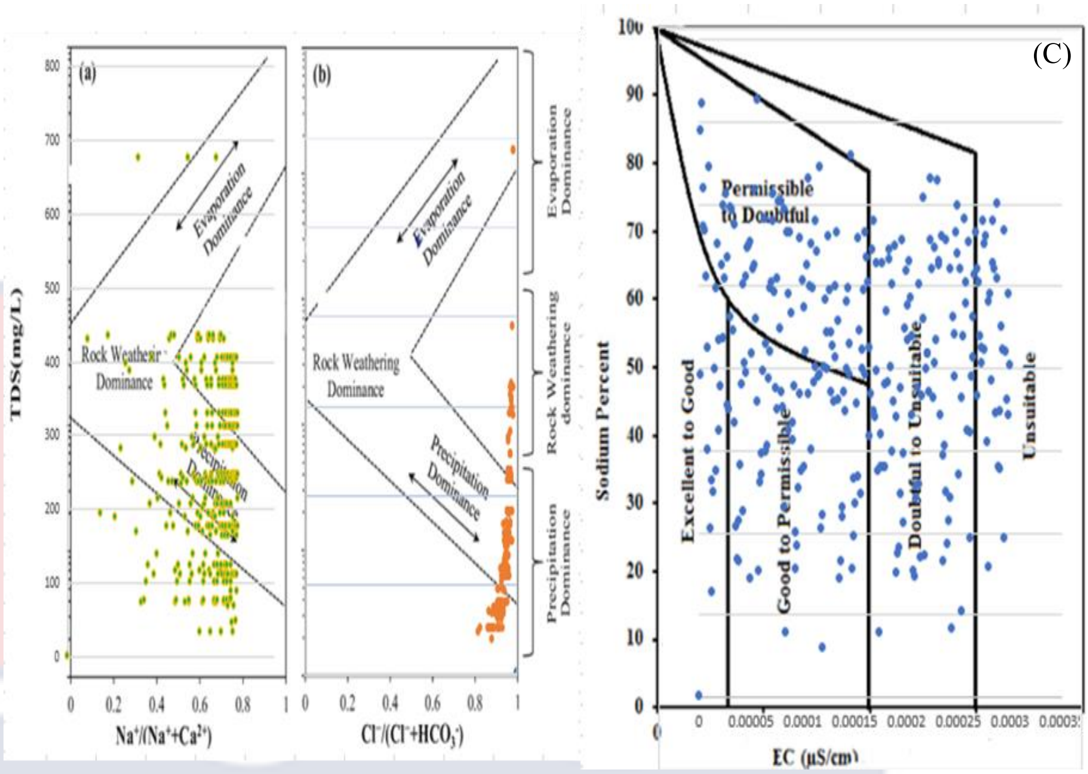

Figure 4.1: Groundwater Classification (Central, Greater Accra and Western Region)

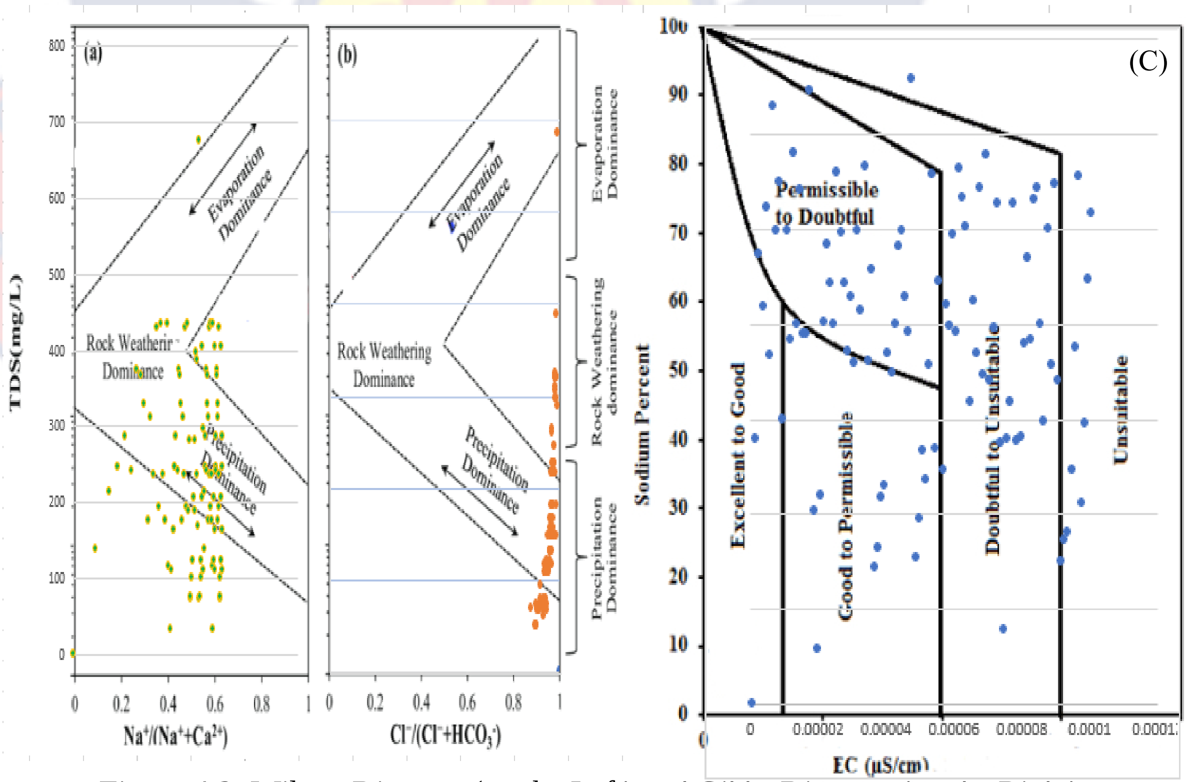

Figure 4.2: Wilcox Diagram (on the Left) and Gibbs Diagram (on the Right) Showing Groundwater Classification for Central Region

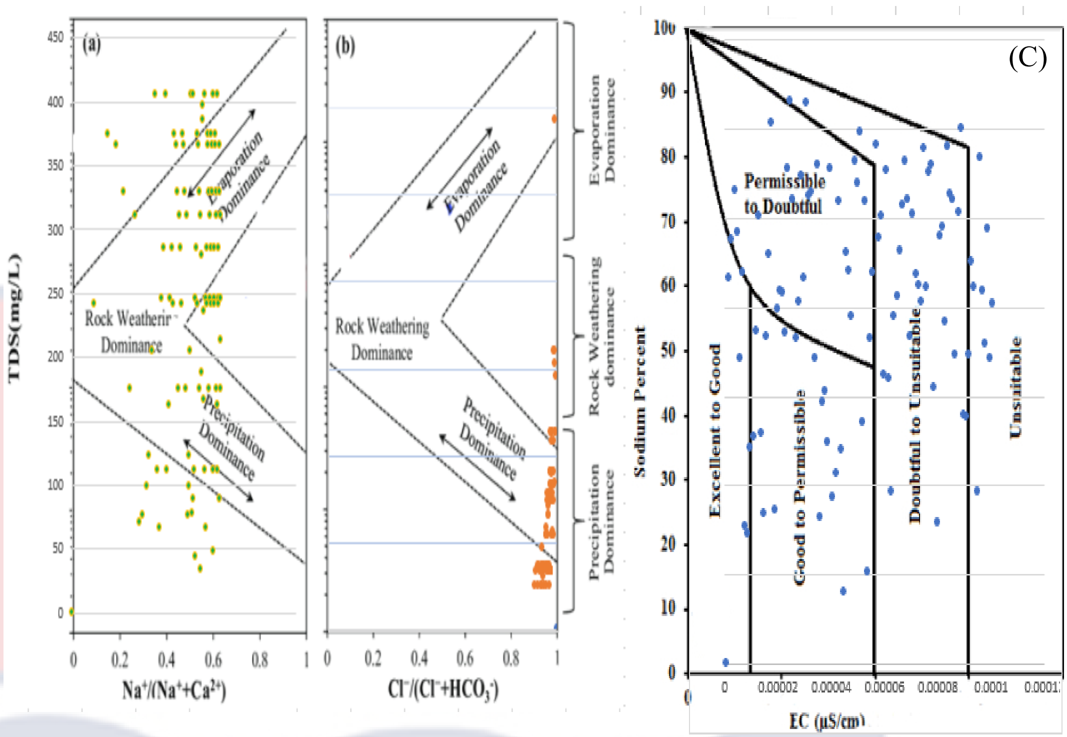

Figure 4.3: Wilcox Diagram (on the Left) and Gibbs Diagram (on the Right) Showing Groundwater Classification for Greater Accra Regions

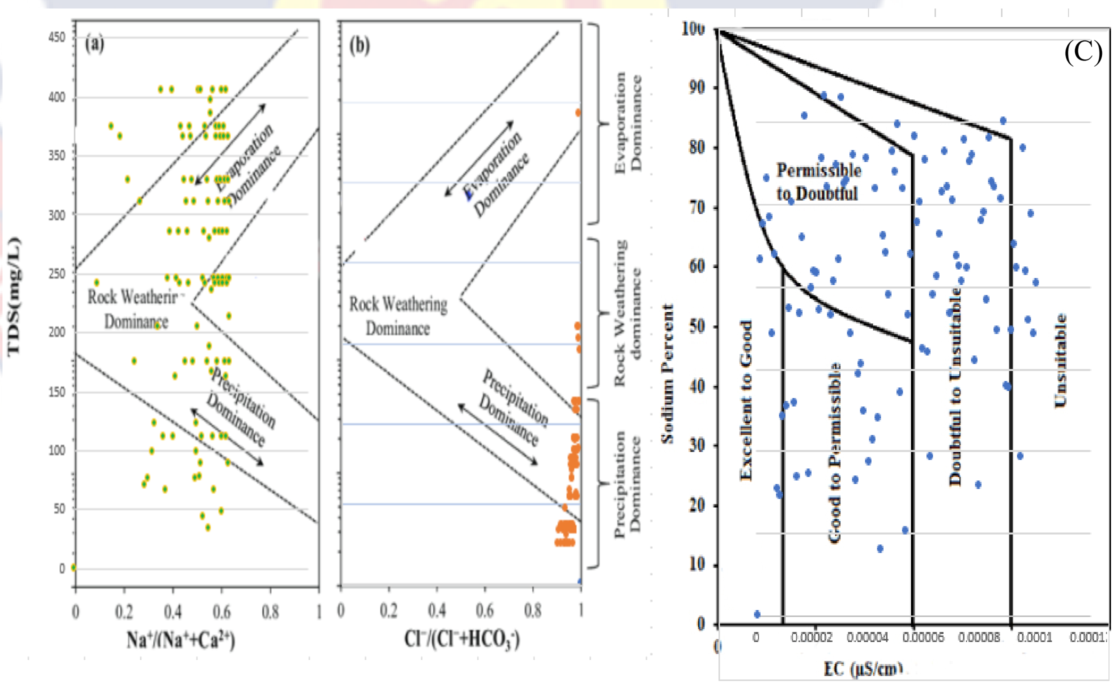

Figure 4.4: Wilcox Diagram (on the Left) and Gibbs Diagram (on the Right) Showing Groundwater Classification for Western Region

Figures 4.1, 4.2, 4.3 and 4.4 present Wilcox and Gibbs diagrams and show the distribution and classification of groundwater parameters. Figures 4.1(a) and 4.1(b) show that most of the groundwater parameters sampled from Central, Greater and Western Regions concentrated under precipitation and rock weathering dominance regions, with few under evaporation dominance. From Figure 4.1(c), few of the parameters concentrated in the region of excellent to good, with most of them under permissible to doubtful, doubtful to unsuitable, good to permissible and unsuitable regions.

As shown in Figures 4.2(a) and 4.2(b), most of the groundwater sampled from Central Region concentrated under precipitation and rock weathering dominance, with few under evaporation dominance. From Figure 4.2(c), only a few of the groundwater parameters analysed concentrated under the Excellent to Good region, with many of them concentrating under permissible to doubtful, doubtful to unsuitable, good to permissible and unsuitable regions.

From Figures 4.3(a) and 4.3(b), only one of the wells sampled from Greater Accra was grouped under the evaporation dominance region, while many were under rock weathering and precipitation dominance. From Figure 4.3(c), only a few of the Wells sampled from the Greater Accra Region were found under Excellent to Good and Good to Permissible regions. Many of the wells sampled classified under Doubtful to Unsuitable and unsuitable regions. From Figures 4.4(a) and 4.4(b), the wells sampled from the Western Region concentrated under precipitation, rock weathering and evaporation dominance regions. From Figure 4.4(c), only a few of the wells sampled from the Western

Region grouped under the Excellent to Good region, with many under Good to Permissible, Unsuitable and Permissible to Doubtful.

Table 4.2: Comparison of Groundwater Parameters among Central, Greater Accra and Western Regions (N = 300)

	Central Region			Greater Accra			Western		
Parameters	Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DO	7.16	1.84	11.20	7.10	3.39	9.90	7.34	3.88	11.76
pH	6.88	4.80	12.40	6.61	3.40	12.00	6.97	4.80	10.00
Temperature	27.34	25.60	33.60	28.51	25.90	33.60	26.36	23.60	32.60
TH	156.28	13.85	372.00	192.73	16.69	300.00	129.75	12.85	292.00
TDS	452.23	376.47	555.67	456.95	391.67	553.95	432.14	396.83	553.95
Turbidity	7.09	2.84	13.64	8.16	2.43	11.64	9.08	2.89	14.87
EC	432.31	269.45	669.73	418.26	245.86	672.56	428.96	277.38	629.28
Na	232.29	35.00	675.00	244.64	23.00	675.00	249.40	43.00	397.00
K	118.21	21.00	227.00	126.98	60.00	206.00	129.56	36.00	239.00
Ca	160.05	42.00	284.00	185.76	93.00	286.00	169.46	52.00	421.00
Mg	51.72	11.00	132.00	55.06	4.00	119.00	50.02	3.00	127.00
SO_4	75.09	116.00	161.00	67.25	46.40	131.00	98.11	54.00	147.00
PO_4	3.10	0.56	6.30	4.81	3.00	8.32.00	5.11	1.00	8.20
Cl	141.97	39.00	319.00	162.66	50.00	586.00	138.52	47.00	586.00
HCO_3	10.74	3.50	18.00	11.94	4.00	22.00	10.04	3.00	16.00
SiO_3	47.59	8.00	53.00	35.78	18.00	83.00	50.710	12.00	103.00
NO_3	21.96	7.00	42.00	21.94	5.00	44.00	24.60	6.00	76.00

Table 4.2 compares the physicochemical parameters of the groundwater sampled from the three regions. As presented in Table 4.2, the value of dissolved oxygen sampled from Western Region (7.34 mg/l) was the highest compared to Central (7.16 mg/l) and Greater Accra (7.10 mg/l) Regions. The mean pH value (6.69) of the groundwater sampled from the Greater Accra Regions was the lowest compared to Central (6.88) and Western (6.97) Regions. The mean temperature value of Western Region was least (26.36 ^{o}C) compared to Central (27.34 ^{o}C) and Greater Accra (28.51 ^{o}C) Regions. Comparing the total hardness of the groundwater sampled of the three regions, Greater Accra Region recorded the highest at 192.73.10 mg/l, followed by Central (156.28 mg/l) and Western (129.75 mg/l) Regions.

Among the three regions, Greater Accra Region recorded the highest total dissolved substances (456.95 mg/l), followed by Central and Western Regions, connoting that the groundwater is unsafe and not good for drinking. Western Region recorded the highest turbidity at 9.08 NTU. Central Region, on the other hand, recorded higher turbidity concentration compared to Greater Accra Region at 8.16 NTU and 7.09 NTU, respectively. Central Region recorded a mean value of electrical conductivity (432.31 mS/cm) which was the highest compared to Western Region (418.26 mS/cm), making Greater Accra the least electrical conductivity recording region.

Sodium ion concentration (249.40mg/l) sampled from Western Region was the highest, followed by Greater Accra (244.64 mg/l) and Central (232.29 mg/l) regions. The mean range of potassium ions recorded for Greater Accra Region was the highest compared to Western (129.56 mg/l) and Central (118.21 mg/l) regions. Regarding mean calcium ion concentration, Greater

Accra Region (185.70 mg/l) featured the highest, followed by Western (168.46 mg/l) and Central (160.05 mg/l) Regions. The mean magnesium (50.02 mg/l) of Western Region was the lowest compared to Central (51.72 mg/l) and Greater Accra (55.06 mg/l) regions. Among the sulphate ion concentrations of the groundwater sampled from the three regions, Western Region recorded the highest at 98.11 mg/l, followed by Central (75.24 mg/l) and Greater Accra (67.25 mg/l) Regions. The mean phosphate ion concentration sampled from the Western Region (5.12 mg/l) was the highest, followed by Greater Accra (4.81 mg/l) and Central (3.10 mg/l) Regions.

The mean chlorine ionic concentration of the groundwater sampled from Greater Accra Region (162.66 mg/l) was the highest compared to Central (141.97 mg/l) and Western (138.52 mg/l). Among the mean of the hydrogen bicarbonate ion concentration of the three regions, Greater Accra Region (11.94 mg/l) featured the highest, followed by Central (10.74 mg/l) and Western (10.04 mg/l) region. The mean silicate ion concentration of Western Region (50.71 mg/l) was the highest compared to Central (47.59 mg/l) and Greater Accra (35.78 mg/l) regions. Regarding nitrate ion concentration of the groundwater sampled from the three regions, Western Region recorded the highest at 24.60 mg/l, followed by Central (21.96 mg/l) and Greater Accra (21.94 mg/l) regions.

Table 4.3: Analysis of Variance (ANOVA) Comparing the Means of the Groundwater Parameters of the Three Regions

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F-crit*
Between regions	93456612	16	5841038	2291.16	0.00	1.65
Within regions	12958478	5083	2549.38			
Total	1.06E+08	5099				

Where SS = sum of squares, Df = degree of freedom, MS = mean square, F = F-Statistic, P-value = probability value and F-critical = Critical F-value.

Table 4.3 shows the output of one-way ANOVA test. It shows whether there was a significant change between and among the means of the groundwater parameters sampled from the three regions. The sum of squares (SS), the degree of freedom (Df), mean squares (MS) and the p-values were 93456612.00, 16.00, 5841038.00 and 0.00, respectively.

Table 4.4: KMO and Bartlett's Test

KMO Measure of Sampling Adequacy	KMO	0.69
Bartlett's Test of Sphericity	Approx. Chi-Square	543.43
	Degree of freedom (df)	136.00
	Significant level	0.00

Where KMO = Kaiser-Meyer-Olkin

Table 4.4 shows the output of the KMO and Bartlett's correlation tests from PCA. The KMO value obtained was 0.69, greater than the minimum value (0.50) required for Measuring Sampling Adequacy (MSA). There was a significance level of 0.00 ($p < 0.00$) for the Bartlett test. The value of Bartlett's Test of Sphericity was below the significance level (0.05).

Table 4.5: Variation Explained by the Components

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.92	11.28	11.28	1.92	11.28	11.28	1.88	11.06	11.062
2	1.45	8.50	19.79	1.45	8.50	19.79	1.37	8.07	19.134
3	1.32	7.74	27.53	1.32	7.74	27.53	1.31	7.71	26.845
4	1.30	7.65	35.18	1.30	7.65	35.18	1.27	7.49	34.339
5	1.28	7.50	42.68	1.28	7.50	42.68	1.25	7.33	41.672
6	1.13	6.65	49.32	1.13	6.65	49.32	1.15	6.78	48.456
7	1.05	6.17	55.49	1.05	6.17	55.49	1.15	6.75	55.204
8	1.01	5.96	61.46	1.01	5.96	61.46	1.06	6.25	61.456
9	0.98	5.78	67.24						
10	0.89	5.22	72.46						
11	0.86	5.07	77.52						
12	0.85	4.97	82.50						
13	0.81	4.77	87.27						
14	0.76	4.45	91.72						
15	0.68	3.97	95.69						
16	0.60	3.50	99.19						
17	0.14	0.81	100.00						

Extraction Method: Principal component analysis.

Table 4.5 displays the extracted eigenvalues, initial eigenvalues and the rotation sums of square loadings of the eight components. Based on the screen plot of components' eigenvalues, eight components were identified. Components with eigenvalues < 1 were noted to be statistically insignificant and thus set to be removed from the analysis. As displayed in Table 6, the first eight principal components explained nearly 61.50 percent of the variation in the groundwater quality with variable loadings spread over the PCs.

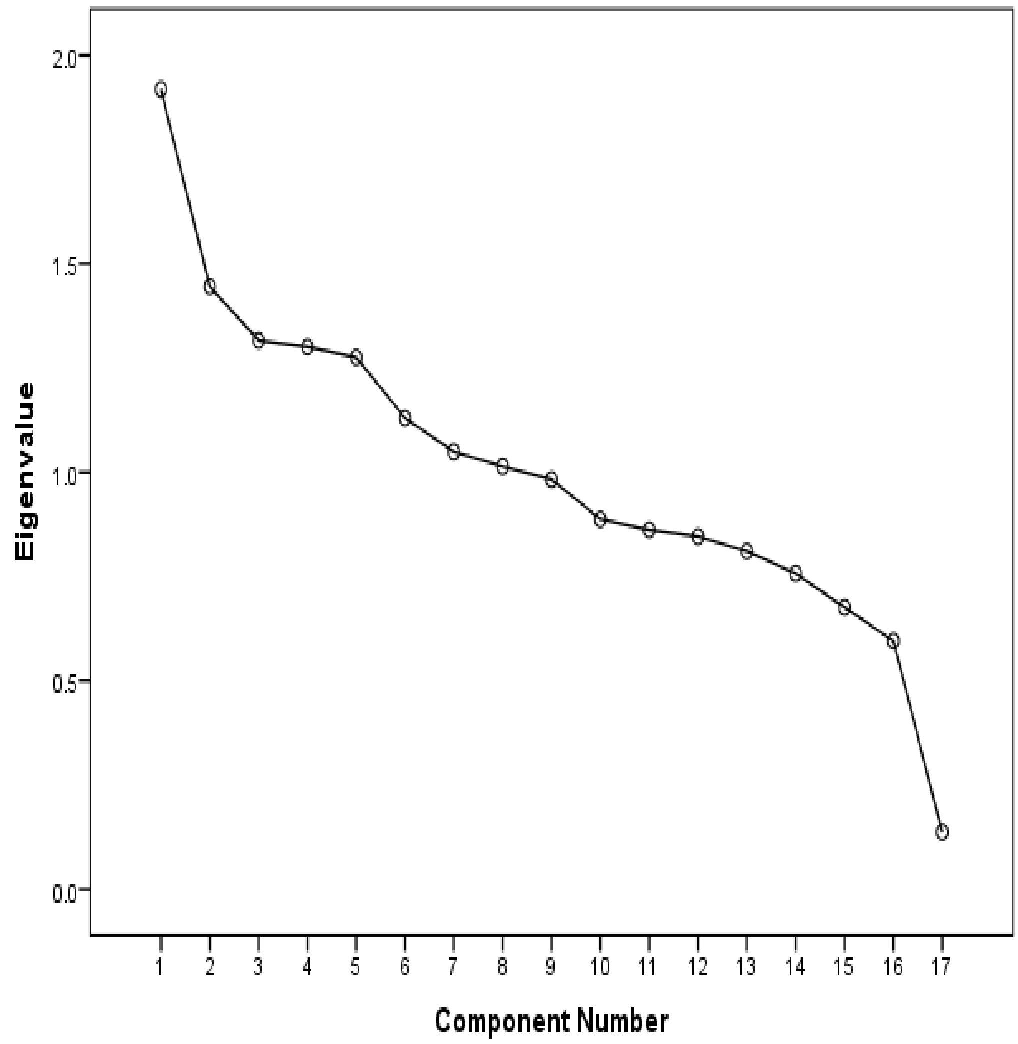

Figure 4.5: Scree plot of Eigenvalues and Number of Principal Components.

Table 4.6: Component Matrix of Groundwater Quality Parameters

	Component							
	1	2	3	4	5	6	7	8
Temperature	0.96							
Turbidity	0.96							
Calcium		0.75						
Potassium		0.74						
Sodium								
Chloride			0.79					
DO			0.53					
Sulphate			0.51					
Bicarbonate				0.73				
Silicate				0.57				
pH								
TH					-0.79			
Nitrate								
Phosphate						0.67		
TDS								
EC							0.81	
Magnesium								0.86

Rotation converged in 8 iterations.

Figures 4.6 and Table 4.6 show the loading plot and component matrix of groundwater quality parameters. Eight parameters influenced groundwater quality in the study regions, that is, temperature, turbidity, calcium, potassium, chloride, dissolved oxygen, sulphate and bicarbonate. The rotated principal component matrix displays the factor loadings between the principal components and the observed groundwater parameters.

From Table 4.6, the loading factors between each principal component and the observed groundwater parameters were greater than 0.50. The first principal component (PC1) showed a strong positive loading on temperature

(0.96) and turbidity (0.96), explaining 11.28 percent of the variance in the groundwater data. The second principal component showed a strong positive loading on Calcium (0.75) and potassium (0.74), accounting for 8.50 percent of the variance in the groundwater data. The principal component three (3) exhibited a moderate positive loading on chloride (0.79), dissolved oxygen (0.53) and sulphate (0.51), explaining 7.74 percent of the variation in the groundwater data observed.

There was a strong positive loading from component four (4) on bicarbonate and silicate at 0.730 and 0.565, respectively, explaining 7.65 percent of the variation in the groundwater data. The principal component five had strong negative loading on total hardness (-0.79), explaining 7.50 of the total variation in the groundwater data. Principal component six loaded positively on phosphate at 0.67. The factor loading between PC 6 explained 6.65 percent of the total variation in the groundwater data. The principal components seven and eight loaded strongly positively on electrical conductivity and magnesium at 0.810 and 0.86, explaining the variation in groundwater data observed at 6.17 percent and 5.96 percent, respectively. Nevertheless, no load between the principal components and observed groundwater parameters indicates that they did not load onto the groundwater data observed.

Discussion on Field Data Analysis

Regional variation of groundwater quality parameters

This section discusses the field data analysis on the impact of the physicochemical parameters assessed on groundwater quality in the study regions. The physicochemical parameters of the groundwater included in the study were pH, dissolved oxygen, temperature, total hardness, total dissolved substances, salinity and electrical conductivity and sodium, potassium, calcium, nitrate, hydrogen bicarbonate, silicate, sulphate, phosphate and chloride ions. These parameters provide a profound understanding of the nature, quality and pattern of changes in the groundwater of the selected regions.

As presented in Table 4.1, the mean dissolved oxygen (DO) at 7.20 mg/l fell within the WHO recommended DO value (> 6.5 mg/l) for excellent safe drinking water. The maximum value (11.76 mg/l) generally indicates that the underground water is good for drinking. Likewise, the high value of 11.76 mg/l of the DO sampled means that some of the groundwater sampled meets the WHO limits and is safe for drinking. The high dissolved oxygen level (11.76 mg/l) might be due to the significant recharge from oxygen-rich surface water or aeration due to the movement of the groundwater through porous rocks or sediments containing air pockets (Jeong et al., 2018). The low oxygen content (1.84 mg/l) shows that the groundwater in the regions is not safe for drinking and might be associated with health-related issues and thus, does not meet the WHO standard values and is unsafe for drinking. This might be due to high underground decomposition of organic matter from improper disposal of domestic, industrial and agricultural waste and high temperatures (through climate change) (Hutchins et al., 2020; Piatka et al., 2021).

The minimum pH value of 5 suggests the presence of anions such as sulphate, nitrate, phosphate and hydrogen bicarbonate in the groundwater, demonstrating an increased level of the acidic content of some of the groundwater sampled. The pH value ranged from 5.00 to 12.40, reflecting that the groundwater sampled to some extent is acidic (low pH value, 5.00) and alkaline (high pH value, 12.40). This means that some of the groundwater sampled is extremely acidic for drinking and associated with health problems like ‘acidosis’. According to Sanganyado and Gwenzi (2019), drinking acidic water increases disease outbreaks and influences antimicrobial effects that affect the healthy life of people. The high pH value of 12.40 means that some of the groundwater sampled from the three regions is alkaline and the groundwater in the regions is, therefore, hard water. High pH values indicate that the underlying rock of the groundwater is mostly made of limestone. According to Oludare (2017), limestone raises pH of water due to the presence of calcium carbonate. It also means that some of the groundwater are not safe for human consumption. Hence, they might have health impacts on the human body even though high alkaline water seems to be good for the human system (Sanganyado et al., 2019). This is because it has pro-ageing, colon-cleansing, immune system support hydration, skin health, and other detoxifying properties and cancer resistance properties (Yao et al., 2022). Chycki et al. (2017) argued that much more alkaline water can lead to weight loss and hydration. Low pH value in the regions might be due to poor agricultural practices such as high use of fertiliser, and agronomic practices such as tillage, continuous cultivation, mining, and industrial activities (Patil et al., 2012). The low and high pH of the regions calls for urgent measures to shield and control activities that lead to both slow and rapid increases in the acidic and alkaline content of the groundwater in the study regions.

The mean (28.40 ^{o}C) temperature (^{o}C) value of the groundwater sampled from the three regions falls within the WHO limits 23 ^{o}C to 33 ^{o}C.

This mean temperature of the study regions might be due to the measurements being done in situ and structures or differences in drilling depths. The mean temperature range (25.00 ^{o}C to 33.10 ^{o}C) indicates that some of the groundwater sampled falls within the WHO standard values. The temperature of the groundwater sampled that exceeded the WHO recommended range might be due to global climate change, the geographic location of the area sampled, the high rate of deforestation that exposes the land directly to the sun's rays and excessive disposal of industrial, agricultural and domestic human waste in the regions. Management and agencies responsible for water management and conservation should develop strategies such as afforestation programmes and other volunteering groups to see through the sustainability and protection of groundwater from excessive sunlight.

Lowland regions should see it as a need and control the irregular disposition of liquid waste. Singh and Gupta (2016) asserted that sewage discharges from households and industries seep down, leading to underground water temperature alteration and low water quality. Accordingly, water temperature affects electrical conductivity, solubility, ionic strength, corrosion and dissolution (Alvarez-Bastida et al., 2018). Total hardness measures all multivalent cations concentrations in water (Wang et al., 2022), and thus, the mean total hardness (TH) of 159.58 mg/l exceeded the WHO standard value for excellent drinking water quality (< 60 mg/l). The mean total harness of the underground water sampled from the three regions connotes that the groundwater is moderately hard with salt content. Likewise, the range of the total from 13.85 mg/l to 372.00 mg/l confirms that some of the groundwater sampled are soft (excellent for drinking), moderately soft (good for drinking, but not excellent), moderately hard and very hard (not good for drinking).

The high total hardness might be due to the influence of the sea and high calcium, iron and magnesium concentrations, limestone or discharges from operating or abandoned mines (Shah et al., 2022). Rakhimova (2022) believes that cations such as magnesium carbonate from dolomite and calcium carbonate from limestone in the soil increase the hardness of the water. The mean total dissolved substance (TDS) of 453.77 mg/l of the three regions exceeded the WHO recommended values for excellent drinking water (< 300 mg/l, excellent). The total dissolved substances of the groundwater sampled from the three regions compared to the WHO standard guidelines indicate that the water quality is poor (> 300 mg/l, poor and not good for drinking). Also, it indicates the gradual accumulation of dissolved ions, especially sulphates.

The range of the total dissolved substances (376.47 mg/l to 555.67 mg/l) indicates that some groundwater is safe for drinking. The value of the total dissolved substances of the groundwater sampled in the regions below 300 mg/l demonstrates that some of the sampled groundwater is excellent for drinking, whereas that above 300 mg/l is not safe for drinking. The high value of the total dissolved substances in the regions might be due to the infiltration of pollutants from landfill leachate, feedlots, or sewage and increased human anthropogenic in the three regions. Total dissolved solids as groundwater pollutants are dangerous for human health, reduce water quality, affect the taste of the water and make water quality unsafe for human use. According to Soleimani et al. (2022), inorganic salts like calcium, chlorides, magnesium, bicarbonates, potassium, sulphates and sodium influence TDS in water. The high presence of total dissolved substances in the groundwater sampled might result from urban run-off, sewage disposal, industrial wastewater, mineral springs, seawater intrusions, carbonate and salt deposits, drinking water treatment chemicals, stormwater and agricultural runoff.

The mean electrical conductivity (426.51 mS/cm) exceeded the WHO recommended value (400 mS/cm). The electrical conductivity of the three regions ranged from 245.86 mS/cm to 695.37 mS/cm, indicating that some of the groundwater sampled from the regions met the WHO limits. The low EC demonstrates that the groundwater is safe for drinking. The low EC concentration in the groundwater sampled might be due to factors such as low levels of dissolved solids, geological composition, distance from pollution sources, rainfall and dilution, and temperature variations. The high EC value indicates the existence of a significant level of dissolved elements, like metal salts and organic material, in the groundwater samples collected from the three different regions (Haritash et al., 2016). In addition, it indicates the influence of seawater and the presence of cations and anions such as calcium, magnesium and chloride ions. This imprints that the groundwater is unfit for human use. It is suggested that the Government of Ghana and its agencies should encourage the citizens in the regions and other regions to adopt measures that contribute to preserving the groundwater quality.

The mean turbidity (8.11 NTU) exceeded the WHO recommended values (< 5 NTU) for excellent drinking water. The mean turbidity compared to the WHO limits indicates that the groundwater sampled is very polluted or not very clear. The turbidity range of the groundwater (2.43 NTU to 13.87 NTU) implies that the groundwater sampled meets the WHO standard values for excellent drinking water. The high turbidity might be due to high temperatures (from climate change), erosion and disposal of wastewater into the underground water (Jain & Singh, 2020; Mishra, 2023). According to Sheikh et al. (2022), turbidity is a concern because it can influence economic income and agricultural production. Consequently, huge money is required to treat the saline water, and crop yield becomes less since plants cannot utilise the soil water to prepare their food, leading to hunger, poverty, and species extinction. The high turbidity value estimated is a concern for individuals, government, and other organisations to tackle. Excessive turbidity increases the level of heavy metals (lead, mercury, and cadmium) in groundwater and the water supply system, affecting human health and other health risks like gastrointestinal diseases (Sonone et al., 2020).

The high mean sodium (Na) ion concentration of the groundwater sampled compared to the WHO recommended value might be due to the influence of the seawater, mineral composition of the groundwater and road salts, sewage, landfills and industrial, agricultural and domestic waste discharges. The sodium (Na) ion concentration range demonstrates that some of the groundwater sampled from the three regions is safe for drinking. The low sodium ion concentration of some of the groundwater sampled indicates health issues such as cardiovascular and neurological issues associated with the wells when people use them. Sorensen et al. (2020) noted that sodium in water helps maintain cardiovascular muscle movement in the body and control diseases such as neurological disorders. Pohl et al. (2013) view that sodium is good for preserving volume and blood pressure regulation, maintaining muscle contraction and transmission of nerve cells and helping balance water, acids and bases in the body, but high sodium consumption would trigger hypertension among humans. This revitalises the need for individual organisations to protect the health of the groundwater in the presence of “illegal mining” and other industrial, household and farming activities that add up to the instability of the sodium content in the groundwater.

The mean potassium ion concentration (124.92 mg/l) of the groundwater from the regions fell below the WHO standard value (300.00 mg/l). This low mean value compared to the WHO standards might be due to the influence of the mineral composition of the regions, other agricultural and industrial activities in the regions, lack of moisture (by climate change) and the high cation exchange capacity of the soil. The range of the potassium ion concentration of the groundwater sampled from the regions indicates that some of the wells sampled met the WHO standard value. The low potassium ion concentration might be due to low pH and oxygen dissolution, a large excess of cations in the soil and a lack of soil moisture in the ground. Low pH, oxygen content, and moisture of groundwater prevent the leaching of potassium into the underground water (Lawniczak et al., 2016; Khan et al., 2018; Albert, 2015). This is a problem because potassium availability in groundwater indicates its availability to plants and the proper health growth of humans and plants (Ayilara et al., 2020). However, it is established that crop yields would reduce if care is not taken. This is because plants absorb their nutrients from the soil in a liquid form.

The mean calcium ion concentration of the groundwater sampled surpassed the WHO accepted value (100.00 mg/l, excellent and safe for drinking), indicating that the groundwater sampled is hard water and thus, not safe for drinking. The high mean calcium ion concentration might be due to the presence of the influence of the ocean water, the mineral composition of the underlying groundwater, and introduction of industrial wastes, and the high-temperature range of the underground water (33.10 mg/l) through climate change, domestic and agricultural solid such as excessive fertiliser application in the environment. High calcium concentration denotes that the groundwater sampled has high electrical conductivity, dissolved substances and salt, resulting in the hardness of the groundwater sampled. The range of calcium ion concentration compared to the WHO standards presages that some groundwater is soft and hard water. Groundwater with a low calcium ion concentration of 42.00 mg/l might be due to the distance of the wells from the seawater. The mineral composition of the underlying rocks has a low concentration of calcium-related compounds such as $CaSO_4$ and a high leaching rate of the soil.

According to Cheng et al. (2022), there is a high rate of calcium ion leaching from rocks and soil of groundwater and underground aquifers even though calcium carbonate is fairly water-insoluble. Calcium carbonate dissolves more easily in water that has a high quantity of dissolved carbon dioxide. Though high calcium ion concentration of groundwater is good for strong bones, body metabolism and other physiological processes, such as cardiovascular contractions, blood coagulation, muscular shrinkage, neuron transmissions and healthy growth of plants, high calcium ion concentration leads to scaling of properties of water and binds with phosphorus, reducing nutrients uptake of plants (Reigl et al., 2022). Calcium deficiency is linked to osteoporosis, nephrolithiasis (kidney stones), colorectal cancer, hypertension and strokes, coronary heart disease, insulin resistance, and overweight. The majority of these diseases have treatments but no cure. However, individuals, governments and agencies should holistically come together to combat activities that contribute to the alteration of calcium contents in groundwater.

The mean magnesium ion concentration (52.27 mg/l) compared to the WHO standard value (50.00 mg/l) is high, indicating that the groundwater sampled is of poor quality for human use. The minimum and maximum magnesium ion concentrations of the groundwater imply that the groundwater is not good for consumption, hence, soft water. The low magnesium content of the groundwater sampled might be due to the acidity nature of the soil, soil cation exchange and mineral composition of the soil. The wells with low magnesium ion concentrations demonstrate that the soil's underlying rocks are free of minerals such as limestone and gypsum (Li et al., 2013). In contrast, the high magnesium ion concentration of the groundwater signifies that the groundwater sampled is hard water and contains high dissolved substances. The increase in the magnesium concentration might be due to the infiltration of high surface water with a high concentration of salt contents into the underground water.

The mean sulphate ion concentration of 92.14 mg/l was lower than the WHO recommended values for excellent drinking water (< 250.00 mg/l, excellent). The mean range (60.00 mg/l to 261.00 mg/l) of sulphate ion concentration sampled from the regions shows that some of the wells meet the WHO recommended values and thus are safe for drinking. The high sulphate ion concentration in some of the wells might be due to cations such as calcium sulphate and magnesium sulphate ions in the wells sampled.

The high sulphate ion concentration of the groundwater sampled from each of the three regions ascertains the acidity of the wells. The groundwater is not safe for human consumption. According to Omer (2019), high sulphate concentration in water makes the water corrosive and to have laxative effects on people and gives water a bitter taste. The mean phosphate ion concentration (3.97 mg/l) was greater than the WHO recommended value (2.00 mg/l). The high phosphate ion concentration in the groundwater sampled from each of the regions might be due to the phosphate erosion of rocks into underground water, chemical fertilisers, manure, and composted materials. Phosphate ions are useful for plant agricultural purposes and health, but Deutsches Arzteblatt International reports showed that excessive consumption of phosphate is harmful to health since it influences cardiovascular diseases in humans (Hahad et al., 2019). Agricultural sectors, organisations, and government agencies should help control the irregular usage of phosphate products in the environment.

The mean range of phosphate ion concentration in some of the groundwater sampled shows that groundwater has low phosphate ion concentration. This might be due to the mineral composition of the underlying soil, low fertiliser usage and domestic, industrial and agricultural production of products that contain phosphate ions. Low phosphate ion concentration in the groundwater sampled might be due to cations such as calcium and irons, which strongly adsorb phosphate ions onto the soil. Adsorption and mineralisation limit phosphate ion movement within or below the root zone. Moreover, soils with high calcium carbonate concentrations can impede phosphorus transport due to the formation of calcium phosphate minerals.

The mean chlorine ion concentration sampled from the groundwater exceeded the WHO recommended value (250.00 mg/l). The high mean values might be due to the mineral composition of the underlying rocks and intrusion of cations such as sodium and magnesium as well as seawater. Chloride ion concentration in seawater is about 19,000 milligrams per litre (mg/l). Chloride concentrations increase dramatically when groundwater becomes contaminated with seawater (Alfarrah & Walraevens, 2018). The low chloride ion concentration of some of the groundwater might be due to the distance of some of the wells away from the seawater, saltwater intrusion, mineral dissolution, industrial and domestic waste and the underlying mineral composition of the area enough cation such as calcium and magnesium.

The mean hydrogen bicarbonate ion concentration estimated was greater than the WHO recommended values. The high hydrogen bicarbonate ion concentration of the sampled regions compared to the WHO recommended value (10.00mg/l) shows that the groundwater is unsafe for human consumption. This might be a result of the high dissolution of carbon dioxide gas and domestic and industrial waste disposal in the environment. The mean range of hydrogen bicarbonate ion concentration compared to the WHO standard values shows variable concentration. The high hydrogen bicarbonate ion concentration might be due to the availability of cations like calcium and magnesium and the mineral composition of the soil. The high hydrogen bicarbonate ion concentration might have contributed to the hardness of the groundwater in the regions.

The mean silicate ion concentration was below the WHO recommended value (100.00 mg/l). The low silicate ion concentration compared to the WHO standard value might be due to the mineral composition of the soil and domestic and industrial waste disposal that inhibits the existence of silicate ions in the underground water. The low silicate ion concentration connotes the need to protect the groundwater in the regions. Silicates are classified as corrosion inhibitors and have protective films on metal surfaces. This makes them harmless to humans; therefore, their presence in the soil needs attention.

The mean range of silicate ion concentration indicates that some of the wells sampled contain high silicate ion concentrations. This might be due to the mineral composition of the soil and the high temperature of the wells. According to Jollivet et al. (2012), deep groundwater aquifer has more silicate ion concentration due to the influence of the high temperature. The high concentration of silicate ions might be due to the mineral compositions of the underlying rocks and the rainfall pattern of the area that instigates the dissolution of silicate ions in the groundwater. The high silicate ion concentration might also be due to the presence of weathering rocks (Dobrzyński, 2005).

Wells with low silicate ion concentration might be influenced by kaolinite rocks. Kaolinite rocks prevent silica solubility in water (Dobrzyński et al., 2005). Dobrzyski et al. (2005) state that the water content deficiency of the aeration region, annual variations of temperature and precipitation bedrock hypersensitivity, and chemical changes stability all affect the amount of silica discharge into the groundwater. The mean nitrate ion concentration exceeded the WHO recommended value (50.00 mg/l) for safe drinking water. The high nitrate ion concentration recorded from the wells sampled might be due to the atmospheric deposition, mineral dissolution and other anthropogenic sources (mining, fertiliser, etc.) and the presence of gypsum in the groundwater aquifer of the regions sampled (Sharma et al., 2020). The low range of these anions imprints that the groundwater is becoming acidic and might be due to uncontrolled fertiliser, industrial waste discharge, domestic waste production, feedlots and sewage. The high nitrate ion concentration of the groundwater in the region calls on government and non-governmental organisations to protect and conserve the groundwater in the region. The availability of clean drinking water and hygiene is essential for meeting the Millennium Development Goals (MDGs). The presence of some of these pollutants in the groundwater spells the impending doom of water scarcity in the regions and beyond. However, protecting and conserving the groundwater quality would help fight against climate change and provide enough water to meet the population growth and industrialisation water needs in the regions.

Classification of groundwater under environmental parameters causing alteration in groundwater quality (based on physicochemical parameters of each region)

The study assessed the factors that influence the concentration of ions in the groundwater. Information theory assumes that groundwater quantity and quality are recurrent movements and consist of complete systems. To understand the processes that influence groundwater quality, the Gibbs diagram has three distinct areas (precipitation, rock and evaporation dominance), while the Wilcox diagram with five distinct regions (Good to Permissible, Doubtful to Undoubtful, Permissible to Doubtful, Excellent to

Good and Doubtful) have been employed to demonstrate the source of ions and explain the general chemistry of the groundwater in the study areas.

Gibbs ratio $Na^+/(Na^++Ca^{2+})$ for cations and $Cl^-/(Cl^-+HCO_3^-)$ and Wilcox ration (sodium percent) for anions of water samples were plotted separately against the respective TDS. These show the type of dominance controlling the groundwater quality. Gibbs diagram shows the chemical interaction between rock-forming minerals of the aquifer and helps evaluate water quality. The following graphs help establish the factors and characteristics of groundwater parameters contributing to groundwater pollution.

Concerning Figures 4.1(a) and 4.1(b), most of the groundwater parameters sampled from the wells in the selected regions (Central, Greater and Western Regions) concentrated under precipitation and rock weathering dominance regions, while few of them fell under evaporation dominance region. Underground weathering rocks and precipitation influence most groundwater in the three regions (Central, Greater Accra and Western Regions), with little influence from the evaporation dominance variables. Few of the parameters of the groundwater from the three regions concentrated in the region of excellent to good, while most of them concentrated under Good to Permissibility, Permissible to Doubt, Doubtful to Unsuitable and Unsuitable regions. This means that the wells sampled were not safe for human consumption. Only a few met WHO recommended criteria for excellent drinking water. Only groundwater parameters in the Excellent to Good and Good to Permissible might meet the WHO recommended values. Nevertheless, those in the other regions, such as Permissible to Doubt, Doubt to Unsuitable and Unsuitable, do not meet the WHO standard value and, thus, are not safe for drinking.

Gibbs and Wilcox diagrams for Central and Greater Accra Regions indicate that most of the groundwater sampled from each region is precipitation and rock weathering, while few of them were under evaporation dominance. Only a few of the groundwater parameters analysed were concentrated under the Excellent to Good region, but many of them were concentrated under Good to Permissible, Permissible to Doubtful, and Doubtful to Unsuitable. Few of the parameters are concentrated under Unsuitable regions, and this is good.

The wells sampled from the Western Region concentrated under precipitation, evaporation and rock weathering regions. This means that groundwater in the Western Region is influenced by the mineral composition of the underlying rocks, precipitation and evaporation. Many of the groundwater parameters sampled from Western Region were concentrated under Doubtful to Unsuitable, unsuitable and Permissible to Doubtful, with few under Good to Excellent. This means that the groundwater in the regions is severely polluted or getting polluted. This might result from domestic actions and activities such as illegal mining, sometimes close to homes.

Comparing physical parameters in groundwater by region (each region)

From Table 4.2, the mean dissolved oxygen of the groundwater sampled from Central, Greater and Western fell within WHO standard values for safe drinking water. The high mean dissolved oxygen values of Western Region (7.43 mg/l) compared to Central and Greater Accra Regions might be due to low temperature (23.00 ^{o}C), and proper disposal of industrial and domestic waste. According to Luo and Zhou (2022), warm water has a low ability to retain a large amount of oxygen. The mean ion concentration of dissolved oxygen of Central, Greater Accra and Western Regions, respectively, suggests that the groundwater sampled from the region does not meet the WHO recommended limits (> 6.5 mg/l) for safe drinking water. The government and its agencies should put mechanisms into place to control activities such as deforestation and unplanned disturbance of the soil to keep the soil particles in shape for proper aeration to occur in the underground water in the regions. This is because high decomposition and infiltration of waste materials, as well as deforestation, contribute to the oxygen reduction in the groundwater.

The mean pH value (6.93) of the Western Region was the highest compared to Central (7.84) and Greater Accra (7.69). The high pH values of Central and Greater Accra Regions might be due to the sea intrusion and high cation such as calcium and magnesium concentration and mineral composition of the underlying rocks. The low mean pH value of the groundwater sampled from Central and Greater Accra Regions might be due to mining and agricultural activities such as fertiliser, pesticide and weedicides application and low cations like magnesium and calcium concentration.

The mean pH value of the groundwater sampled from the Western Region was the lowest, followed by Greater Accra and Central Regions. The high mean temperature value of the groundwater sampled in the Greater Accra Region might be due to the influence of climate change, the presence of pollutants in the water, seawater intrusion with high heat capacity, high emission of aerosols and regular deforestation for settlement exposing the land surface to the excessive heat of the sun. Thus, when the surface of the soil is heated, the heat is transferred by geothermal conversion and conduction. Also, the low mean temperature value of the Western Region compared to Central and Greater Accra Regions might be the low influence of the seawater on the groundwater, reducing the heat capacity of the groundwater (Riedel, 2019).

The high total hardness of the groundwater in Greater Accra Region (452.20 mg/l) compared to Central and Western Regions might be due to the influence of the seawater, mineral composition of the soil, and presence of cations such as sodium and magnesium. The high total hardness of each region might be due to the mineral composition of the regions, fertiliser application, and domestic and industrial introduction of wastewater with high concentrations of cations such as sodium and calcium. The total hardness concentration of the groundwater based on the regions studied follows as Greater Accra Region > Central Region > Western Region.

The high total dissolved substances concentration of the groundwater sampled from Greater Accra Region might be the influence of illegal mining, agricultural and industrial activities, as well as, the introduction of domestic solids and ligiids wastes and shallowness of some parts of the region (some of the wells were not deep) and influence of the seawater. Based on the means of the three regions, Greater Accra Region recorded the highest (mg/l), followed by Western, Central and Western regions. The high turbidity content of the groundwater sampled from Western Region might be due to the seawater intrusion, mineral composition of the soil and irregular disposal of domestic, agricultural and industrial wastes.

The high value of the electrical conductivity of the groundwater sampled from Greater Accra Region compared to Central and Western Regions might be due to the high salt content of the underground water, the influence of the seawater and the presence of cations and anions such as calcium, magnesium and chloride ions. It might also be due to the sea intrusion and presence of cations and anions, as well as, poor disposal of domestic, agricultural and industrial wastes into the environment. In sum, Central Region featured the highest electrical conductivity, followed by Western and Greater Accra Region.

The high sodium ion concentration of Western Region might be due to the influence of the seawater, industrial, agricultural and domestic waste discharges and the mineral composition of the groundwater. The low sodium ion concentration of some of the groundwater sampled indicates health issues associated with the groundwater in the areas. The high potassium ion concentration of the Western compared to Central and Greater Accra Regions might be due to the influence of the mineral composition of the underlying rocks. The low potassium ion concentration of Central and Greater Accra Regions might be due to low pH and oxygen dissolution, a large excess of cations in the soil, the lack of soil moisture and the high cation exchange capacity of the soil

The high calcium values of Greater Accra Region compared to Central and Western Regions might be due to the presence of seawater and mineral composition of the underlying groundwater. Groundwater with low calcium ion concentration is influenced by the distance of the wells from the seawater. It also means that the mineral composition of the underlying rocks has a low concentration of calcium-related compounds such as $CaSO_4$, and a high leaching rate of the soil. The low magnesium content of the groundwater sampled from Western Region might be due to the acidic nature of the soil, soil cation exchange and mineral composition of the soil.

The wells with low magnesium ion concentrations mean that the underlying rocks of the soil are free of minerals such as limestone and gypsum. The high magnesium ion concentration of Central and Greater Accra Regions implies that groundwater is not good for drinking. The high magnesium ion concentration of the groundwater signifies that the groundwater sampled is hard water and contains high dissolved substances, low pH, and high temperature which might be due to climate change and domestic, industrial and agricultural waste production in the environment. The increase in the magnesium cation of the groundwater might be due to the infiltration of high surface water with a high concentration of salt contents into the underground water.

The high sulphate ion concentration of Western Region compared to Central and Greater Accra Regions might be due to the existence of cations such as calcium sulphate ions in the wells sampled. The high sulphate ions concentration in some of the groundwater might be due to mineral dissolution, anthropogenic sources (mining, fertiliser, etc.), atmospheric deposition and the presence of gypsum in the aquifer. The high sulphate ion concentration ascertains the acidity of the wells and thus, the groundwater is not safe for human consumption. The low phosphate ion concentration in the groundwater sampled from the Central Region might be due to the presence of cations such as calcium and irons, which strongly adsorb phosphate ions onto the soil.

The high phosphate ion concentration in the groundwater sampled from each of the regions might be due to the phosphate erosion of rocks into underground water, chemical fertilisers, manure and composted materials. Phosphate ions are useful for plant agricultural purposes and health, but, according to Hahad et al. (2019), excessive consumption of phosphate is harmful to health since it influences cardiovascular diseases in humans.

The high chlorine ion concentration of Greater Accra Region might be due to the mineral composition of the underlying rocks, and the existence of positive ions such as sodium. Hence, chloride ion concentration in seawater is about 19,000 milligrams per litre (mg/l). Nevertheless, when groundwater becomes contaminated with seawater, chloride concentrations increase dramatically (Alfarrah et al., 2018). The mean range of chloride ion concentration sampled from the regions shows that some of the groundwater was of low chloride ion concentration. The low chloride ion concentration of some of the groundwater recorded for Central and Western Regions might be due to the distance of some of the wells away from the seawater, saltwater intrusion, mineral dissolution, industrial and domestic waste and that underlying mineral composition of the area enough cation such calcium and magnesium. Regarding chloride ion concentration, Greater Accra featured the highest followed by Central and Western Regions.

The high hydrogen bicarbonate ion concentration of the sampled from Western Region compared Central and Western Regions indicates that the groundwater is not safe for human consumption. The high hydrogen bicarbonate ion concentration in Western Region might be due to the high dissolution of carbon dioxide gas, and domestic and industrial waste disposal into the environment. furthermore, it might be due to the presence of cations such as calcium and magnesium and the mineral composition of the soil. The high concentration of hydrogen bicarbonate ion concentration might have contributed to the hardness of the groundwater in the regions. The low silicate ion concentration Greater Accra Region compared to Central and Western Regions might be due to the mineral composition of the soil and domestic and industrial waste disposal that inhibits the presence of silicate ions in the groundwater.

The high concentration of silicate ions in Western Region compared to Greater Accra and Central Regions might be due to the mineral compositions of the underlying rocks and rainfall pattern of the area that instigates the dissolution of silicate ions in the groundwater. Wells with low silicate ion concentration might be influenced by kaolinite rocks. The low silicate ion concentration connotes the need to protect the groundwater in the regions. According to Jollivet et al. (2012), deep groundwater aquifer has more silicate ion concentration and thus, it is due to the influence of the high temperature.

The high nitrate ion concentration recorded from the wells sampled from the Western Region compared to Central and Greater Accra Regions might be due to the atmospheric deposition, mineral dissolution and other anthropogenic sources (illegal mining, fertiliser, etc.) as well as the presence of gypsum in the groundwater aquifer of the regions sampled (Sharma et al., 2020). The low anions indicate that the groundwater has become acidic and might be due to the influence of uncontrolled fertiliser, industrial waste discharge and domestic waste production, as well as feedlots and sewage.

The fact that the underground water hinges on the soil conformation, the bedrock and the mineral compositions of the soil, anthropogenic activities are impacting the underground water and its quality. The Government of Ghana and its agencies should help enforce rules and regulations that protect underground water. According to Wang et al. (2022), underground water forms part of the natural cycle of the earth and freshwater and it needs to be protected and sustained. There would be disparities in access to better water quality and sanitation among rich and poor, men and women and rural and urban locations. People in Ghana might depend on unsustainable and unimproved water sources. This highlights the fact that more rapid and exigent actions are needed to improve and protect the groundwater in the regions. To resolve the disparities in access to water and sanitation, Ghanaians should embrace SDG 6, which calls for the provision and continuous management of water and sanitation for everyone, ensuring fair and universal use of safe and cheap water supply for everyone by 2030.

Determining the significant differences between and among the physicochemical parameters of the groundwater of the three regions

The study determined if there was a statistically significant difference between and among the groundwater parameters of the study regions. ANOVA analysis was run on the averages of the groundwater parameters. Comparing the alpha-value (0.05) to the p-values (0.00), there was a significant difference between and among the means of the groundwater parameters sampled from the study regions. Based on the test statistic, the null hypothesis is false and thus, would be rejected (0.00 < 0.05), indicating that differences among the means of the groundwater parameters are not due to chance.

The significant differences in groundwater parameters among the Central, Greater Accra, and Western Regions might be influenced by the geological formations and seasonal variations within the regions (Asomaning et al., 2023; Mensah et al., 2023). These factors contributed to spatial heterogeneity in the groundwater quality and influenced the groundwater parameters, including total hardness, electrical conductivity, ion concentrations, turbidity, and total dissolved solids, in the regions. For example, the mean dissolved oxygen (DO) in Central Region (7.16 mg/l) was lower than that of Greater Accra (7.1 mg/l) and Western (7.34 mg/l) Regions. pH levels were relatively consistent, with Central Region exhibiting a mean of 6.88, compared to 6.61 in Greater Accra Regions and 6.97 in Western Region. Temperature averages at 27.34 °C in Central Region, 28.51°C in Greater Accra and 26.36 °C in Western Regions, indicating slight differences in thermal characteristics. Total hardness (TH) was highest in Greater Accra Region (192.73 mg/l) compared to Central (156.28 mg/l) and Western (129.75 mg/l), suggesting varying mineral content. Total dissolved solids (TDS) and electrical conductivity (EC) levels were relatively similar across the regions. Greater Accra Region recorded the highest mean sodium (244.64 mg/l) and chloride (162.66 mg/l) ions compared to Western and Central at (249.4 mg/l and 138.52 mg/l) and (232.29 mg/l and 141.97 mg/l), respectively.

PCA estimating the loading effects of groundwater parameters

PCA indicates that the groundwater data gathered for the study is an eight-component system. The first eight principal components explained nearly 61.50 percent of the variation in the groundwater quality with variable loadings spread over the PCs. However, this is associated with pollutants in the groundwater and the interdependent association between the groundwater components. From the correlation matrix, the parameters were strongly positively correlated. The correlation matrices were > 0.5, implying that the PCA analysis could be conducted.

The KMO value obtained was 0.691, greater than the minimum value (0.50) required for Measuring Sampling Adequacy (MSA). This indicates that the dataset fits the principal component analysis and implications. The Bartlett test has a likelihood value of < 0.001, implying the rejection of the null hypothesis. Nonetheless, when the Bartlett test correlation matrix is an identity matrix, the dimension cannot be reduced. The significance level of Bartlett's test ($p<0.001$) affirms there are statistically significant associations among the groundwater parameters.

The loading factors between each principal component and the observed groundwater parameters were greater than 0.50. The first principal component (PC1) showed a strong positive loading on temperature (0.957) and turbidity (0.956), explaining 11.28 percent of the variance in the groundwater data. This positively strong loading might have been due to the influence of domestic, industrial and agricultural activities, the underlying geological characteristics of the underground rocks surrounding the groundwater, and the seawater's influence. The variation in the temperature might be due to climate change and the introduction of substances that can trap heat into the environment (as a result of domestic, industrial and agricultural dispositions).

The second principal component showed a strong positive loading on calcium (0.747) and potassium (0.735), accounting for 8.50 percent of the variance in the groundwater data. The positive load of the second principal components of calcium and potassium might be due to anthropogenic activities (like farming) and weathering processes of the underlying rocks. The principal component three exhibited a moderate positive loading on chloride ion (0.785), dissolved oxygen (0.534) and sulphate (0.511), explaining 7.74 percent of the variation in the groundwater data observed. This variation between the Dissolved oxygen and chloride and sulphate ions could be attributed to agricultural activities such as uncontrolled application of chemical fertilisers and frequent disposal of domestic wastes in the regions and run-off from such non-point sources pollutants into the underground water system.

Wells with a high concentration of total dissolved substances might be due to the alkalinity of the soil, mineral composition of the soil, high dissolved ions such as chlorine and calcium and the intrusion of the seawater. The variation in the chloride ion concentration might be due to the distance of some of the wells away from the seawater, saltwater intrusion, mineral dissolution, industrial and domestic waste and the underlying mineral composition of the area enough cations such as calcium and magnesium.

There was a strong positive loading from component four on Bicarbonate and Silicate at 0.730 and 0.565, respectively, explaining 7.65 percent of the variation in the groundwater data. This variation might be due to the presence of cations such as calcium and magnesium and the mineral composition of the soil. The load of the fourth principal component onto hydrogen bicarbonate ion might be due to the hardness of the groundwater, the mineral composition of the soil and the high temperature of the wells. The variation in the silicate ions is said to have been influenced by the mineral composition of the soil and domestic and industrial waste. The principal component five had strong negative loading on total hardness (-0.793), explaining 7.50 of the total variation in the groundwater data. The variation in the total hardness was due to the influence of seawater and the presence of cations such as magnesium and calcium.

Principal component six loaded positively on Phosphate at 0.0.671. The factor loading between PC 6 explained 6.65 percent of the total variation in the groundwater data. Also, the variation in the phosphate ions might be due to the mineral composition of the underlying soil, low fertiliser usage and domestic, industrial and agricultural production of products that contain phosphate ions. It might also be due to the phosphate erosion of rocks into underground water, chemical fertilisers, manure, and composted materials. The principal components seven and eight loaded strongly positively on electrical conductivity and magnesium at 0.810 and 0.863, explaining the variation in groundwater data observed at 6.17% and 5.96 percent, respectively.

The variation in the magnesium ion concentration of the groundwater signifies that the groundwater sampled is hard water and contains high dissolved substances, low pH, and high temperature (due to climate change) and domestic, industrial and agricultural waste production in the environment. The variation in electrical conductivity might be due to the presence of cations such as calcium and magnesium and the mineral compositions of the soil.

Nevertheless, there was no load between the principal components and observed groundwater parameters like pH, total hardness, sodium, nitrate and TDS, indicating that they did not load onto the groundwater data observed. They do not contribute to or explain the variation in the groundwater data.

Summary of Chapter

This chapter presented results and discussions on results from the field observation. The study discussed descriptive statistics of the physicochemical parameters and other analyses like PCA and ANOVA, as well as tables and charts. Wilcox diagram was employed to identify the main environmental factors (precipitation, rock and evaporation) that influence the wells sampled. Gibbs's diagram was used to determine the suitability of the wells for drinking. One-way ANOVA test demonstrated significant change between and among the means of the groundwater parameters sampled from the three regions.

CHAPTER FIVE

RESULTS AND DISCUSSION

Introduction

This chapter presents and discusses results obtained from the ML prediction analysis. It presents values of the predicted chemical parameters after training using test variables. Decision tree regression (DTR) and polynomial regression (PR) models were employed in this study to analyze the chemical parameters of the groundwater sampled from the study regions. This relates to the importance of checking the quality and drinkability of the groundwater in the study regions.

Research Objective 1: To develop a model and determine its performance in predicting groundwater chemical parameters. Research Objective 1 employed DTR and PR to predict groundwater chemical parameters using easily measured parameters. It demonstrated how DTR and PR were used to predict groundwater chemcial parameters. Results are presented in Tables 5.1 and 5.2.

Table 5.1: Predicted Mean Concentration of Chemical Parameters of Groundwater Quality and their Accuracies (All Regions)

Ions	Predicted Ion Concentration/mg/l	Prediction Accuracy
Na^{+}	244.00±101.12	0.83
K^{+}	121.00±19.48	0.92
Mg^{2+}	53.00±32.31	0.90
Ca^{2+}	172.00±51.05	0.96
HCO_3^{-}	12.00±27.81	0.81
Cl^{-}	147.00±78.64	0.80
NO_3^{-}	42.00±33.73	0.85
SiO_3^{-}	48.00±2.73	0.82
PO_4^{3-}	4.00±13.53	0.84
SO_4^{2-}	82.00±11.89	0.95

Table 5.1 presents the predicted groundwater chemical parameters and their accuracy. From the table, the predicted sodium (Na^{+}), potassium (K^{+}), magnesium (Mg^{2+}) and calcium (Ca^{2+}) concentrations were at accuracies of 0.83, 0.92, 0.90 and 0.96, respectively. Likewise, anion concentrations of hydrogen bicarbonate (HCO_3^{-}), chloride (Cl^{+}), nitrate (NO_3^{-}), silicate, phosphate (PO_4^{2}) and sulphate of 12.00 mg/l, 147.00 mg/l, 51.71 mg/l, 4.00 mg/l and 82.00 mg/l were predicted with accuracies of 0.81, 0.80, 0.85, 0.82, 0.84 and 0.95. The model's performance for calcium ions is the highest, followed by sulphate, potassium, magnesium, nitrate, phosphate, sodium, silicate, hydrogen bicarbonate, and chloride ions.

Table 5.2: Observed and Predicted Values of Machine Learning Model

Ions	Observed concentration/mg/l	Predicted concentration/mg/l
Na^{+}	242.44±113.28	244.00±101.12
K^{+}	124.92±35.37	121.00±19.48
Mg^{2+}	52.27±28.19	53.00±32.31
Ca^{2+}	171.76±48.06	172.00±51.05
HCO_3^{-}	10.91±3.30	12.00±27.81
Cl^{-}	147.72±80.76	147.00±78.64
NO_3^{-}	42.83±22.83	44.00±33.73
SiO_3^{-}	48.69±48.69	48.00±2.73
PO_4^{3-}	3.97±1.02	4.00±13.53
SO_4^{2-}	82.14±17.94	82.00±11.89

Table 5.2 compares the observed and predicted ion concentrations of the groundwater of the study regions. From the table, the predicted cationic sodium (Na^{+}) and calcium (Ca^{2+}) concentrations of (244.00 mg/l) and 171.76

mg/l were higher than the observed concentrations. The observed potassium (K) and magnesium (Mg) ion concentrations of 124.92 mg/l and 53.00 mg/l were less than the experimental values. Similarly, the predicted anions concentrations such as chlorine (Cl) 147.00 mg/l, silicate 48.00 mg/l and sulphate 82.00 mg/l were less than the observed values, while predicted hydrogen bicarbonate (HCO_3^-), phosphate (PO_4), nitrate (NO_3^-) and silicate ion concentrations were higher than observed concentration.

Table 5.3: Regression Analysis of Chemical Parameters (n = 10)

Regression Statistics	Values
R^2	0.999142
Adjusted R^2	0.874142
Standard Error	3.349167

Table 5.3 shows a regression analysis of chemical parameters. From the table, determinants like R^2 and adjusted R^2 were noted to be 0.999142 and 0.874142, respectively, with a standard error of 3.35.

Regression Graph of Predicted Per Sampling Location (n =300)

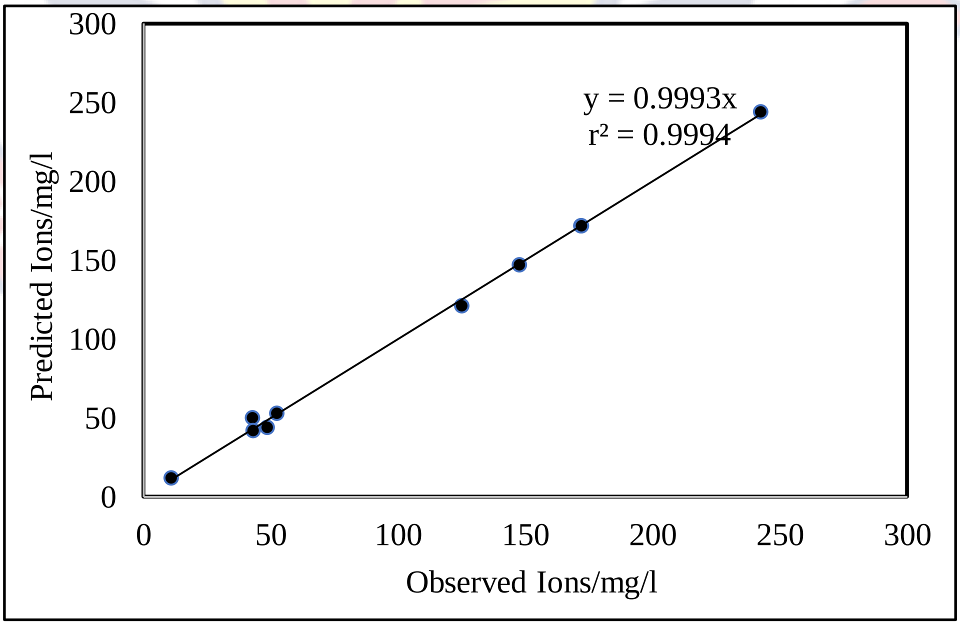

Figure 5.1: Regression Graph (Observed Ions versus Predicted Ion) (n = 10)

Figure 5.1 represents a regression graph of the observed and predicted ions. As displayed in Figure 5.1, the root square is 0.9994, and the equation of the graph is y = 0.993x. The R^2 indicates that SLA can be used to predict. According to Valentini et al. (2021), a low R^2 indicates the failure of a model to be used for prediction.

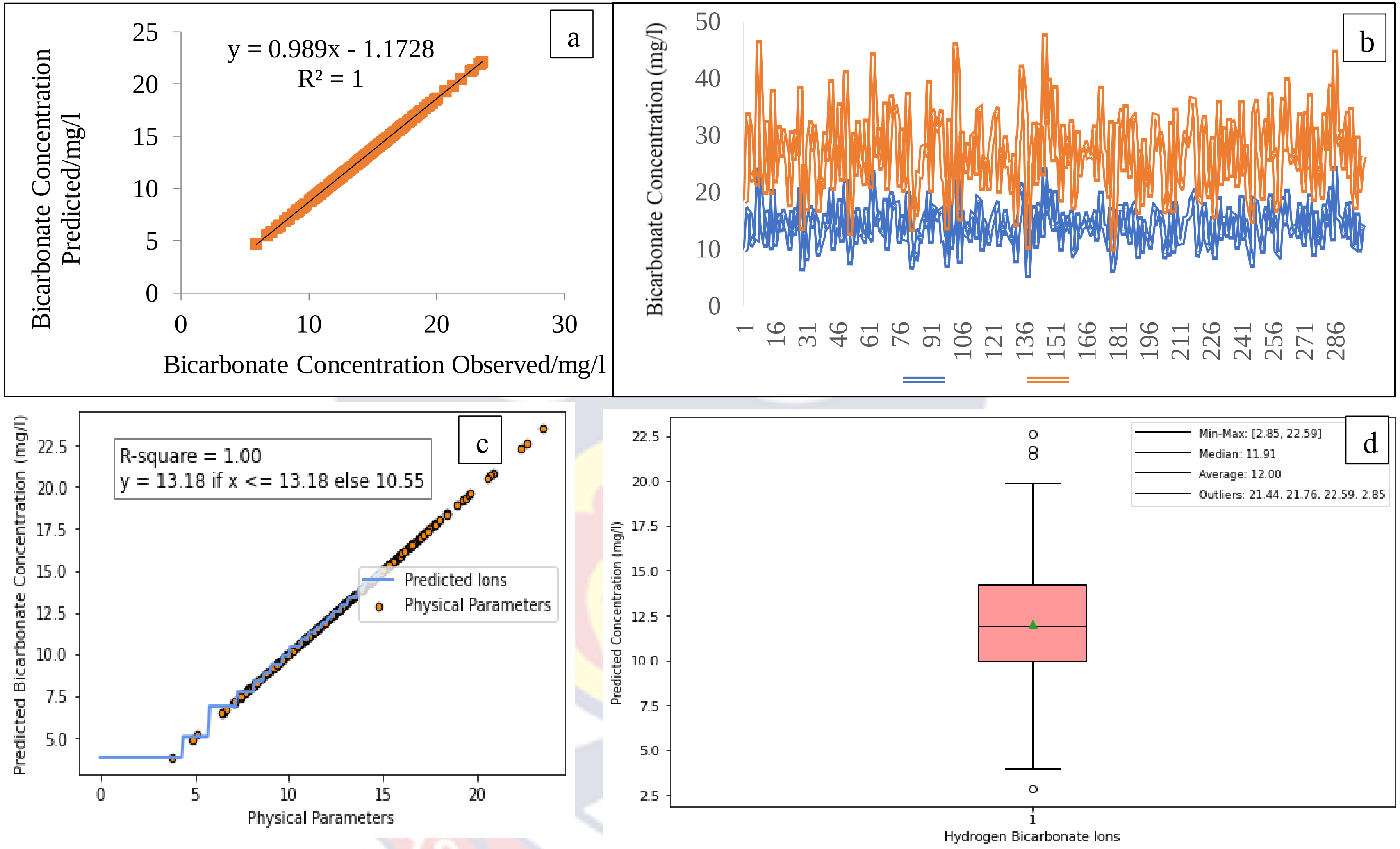

Figure 5.2: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Bicarbonate Concentration Prediction (n =300)

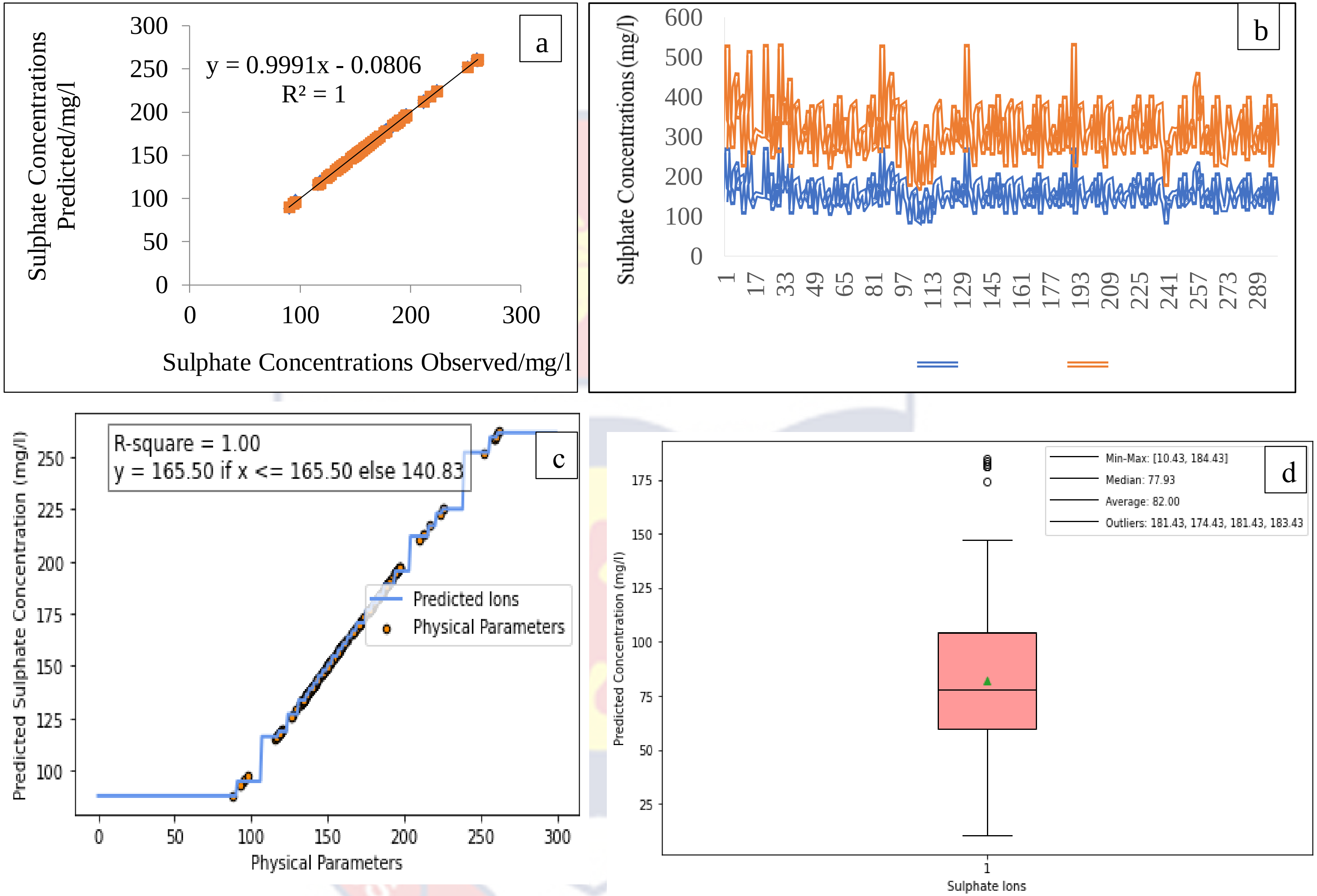

Figure 5.3: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Sulphate Ion Concentration Prediction (n =300)

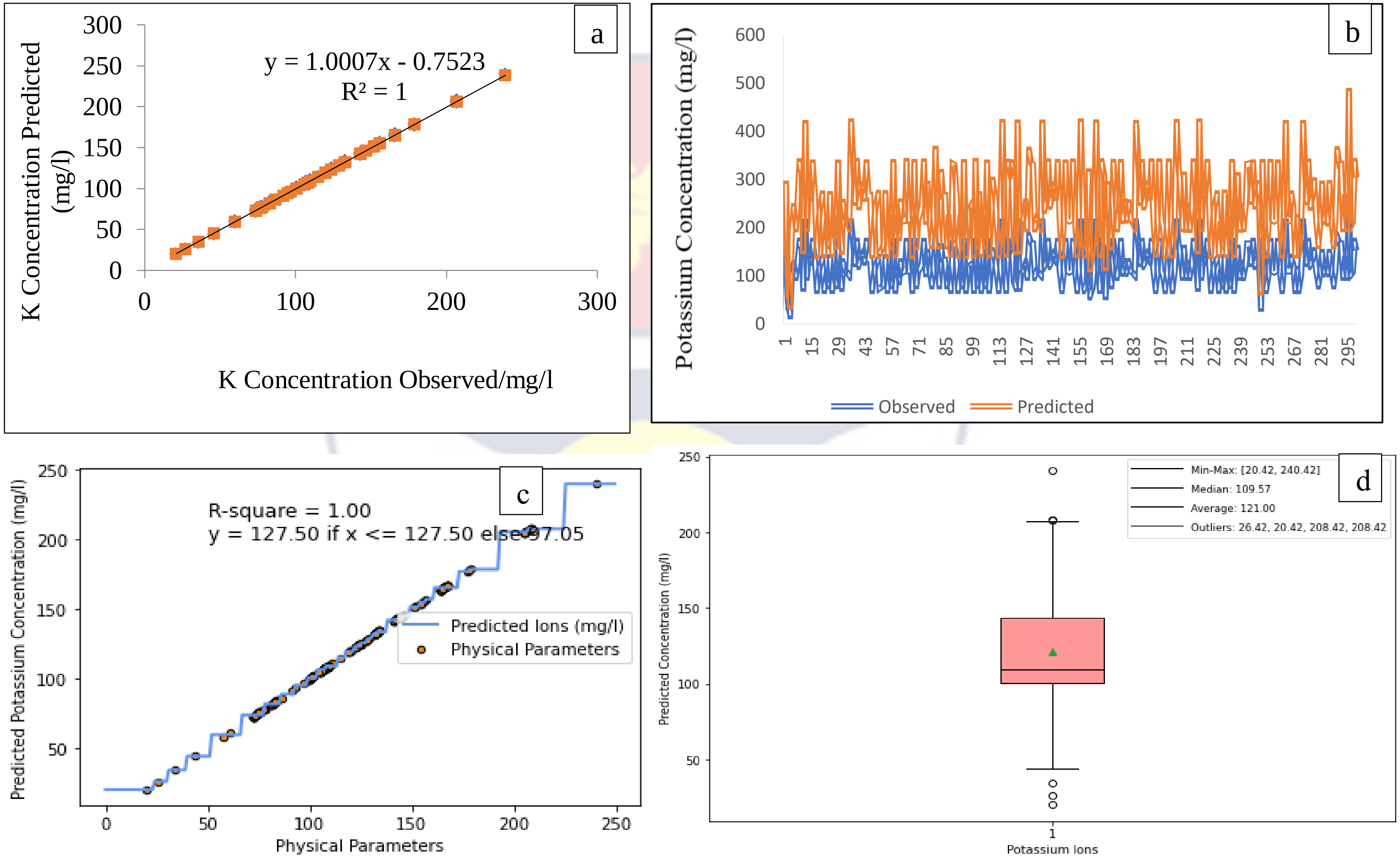

Figure 5.4: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Potassium Ion Concentration Prediction (n =300)

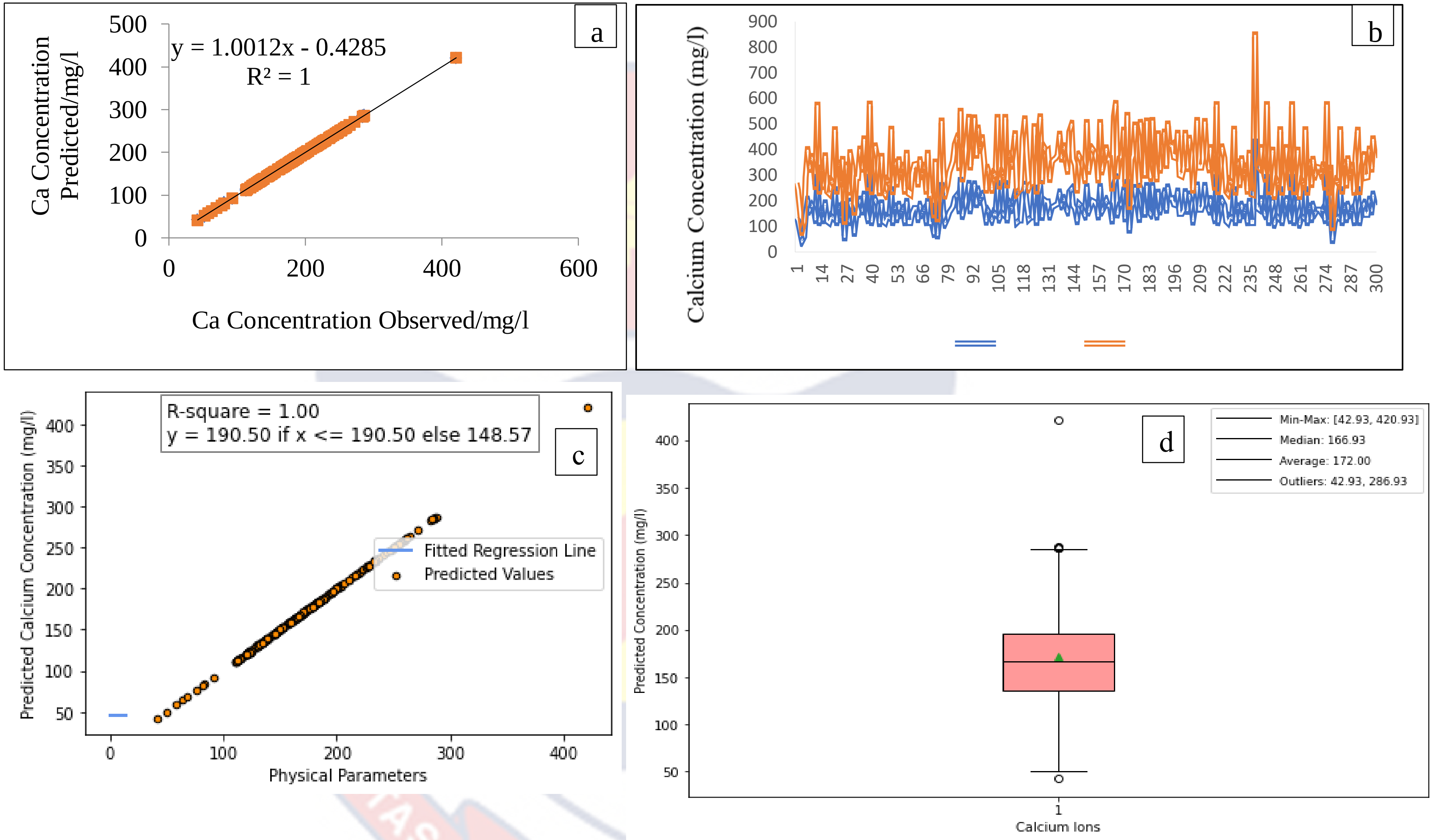

Figure 5.5: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Calcium Ion Concentration Prediction (n =300)

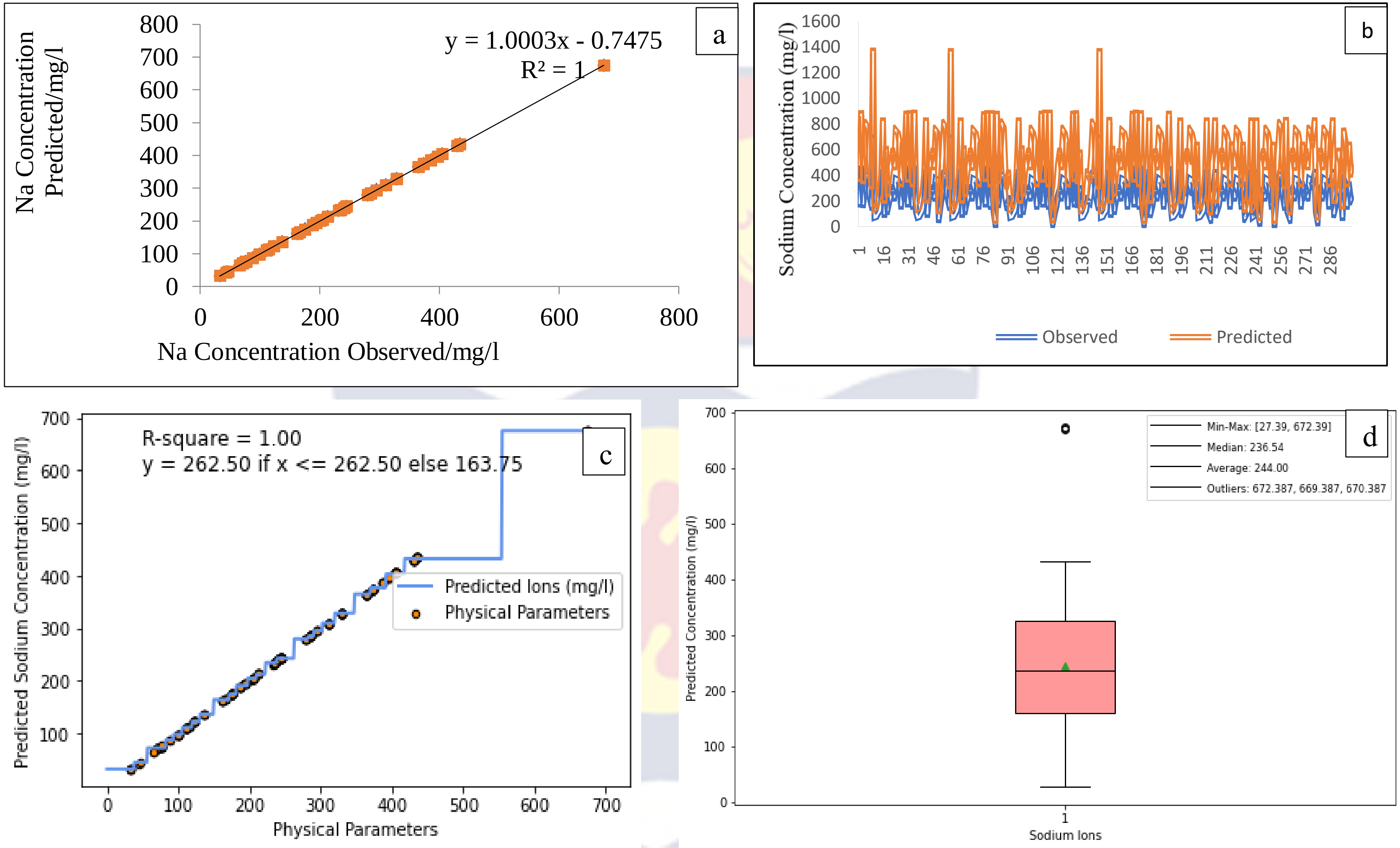

Figure 5.6: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Sodium Ion Concentration Prediction (n =300)

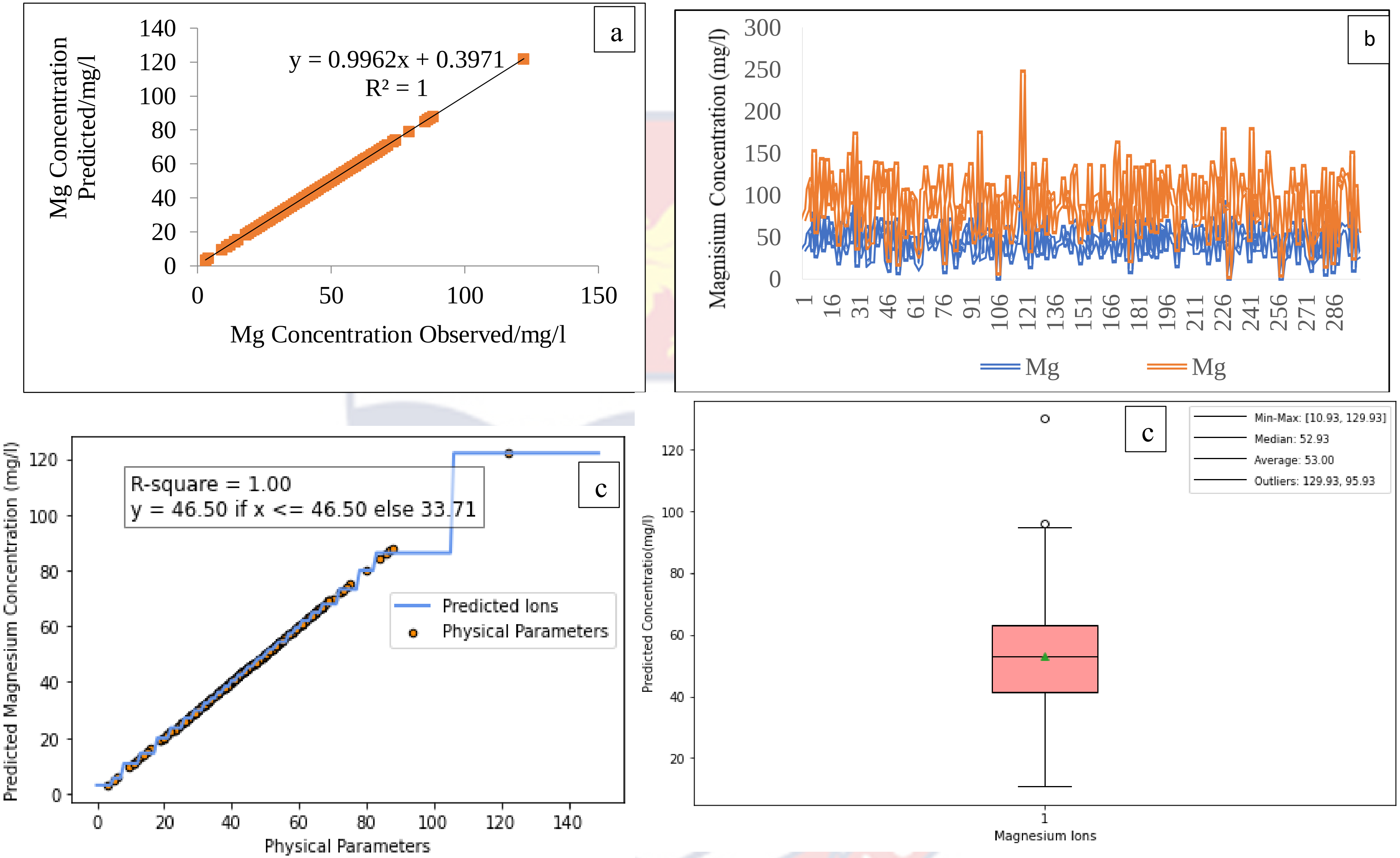

Figure 5.7: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Magnesium Ion Concentration Prediction (n =300)

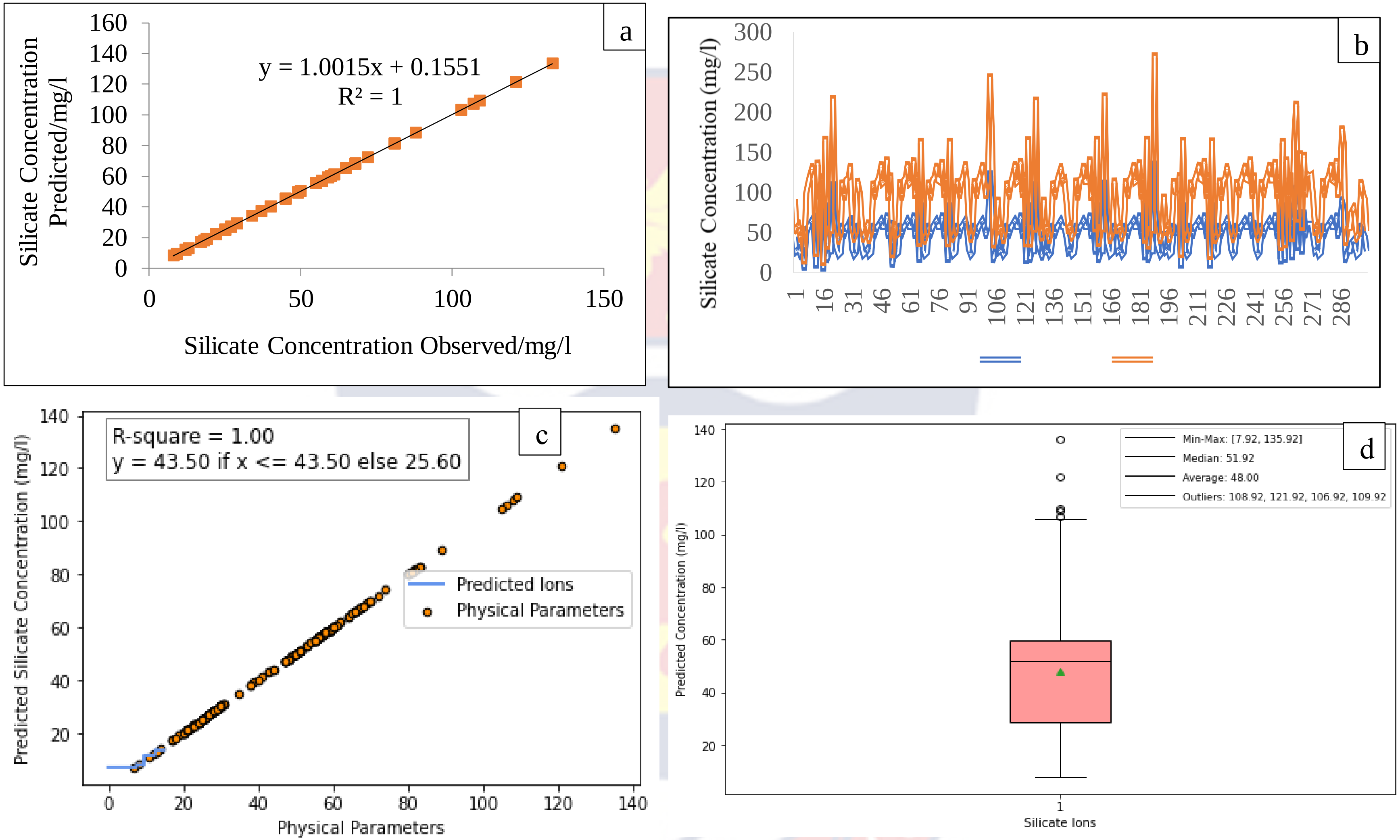

Figure 5.8: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Silicate Ion Concentration Prediction (n =300)

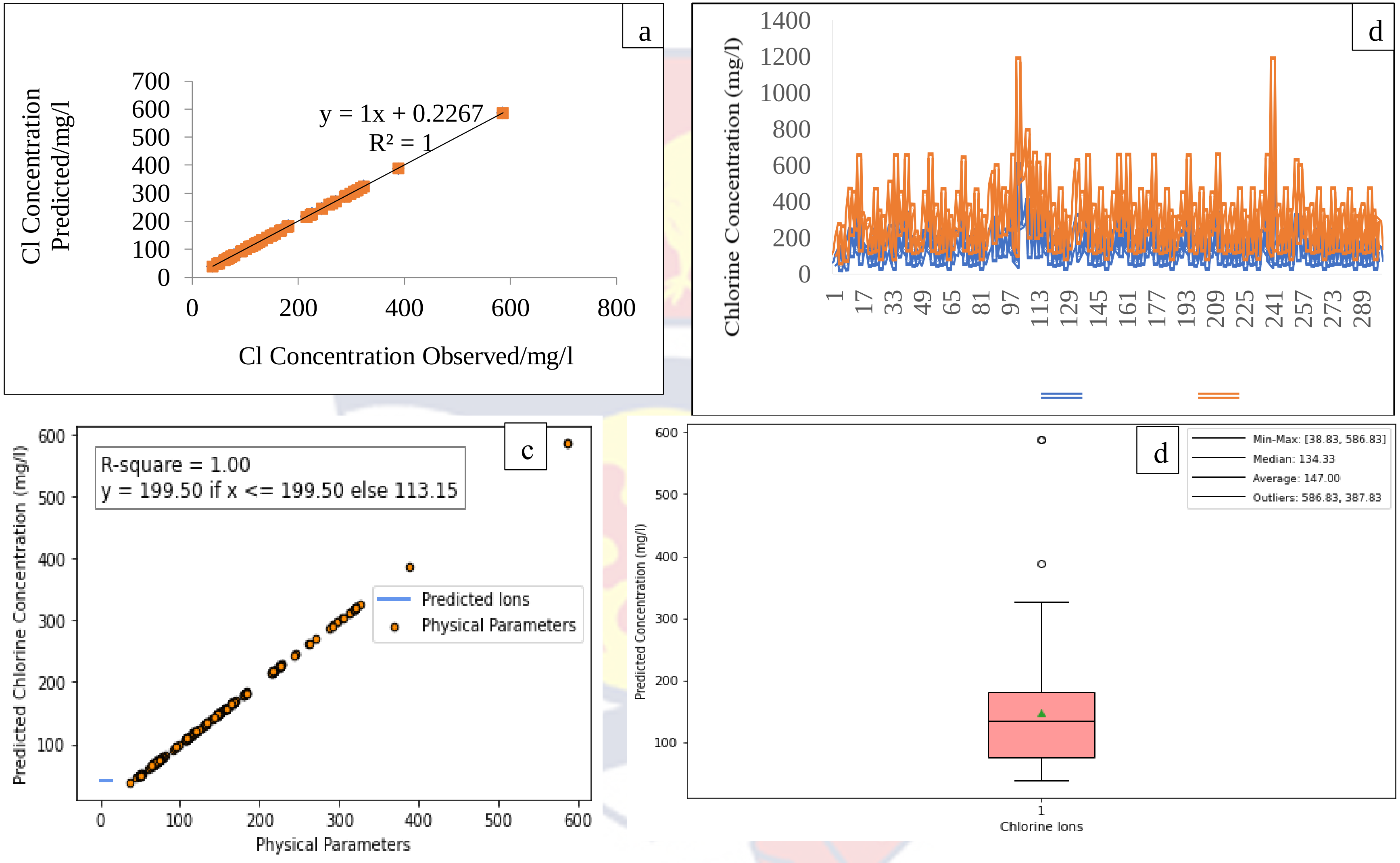

Figure 5.9: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Chloride Ion Concentration Prediction (n =300)

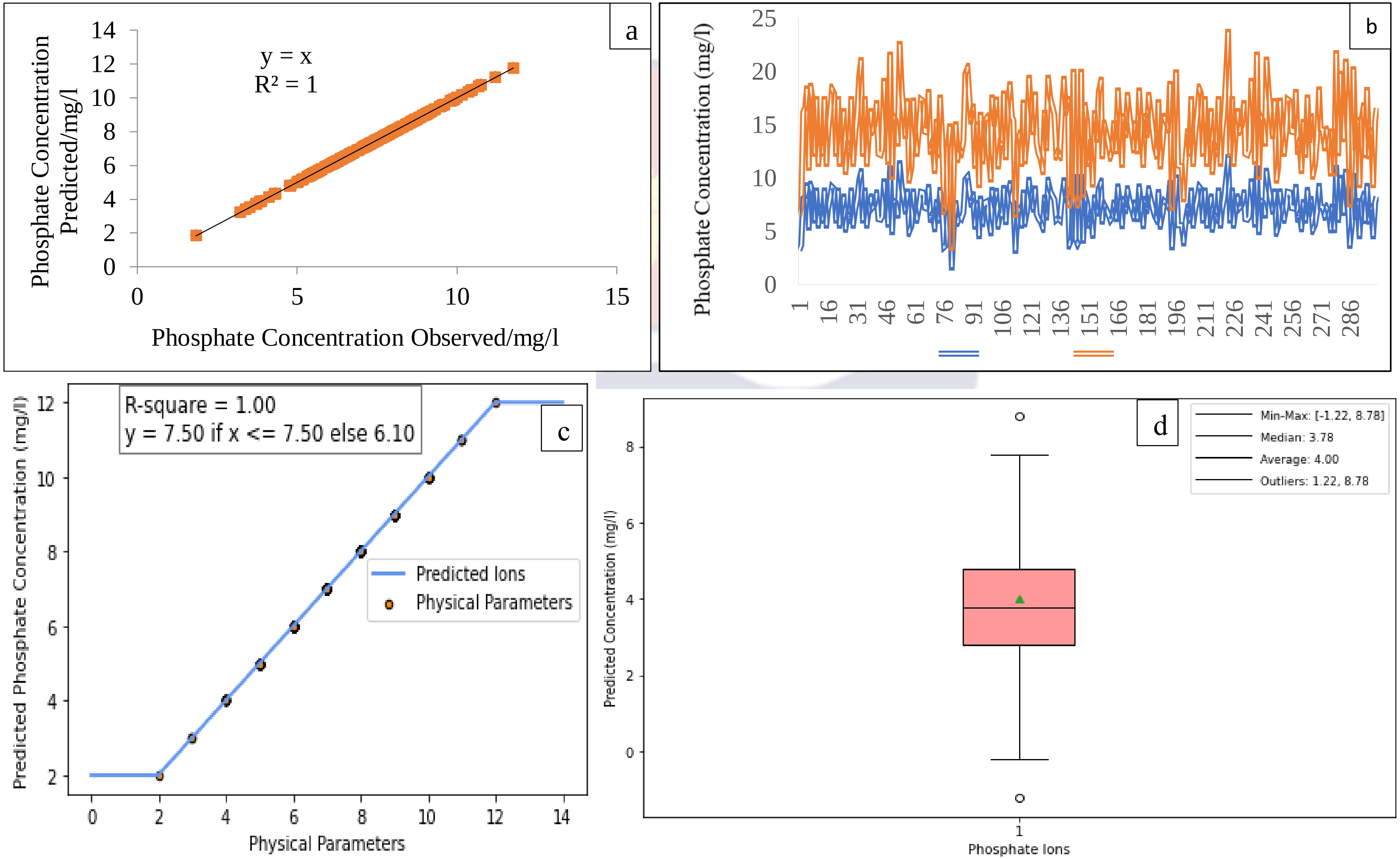

Figure 5.10: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Phosphate Ion Concentration Prediction (n =300)

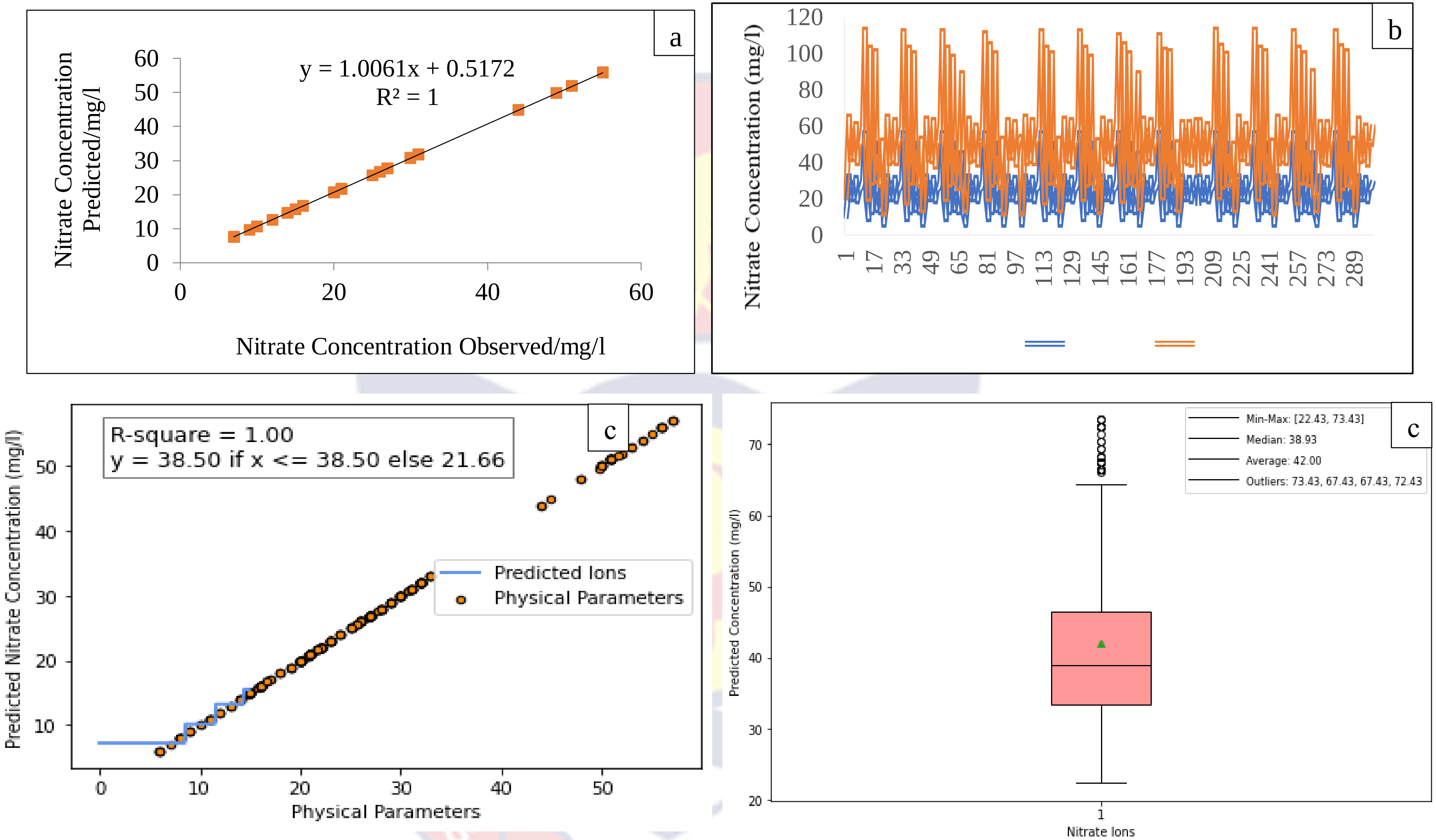

Figure 5.11: Polynomial Regression (a), Stacked Line (b) and DTR (c) Graphs and Boxplot (d) Showing Nitrate Ion Concentration Prediction (n =300)

Figures 5.2, 5.3. 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10 and 5.11 show polynomial regression (a), stacked line (b) and DTR (c) graphs and boxplot of the predicted chemical concentrations (HCO_3^-, SO_4^{2-}, K^+, Ca^{2+}, Na^+, Mg^{2+}, SiO_3^-, Cl, NiO_3^-, and PO_4^{3-}). The polynomial regression and DTR graphs and boxplots display equations, R^2 and best-fit lines. The equations show the non-linear relationship between the observed and predicted values, that is, the possibility of the observed concentrations influencing the predicted chemical concentration of the groundwater sampled from the study areas.

Model Evaluation Matrix

Models have deficiencies, and it is always appropriate to validate their performance (Sun et al., 2012). The model for the study was trained and used to predict the chemical parameters of the groundwater parameters of the study regions. Following groundwater quality parameter predictions, the prediction efficiency or performance of the model was evaluated using RMSE. The RMSE estimates the model's error, influenced by vertical distances between the dataset and the regression line. The smaller the RMSE, the higher the performance of the model and its predictability. The estimated RMSE is presented in Table 5.4.

Table 5.4: Root Mean Square Error Estimation for the Observed and Predicted Groundwater Quality Ion Concentration (All Regions)

	Concentration/mg/l							
Ions	Observed concentration	Predicted concentration	(O-P)/mg/l	MAS	(O-P)(O-P)	MAE	RMS	RMSE
Na^{+}	242.44	244	-1.56	1.56	2.4336	1.63	2.86	1.69
K^{+}	124.92	121	3.92	3.92	15.3664			
Mg^{2+}	52.27	53	-0.73	0.73	0.5329			
Ca^{2+}	171.76	172	-0.24	0.24	0.0576			
HCO_{3-}	10.91	12	-1.09	1.09	1.1881			
Cl^{-}	147.72	147	0.72	0.72	0.5184			
NO_3^{-}	42.83	50	-7.17	7.17	51.4089			
SIO_3^{-}	48.69	48	0.69	0.69	0.4761			
PO_4^{3-}	3.97	4	-0.03	0.03	0.0009			
SO_4^{2-}	82.14	82	0.14	0.14	0.0196			
				16.29	**72.00**			

Where MAS = Absolute difference between observed (O) and predicted (P) values,

MAE = Mean absolute error

RMSE = Root mean square error

From Table 5.4, MAE, RMS and RMSE of the predicted ions determine the performance of MLA in predicting the ions. From the table, the estimated MAS for the cations such as sodium (sodium, potassium, magnesium and calcium were 1.56, 3.92, 0.73 and 0.24, respectively. Anions such as hydrogen bicarbonate, chlorine, nitrate, silicate, sulphate and phosphate featured RMS as follows 1.09, 0.72, 7.17, 4.69, 0.97 and 0.02. The model performance for sulphate and calcium ions was the finest, followed by chloride, magnesium, phosphate and hydrogen bicarbonate ions. The model performs poorly for nitrate ions compared to the other ions. In general, the model achieves a very low RMSE of 1.690, showing the predictability of the model in forecasting groundwater quality parameters with great accuracy and precision.

Research Objective 2: To determine the predictability of chemical parameters of groundwater using the easily measured parameters.

Research Objective 2 validated the performance of ML in groundwater quality prediction using easily measured parameters. To achieve this objective, the performance of the ML was compared with aqueous geochemical models such as Visual Minteq, Phreeq C and Wateq4F. The result is presented in Tables 5.5 and 5.6.

Table 5.5: Observed and Predicted Values of the Models (Visual Minteq, Phreeq C and Wateq4F with Machine Learning Model)

		Model Predicted Values/mg/l			
Ions	Observed Values/mg/l	Visual Minteq	Phreeq C	Wateq4F	Machine Learning
Na^+	242.44	241.85	331.05	241.85	244.00
K^+	124.92	124.10	123.55	124.06	121.00
Ca^{2+}	171.76	135.42	141.19	144.52	172.00
Mg^{2+}	52.27	21.74	9.50	87.64	53.00
Cl^-	147.72	147.83	141.02	147.83	147.00
HCO_3^-	10.91	10.40	10.40	10.40	12.00
SO_4^{2-}	82.14	63.93	64.26	63.83	82.00
PO_4^{3-}	3.97	3.93	3.93	3.93	4.00
NO_3^-	42.83	41.67	41.67	41.67	44.00
SiO_3^-	48.69	31.55	51.65	53.13	48.00

Table 5.5 compares the observed and predicted values of the Visual Minteq, Phreeq C, Wateq4F and ML. From Table 5.5, sodium ion concentration predicted by Visual Minteq (241.85 mg/l) and Wateq4F (241.85 mg/l) were very close to the observed value (242.44 mg/l) compared to Phreeq C (331.05 mg/l) and ML (244.00 mg/l). The predicted potassium ion concentrations of Visual Minteq and Wateq4F at 124.10 mg/l were very close to the observed value (124.92 mg/l) compared to Phreeq C (123.55 mg/l) and ML (121.00 mg/l). The predicted calcium, magnesium, chloride, sulphate and silicate ion concentrations of the ML model at 172.00 mg/l, 53.00 mg/l, 147.00 mg/l, 82.00 mg/l and 48.00 mg/l, respectively, were close to the observed values compared to the Visual Minteq, Phreeq C and Wateq4F.

Table 5.6: Comparing Machine Learning Model's Performance with Visual Minteq, Phreeq C and Wateq4F using Statistical Parameters of Model Prediction

Statistical Parameters	Visual Minteq	Phreeq C	Wateq4F	ML
R-square	0.997	0.999	0.972	0.999
RMSE	16.970	33.160	15.330	1.690
D-statistics	0.987	0.960	0.988	1.000

Table 5.6 compares the performance of the AI model with Visual Minteq, Phreeq C and Wateq4F, using the statistical parameters including R^2, RMSE and d-statistics. As displayed in Table 5.6, the estimated R^2 values of Visual Minteq were 0.997, Phreeq C 0.999, Wateq4F 0.972 and ML model 0.999. The estimated RMSE for Visual Minteq, Phreeq C, Wateq4F and ML model were 16.97, 33.16, 15.33 and 1.690, respectively. Also, d-statistics for Visual Minteq were 0.987, Phreeq C (0.960), Wateq4F (0.988) and model (1.00).

Research Objective 3: To establish if there were statistically significant mean differences between and among the predicted groundwater chemical parameters of the study regions. Research Objective 3 determined whether there are significant differences in the mean values of predicted groundwater chemical parameters across the study regions. The chemical parameters of the groundwater for each region were predicted and compared. ANOVA was used to determine the statistical significance level among the three regions. Results are presented in Tables 5.7 and 5.8.

Table 5.7: Comparing ML Predicted Ions of Each Region

	Central (mg/l)		Greater Accra (mg/l)		Western (mg/l)	
Ions	Observed	Predicted	Observed	Predicted	Observed	Predicted
Na+	232.29	234	144.43	144	249.04	155
K+	118.49	118	126.98	124	129.56	128
Mg2+	51.72	50	55.06	51	50.02	47
Ca2+	160.05	160	185.76	162	168.46	168
HCO3-	10.74	14	11.23	16	13.04	13
Cl-	141.97	138	68.94	65	138.52	138
NO3-	21.96	27	17.64	18	24.6	22
SiO3-	47.59	55	35.78	35	50.71	49
PO_4^{3-}	3.1	3	4.81	4	5.12	6
SO42-	75.24	78	67.25	66	98.11	98

Table 5.7 compares the observed and predicted ion concentrations for each region, namely Central, Greater Accra, and Western Regions. In the Central Region, the observed sodium (232.29 mg/l), hydrogen bicarbonate (10.74 mg/l), nitrate (21.96 mg/l), silicate (47.97 mg/l) and sulphate (75.24 mg/l) ions were less than the correspondent ML prediction at 232.00 mg/l, 14.00 mg/l, 27.00 mg/l and 78.00 mg/l, respectively. The observed potassium (118.49 mg/l), magnesium (51.72 mg/l), calcium (160.05 mg/l) and phosphate (3.10 mg/l) were greater than the ML predicted ions.

The observed sodium concentration in Greater Accra Region stands at 144.43 mg/l, closely aligned with the ML-predicted value of 144 mg/l. The observed potassium (126.98 mg/l) was less than the ML predicted value (124 mg/l.) The observed magnesium (55.06 mg/l) and calcium (187.76 mg/l) were greater than the ML ion (51 mg/l and 162 mg/l, respectively). Hydrogen bicarbonate, chloride, nitrate, silicate, phosphate, and sulphate ions exhibit relatively close alignment between observed and predicted concentrations, with minor variations observed. In Western Region, the observed sodium concentration (249.04 mg/l) was higher than the ML-predicted value (155 mg/l). The observed potassium ion (129.56 mg/l) was less than the ML predicted ion (128 mg/l). Magnesium ions (50.02 mg/l) were slightly higher compared to the ML prediction (47 mg/l). Calcium concentrations exhibited a close alignment between observed (168.46 mg/l) and ML-predicted (168 mg/l) values. Additionally, hydrogen bicarbonate, chloride, nitrate, silicate, phosphate, and sulphate ions show relatively consistent alignment between observed and predicted concentrations, with minor variations observed.

Regression Plotted Graphs of Each of the Region: observed Values Verses Model Predicted Values (n = 10)

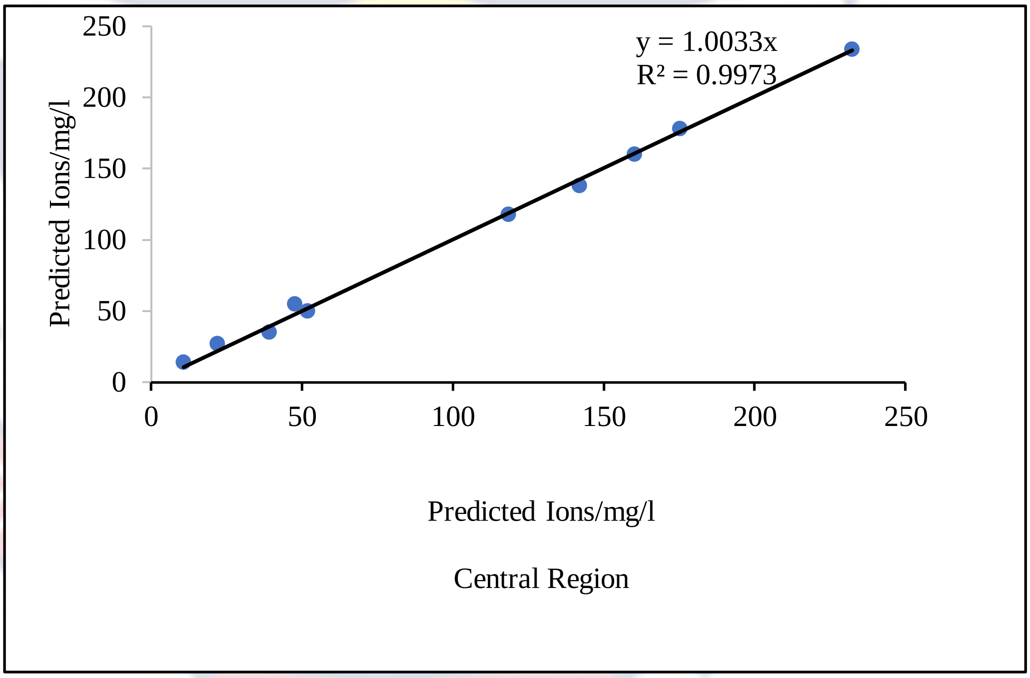

Figure 5.12: Predicted Mean Concentration of the Chemical Parameters of the Groundwater Sampled from Central Region (n = 10)

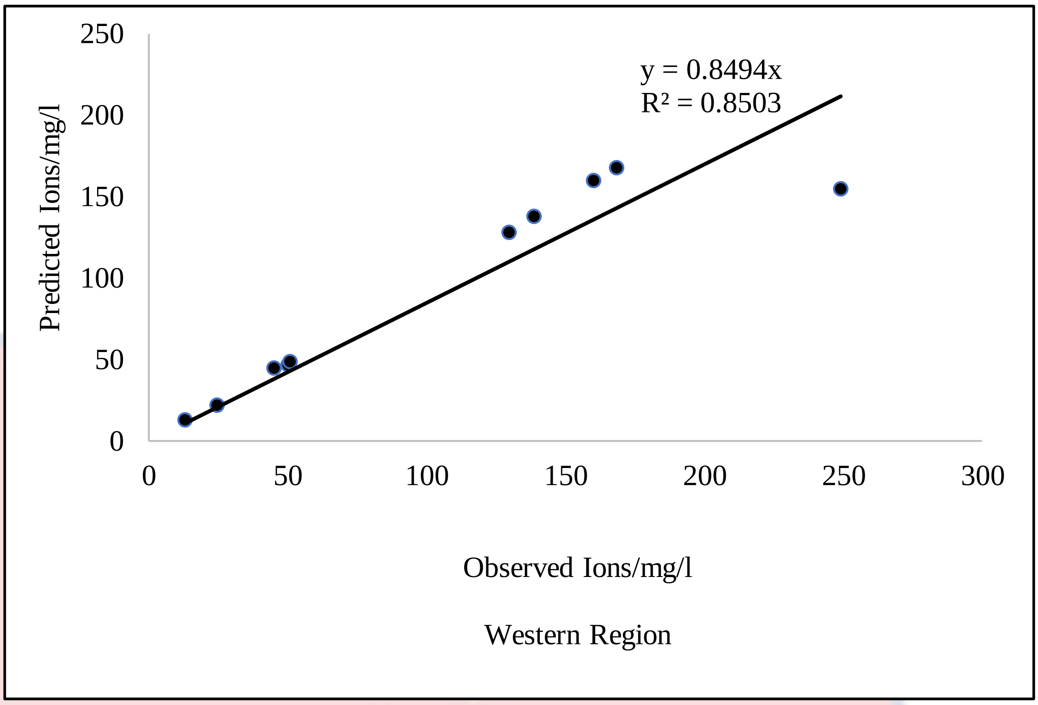

Figure 5.13: Predicted Mean Concentration of the Chemical Parameters of the Groundwater Sampled from the Western Region (n = 10)

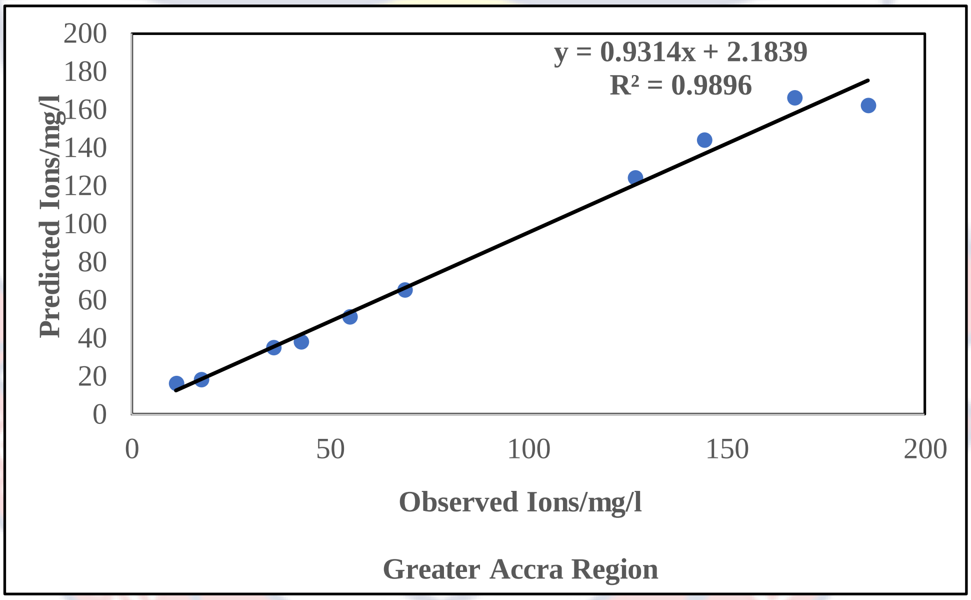

Figure 5.14: Predicted Mean Concentration of the Chemical Parameters of Groundwater Sampled from the Greater Accra Region (n = 10)

Figures 5.12, 5.13 and 5.14 display regression plots of the observed ions and the predicted ions. From the figure, the estimated R^2 and regression equations for Central, Greater Accra and Western Regions were 0.991, 0.9896 and 0.9565, and y = 1.0033x, y = 0.93142 + 2.1839 and y = 0.8494x were determined for Central, Greater Accra and Western Regions, respectively.

Table 5.8: ANOVA Test Showing Mean Differences in Predicted Groundwater Chemical Parameters among the Three Regions

ANOVA

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	1017.185	2	508.5926	0.160395	0.852713	3.402826
Within Groups	76101.11	24	3170.88			
Total	77118.3	26				

Where SS = sum of squares, Df = degree of freedom, MS = mean square, F = F-Statistic, P-value = probability value and F-critical = Critical F-value.

The 5.8 demonstrates whether there are mean differences in predicted groundwater chemical parameters among the three regions. Between groups variation, SS (1017.19) and 2 degrees of freedom yielded a mean square of 508.5926. Within group variation, SS (76101.11) and 24 degrees of freedom resulted in a mean square of 3170.88. The total variation across all groups amounts to 77118.3, encompassing a total of 26 degrees of freedom. The calculated F-statistic of 0.160395, accompanied by its corresponding P-value of 0.852713. This indicates that there is a lack of statistically significant differences in predicted groundwater chemical parameters among the regions at the designated significance level. This inference is drawn through comparison with the critical F-value of 3.402826, showing there is an absence of substantial variations in groundwater chemistry across the studied regions.

Groundwater Quality Indexes for the Three Regions

The study determined the Groundwater quality index to ascertain the effects of the chemical groundwater parameters on the water quality, using the predicted groundwater data in Table 5.5. The water quality index was determined from the ten (10) parameters of the study water quality parameters based on theoretical significance. The study used the formula,

$$WQI = {qiwi}/{\sum wi} \qquad [5.1]$$

for the calculation of the groundwater quality index.

wi is the unit weight,

qi is the water quality rating of 10 groundwater quality variables used in the estimation of the water quality index.

PO_4^{3-} and HCO_3^- ions were not included in the calculation of the groundwater quality index since they influenced the water quality index abnormally.

Table 5.9: Water Quality Index (WQI) Generation

Predicted Ions/mg/l	WHO	1/Sn		Σ1/Sn	K = 1/(Σ1/Sn)	S = K/Sn	Ideal Values	Predicted Values	Vn/Sn	Vn/Sn X 100
Na^{+}	200.00	0.01	10.25	0.10	0.00	0.00	244	1.22	122.00	0.06
12kl2Cl-	250.00	0.00	10.25	0.10	0.00	0.00	147	0.59	58.80	0.02
K^{+}	300.00	0.00	10.25	0.10	0.00	0.00	121	0.40	40.33	0.01
Mg^{2+}	50.00	0.02	10.25	0.10	0.00	0.00	53	1.06	106.00	0.21
SO_4^{2-}	250.00	0.00	10.25	0.10	0.00	0.00	82	0.64	63.60	0.02
Ca^{2+}	100.00	0.01	10.25	0.10	0.00	0.00	172	1.72	172.00	0.17
PO_4^{3-}	2.00	0.50	10.25	0.10	0.05	0.00	4	26.50	2650.00	129.21
HCO_3^{-}	10.00	0.10	10.25	0.10	0.01	0.00	12	1.20	120.00	1.17
SiO_3^{2-}	100.00	0.01	10.25	0.10	0.00	0.00	44	0.57	57.00	0.06
NO_3^{-}	50.00	0.02	10.25	0.10	0.00	0.00	42	1.00	100.00	0.20
	Σ1/Sn =	0.68			0.07		26.63			131.13

Table 5.9 presents the calculation of WQI of the study regions (Central, Greater Accra and Western Regions). The calculated WQI for the three study regions combined was 131.13.

Comparison of Groundwater Quality Indices (Central, Greater Accra and Western Regions)

The study compared WQI of the study regions. The formula as shown in Table 5.9 was used to calculate WQI of the study regions for comparison. The result is presented in Figure 5.15.

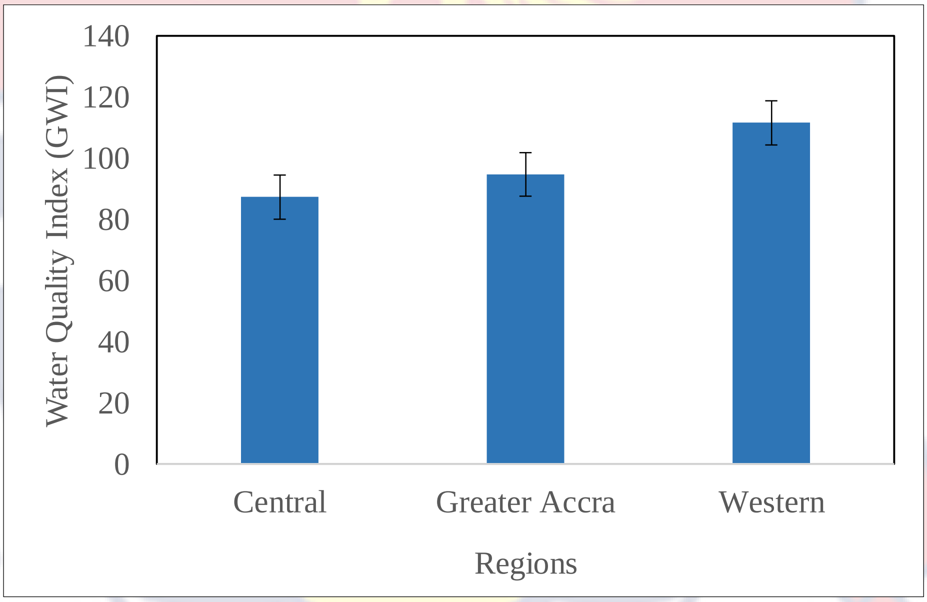

Figure 5.15: Calculated Water Quality Index of Each Region (Central, Greater Accra and Western Regions).

Figure 5.15 compares WQI of the study regions (Central, Greater Accra and Western Regions). WQI of Western Region (111.56) was the highest compared to Greater Accra (94.75) and Central Regions (87.32). Western Region recorded higher WQI compared to Central Region. Central Region featured the least WQI and thus, the purest water quality.

Table 5.10: Regional Groundwater Quality Classification Comparison (%)

Regions	Suitability Classification					
	Unsuitable	Very poor	Poor	Good	Excellent	Total
CGW	6	5	11	64	14	100
Central	2	4	3	72	19	100
Greater Accra	13	7	16	57	7	100
Western	7	12	22	49	10	100

Where GCW = Central, Greater Accra and Western Regions

Table 5.10 displays the groundwater quality classification of the study regions. The classification scheme consisted of unsuitable, very poor, poor, good and excellent. The groundwater classification of the study regions was done based on the WHO water quality guidelines. It shows the degree to which each of the wells sampled has been polluted or if the wells sampled meet the WHO standards for quality water for human consumption.

As represented in Table 5.8, out of the 300 wells sampled of the study regions (Central, Greater Accra and Western Regions), 6 percent of the wells were not suitable, 5 percent were very poor, 11 percent were poor, 64 percent were good and 14 were excellent for drinking. Among the three regions, more of the wells of Greater Accra Region (13%) were unsuitable for human consumption compared to Western and Central Regions, at 7 percent and 2 percent, respectively. Western Region recorded the highest very poor wells (12%), followed by Central (7%) and Greater Accra (4) Regions. Wells with the highest poor quality were recorded for Western Region (22%) and then Greater Accra (16%) and Central (3%) Regions.

Central Region featured the region with the highest number of wells good for human consumption (72%), making the Western (49%) the least region with groundwater quality good for drinking after Greater Accra Region (57%). Furthermore, 19% of the wells of Central Region (13%) were excellent for human consumption compared to Greater Accra (7%) and Western (10) Regions.

Discussion on Model Prediction Analysis

This section presents discussions on the model prediction and the analysis of the chemical parameters of groundwater quality using easily measured data. The data were analysed using SL algorithm, particularly DTR and PR. The values of the predicted ions show the chemical parameters of the groundwater sampled after training, using the test variables.

Predicted concentration of groundwater quality ions with accuracy (all regions)

Regarding Table 5.1, the model predicted groundwater chemical parameters demonstrated very high accuracy. The accuracy of the cations such as Ca^{2+} (0.96), K^{+} **(**0.92), Mg^{2+} (0.90) and Na^{+} (0.82) ion concentrations showed very high performance. For the cations, the model had a very high performance for calcium ions compared to potassium, magnesium and sodium ions, indicating that the model is best fit for predicting calcium ion concentrations. Likewise, the accuracy of anions such as $SO4^{2-}$ (0.95), $PO4^{3-}$, NO_3^{-} (0.84) and (0.83) displayed very high performance. Regarding the anions, the sulphate ion prediction is the highest compared to phosphate, nitrate, silicate, chloride, and hydrogen bicarbonate ions. Generally, the model had a very high performance for ions like calcium, sulphate, potassium, magnesium, nitrate, phosphate and sodium.

The model showed low performance for hydrogen bicarbonate, chloride and silicate ions. The model has good performance cations compared to the anions. Comparing this result to Joslyn et al. (2018), ML model has high accuracy and performance and this might be due to the difference in location and the kind of groundwater parameters estimated. Joslyn analysed two parameters, while this study employed seven easily measured groundwater parameters to predict the chemical characteristics of the groundwater sampled. This might account for the overall significant change in the model's accuracy.

Comparing the accuracy of temperature (0.9789) and dissolved oxygen (0.9889) estimated by Joslyn to the accuracy of calcium (0.96) and sulphate (0.95) ion concentration predicted in this study, there is a significant difference. Concerning the accuracies of the predicted ions, the performance of the model for Ca^{2+} > $SO4^{2-}$ > K^{+} > Mg^{2+} > NO_3^{-} > PO_4^{3-} > Na^{+} > SiO_3^{-} HCO_3^{-} > Cl^{-}. Based on the findings, the question of whether the model can be used to estimate the groundwater chemical parameters is explained by the accuracy of each parameter. Based on the accuracy of the predicted ions and as compared to Joslyn, the model can be used to predict the chemical parameters of groundwater using easily measured parameters to very high accuracy.

Comparison observed and model predicted ion concentration of groundwater (all regions)

Observed and predicted values of machine learning model

From Table 5.2, ML estimated the ionic concentration of the chemical parameters in whole numbers. ML approximated the values greater than 0.49 to 1 and less than 0.49 to 0. The predicted cationic sodium (Na^{+}) and calcium (Ca^{2+}) concentrations of (244.00 mg/l) and 171.76 mg/l were higher than the observed concentrations, with very significant differences. The observed potassium (K) and magnesium (Mg) ion concentrations of 124.92 mg/l and 46.00 mg/l were less than the experimental values.

Although ML predicted ion concentration of potassium and magnesium might be less than the observed. This does not imply that the model is of poor performance. The significant difference between the model-predicted and observed values was not apparent from the model's prediction but from the approximation of the predicted values, variation in the predicted values and outliers. Similarly, the predicted anions concentrations such as chlorine (Cl) 147.00 mg/l, sulphate 172.00 mg/l and silicate (48.00 mg/l) ions were less than the observed values, while predicted phosphate 4.00mg/l, hydrogen bicarbonate (HCO_3^{-}) and nitrate (NO_3^{-}) concentrations were higher than the observed concentration. Comparing the observed and model-predicted values, it is established that the model has good performance for predicting groundwater quality parameters.

Regression analysis: Determining the predictability of the chemical parameters using groundwater physical parameters

From Table 5.3, R^2 and adjusted R^2 of 0.999142 and 0.874142, respectively, proved that the model has high accuracy and can be used to investigate the chemical parameters of groundwater using easily measured parameters. According to Raheja et al. (2022), when either the root or the adjusted root square is higher (close to 1), the model has a higher predictability chance and thus, it can be reliable for other predictions where necessary.

Based on the regression analysis and the adjusted R^2, it is established that the model has high performance and accuracy, or can predict and forecast groundwater quality with minimal error. Comparing this finding to Kouadri et al. (2020), the model has achieved higher prediction accuracy and performance. Singha et al. (2021) who predicted and modelled the state of safe drinking water sources had R^2 of 0996. In comparison with this current study (R^2 = 0.999142), ML yielded better R^2 and this indicates that MLA could be used to model and predict groundwater quality using easily measured parameters.

Model evaluation, using RMSE

Models have deficiencies, and it is always appropriate to validate their performance (Sun et al., 2012). The model for the study was trained before predicting the chemical parameters of the groundwater sampled. Following the groundwater quality parameter predictions, the prediction efficiency or performance of the model was validated using RMSE. The smaller the RMSE, the higher the performance of the model and its predictability.

From Table 5.4, the estimated RMSE for the cations such as sodium, potassium, magnesium and calcium were 1.56, 3.82, 0.73 and 0.24, respectively. Anions such as chlorine, hydrogen bicarbonate, sulphate, silicate, phosphate and nitrate featured MAS as follows 0.72, 1.09, 0.14, 0.69, 0.00 and 7.17, respectively. Based on the RMSE, the model performance for sodium is the finest, followed by calcium, chlorine and hydrogen bicarbonate. The model performed poorly for silicate ions compared to sodium, calcium, chlorine and hydrogen bicarbonate. In general, the model achieves a low RMSE (1.69) and it was found to predict groundwater quality parameters with great accuracy and precision. This finding of the study supports a study by Elbeltagi et al. (2021) who found that AI tools could be used to model and predict water quality. Comparing the RMSE of this current study with Elbeltagi et al. (2021) who had RMSE of 0.6356, ML is said to produce better results.

From Figures 5.2, 5.3. 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10 and 5.11, determinants of ML such as R^2 values were > 95 percent, demonstrating that ML could be used to predict the groundwater chemical parameters using easily measured parameters at very high accuracy. MLA can be used to predict groundwater chemical parameters with high accuracy using easily measured parameters for each of the regions.

Comparing the observed and predicted values of Visual Minteq, Phreeq C and Wateq4F with ML

As presented in Table 5.5, Visual Minteq and Wateq4F could be used to predict sodium, potassium, hydrogen bicarbonate and Nitrate ion concentrations with high accuracy compared to Phreeq C and ML. On the other hand, ML could be a very good model for predicting calcium, magnesium, chloride, sulphate and silicate ion concentrations compared to the Visual Minteq, Phreeq C and Wateq4F. ML although might have low predicting performance for sodium, potassium, hydrogen bicarbonate and Nitrate ion concentrations does not dispute the fact that it could be used for predicting groundwater quality.

Comparing ML model's predicting ability with Visual Minteq, Phreeq C and Wateq4F

Objective 2 compared the predictability and performance of ML with Visual Minteq, Phreeq C and Wateq4F. R^2, RMSE and the d-statistic of the models were determined. As presented in Table 5.6, the estimated R-square value for Visual Minteq was 0.997, Phreeq C (0.999), Wateq4F (0.972) and Machine Learning model (0.999) as well as the d-statistic of the models [Visual Minteq (0.987), Phreeq C (0.960), Wateq4F (0.988) and Model (1.00)] indicated that the models have high predicting ability and performance.

The estimated RMSE for Visual Minteq, Phreeq C, Wateq4F and ML model consisted of 16.97, 33.16, 15.33 and 1.690, respectively. The RMSE of ML connote that it could help predict groundwater quality to high accuracy. Based on the R^2, RMSE and d-statistic, ML had the highest prediction ability and performance, followed by the Wateq4F, Visual Minteq and Phreeq C.

Comparing ML predicted ions of each region

ML model exhibited accurate predictions for sodium and calcium in Central Region, with observed and predicted values closely aligned. ML performed well in predicting chloride ions, with minor discrepancies observed. Hydrogen bicarbonate ions also showed relatively accurate predictions, despite slight deviations from the observed values. This suggests a commendable performance of the ML predicting chemical ions in the Central Region using easily measured parameters. There were discrepancies between observed and predicted values in the various ions. For example, sodium, hydrogen bicarbonate, nitrate, silicate, and sulphate ions exhibit lower observed concentrations compared to their corresponding ML predictions, while potassium, magnesium, calcium, and phosphate ions show higher observed concentrations.

In Greater Accra Region, the ML demonstrated better accuracy in predicting sodium and chloride concentrations, with observed and predicted values closely matching. Hydrogen bicarbonate concentrations also showed satisfactory predictions, albeit with minor variations. These results indicate the ML's effectiveness in estimating important ion concentrations. The observed concentrations of potassium, magnesium, and calcium ions deviate from the ML predictions, indicating potential discrepancies in the model's estimation. For other ions such as sodium, hydrogen bicarbonate, chloride, nitrate, silicate, phosphate, and sulphate, the observed and predicted concentrations demonstrate closer alignment. This suggests better accuracy in ML's predictions for these ions within this region.

The ML performed well in predicting chloride concentrations, with observed and predicted values closely aligned. Hydrogen bicarbonate concentrations also show consistent predictions, highlighting the model's capability to estimate these ions accurately. Although there may be some deviations, particularly in calcium concentrations, overall, ML demonstrated satisfactory performance in estimating key ion concentrations in the Western

Region. Notable differences were observed between observed and predicted sodium and potassium concentrations and this indicates potential challenges in accurately estimating these ions. The ML performed relatively well in predicting hydrogen bicarbonate, chloride, nitrate, silicate, phosphate, and sulphate ions, with minor variations between observed and predicted concentrations.

Regarding the performance of the ML in Central, Greater Accra, and Western Regions, ML exhibited relatively better performance in the Central Region compared to the Western or Greater Accra Region. ML demonstrated performance in predicting ions, including sodium, calcium, chloride, and hydrogen bicarbonate in the Central Region. Following the Central Region, ML exhibited relatively similar performance in Greater Accra and Western Regions.

ANOVA test showing mean differences in predicted groundwater chemical parameters among the three regions

From Table 5.8, the P-value (0.852713) was greater than the alpha value (0.05). This demonstrates that there is a lack of statistically significant differences in predicted groundwater chemical parameters among the regions. This implies that the predicted groundwater chemical parameters do not significantly differ among the Central, Greater Accra, and Western Regions. ML can be used for predicting groundwater chemical parameters using easily measured parameters across different geographical areas.

WQI of Central, Greater Accra and Western Regions

WQI of the three regions was 131.13, higher than the WHO recommended value for excellent drinking water. This high index implies that the groundwater has received progressive pollution. The water quality is reduced and barely good for human consumption and it has seldom departed from tolerable levels with only minor degradation. Water with this water quality needs consensus efforts of the government, individuals and non-governmental organisations to restore its ecological integrity.

The high-water quality index of Western Region (111.56) compared to Greater Accra (94.75) and Central (87.32) Regions depicts that the underground water in the regions is barely safe for drinking. Similarly, it signifies that the groundwater quality of Greater Accra and Western Regions is safer than the Western Region. Moreover, this high value reflects activities of the degradation of bedrock and the impacts of human activities on groundwater quality and contamination. However, the anthropogenic activities in the area might include illegal mining and improper disposal of domestic, agricultural and intertrial wastes in the region. Based on the WHO recommended values for safe drinking (100.1–200, Poor water), the groundwater in the region is graded ‘Poor’. This implies that the water quality in the region has been reduced and is barely good (Mohammad et al., 2020).

WQI of Central and Greater Accra Regions based on the WHO grading system is graded ‘Good water’ but not excellent for drinking. The region with the lowest water quality index is Central Region and thus, the purest water quality among the three regions, followed by Greater Accra Region, making Western Region the last. Nevertheless, it connotes that the water quality is with only minor degradation; conditions seldom depart from tolerable levels. Therefore, the water quality needs to be protected or sustained in order to maintain its quality, fitting for consumption and other purposes. The water quality indexes of each of the three regions were higher than the WHO recommended values of excellent drinking water (< 50, Excellent water). It is recommended effective monitoring programme should be implemented in the regions to help sustain the groundwater quality and sound mining operations. Hence, sustaining groundwater quality in the regions is as protecting health, sustaining water supply systems and meeting the water demands of the population growth in the regions.

Groundwater quality classification by regions

This study classified the wells of the study regions based on the following classification scheme; unsuitable, very poor, poor, good and using the WHO water quality guidelines. Any of these classification schemes depict the degree to which the wells sampled were polluted and whether they are safe for human consumption. When the classification scheme is 'excellent' means the quality of the wells is almost unaltered and is in a perfect state. 'Good' implies that the quality of the wells is maintained with only minor degradation conditions, which seldom depart from tolerable levels. 'Poor' connotes that the quality of the water is almost always reduced and it is barely good. 'Very poor' means that the quality of the wells is not near to be clean, while 'unsuitable' means that the quality of the wells is often compromised, and the characteristics are far from optimum. Furthermore, the wells classified 'unsuitable, poor and very poor' means that they are associated with health-related diseases, including cardiovascular, circulatory and respiratory diseases.

Regarding Figure 5.10, 6 percent of the wells in the study regions (Central, Greater Accra and Western Regions) were unsuitable, meaning they are not suitable for drinking purposes and the quality of the wells is often compromised, and the characteristics are far from optimum. Moreover, 5 percent classified 'very poor' shows that some of the wells are not near to be clean, while 11 percent classified 'poor' denotes that the quality of the water is almost always reduced and it is barely good. The wells classified 'good' affirm that the quality of the wells is maintained with only minor degradation conditions, which seldom depart from tolerable levels, whereas 14 of the wells classified as 'excellent' informs that the quality of the wells is almost unaltered and is in perfect state. These findings reaffirm the study of Lutterodt et al. (2018) who claimed that groundwater in Accra, particularly Dodowa of Ghana is unsuitable for drinking purposes. The groundwater should be treated before consumption and much effort should be made to reduce contamination from local sources of pollution and create restricted spaces surrounding groundwater collection points. The states of the wells being classified as 'unsuitable, poor, and very poor' calls for holistic efforts of government, non-government and individuals to protect and sustain the quality of the groundwater in the study regions and beyond.

Summary of Chapter

This chapter presented results and discussions ML prediction. The ML predicted data analysis the discussion was centred on analysis such DTR and polynomial regressions and groundwater quality indexes, as well as model validating tools such as RMSE. Based on the WHO classification scheme, 6 percent of the wells sampled were unsuitable, 5 percent very poor, 11 percent poor, 64 percent good and 14 excellent for drinking. Statistical parameters such as R-square, RMSE and d-statistics were used to validate the model's performance. The predicting ability of ML developed was compared with aqueous geochemical models like Visual Minteq, Phreeq C and Wateq4F using R^2, RMSE and d-statistics. Based on ML validating tools such as RMSE, the study found that ML model could be used to predict groundwater quality using the physical parameters.

CHAPTER SIX

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

This study employed supervised learning algorithms like DTR and PR to assess and predict groundwater quality. This section presents the summary of the study, key findings based on field data analysis and model predicted data analysis, conclusion, and recommendations including direction for further research.

Summary

The study employed SLA, specifically DTR and PR approaches to predict the groundwater quality of Central, Greater Accra and Western Regions. An experimental research design (factorial design) was used to conclude (predict groundwater quality), validate the hypothesis (analysis testing) and establish the causality (what contributes to groundwater pollution). Three hundred wells (from 30 towns across Central, Greater Accra and Western Regions) were sampled and analysed.

To predict the groundwater quality of the study regions, 80 percent of the secondary dataset was used to train the model and tested with 20 percent of the secondary data and 100 percent of the field dataset. The findings were displayed utilising descriptive statistics (standard deviation, mean, minimum, maximum and skewness). ANOVA) and PCA were used to analyse the primary data. RMSE and regression analyses (R^2) and d-statistics were used to analyse research objectives 1 and 2. Research objective 3 was analysed using ANOVA. The performance of the SML was compared to aqueous geochemical models (Visual Minteq, Phreeq C, and Wateq4F) using RMSE, d-statistics and R^2. This helped to establish the ability of ML model in predicting groundwater quality using easily measured parameters.

Key Findings (Based on Field Data Analysis)

The mean dissolved oxygen (7.20 mg/l), pH (6.82), temperature (28.40 ^oC) and electrical conductivity (426.51 mg/l) fell within the WHO recommended values (>6.5-8 mg/l, 6.5-8.5, 23.6 ^oC-33.10 ^oC and < 500 mS/cm, respectively) for safe quality drinking water. The mean total dissolved substances (453.77 mg/l) and turbidity (8.11 NTU) exceeded the WHO limits (< 300 mg/l and < 5 NTU, respectively) for excellent quality drinking water. Moreover, cations such as sodium (242.44 mg/l), calcium (171.76 mg/l) and magnesium (52.27 mg/l) exceeded the WHO recommended limits (200 mg/l, 100 mg/l and 50.00 mg/l), respectively), while potassium (124.92 mg/l) fell below the WHO recommended value (300 mg/l) for safe quality drinking water.

Anions such as sulphate (82.14 mg/l), chloride (147.72 mg/l), silicate (48.69 mg/l) and nitrate ions (42.38 mg/l) fell below the WHO recommended values (250 mg/l, 250 mg/l, 100 mg/l and 50 mg/l, respectively) for safe quality drinking water, while mean phosphate (3.17 mg/l) and hydrogen bicarbonate (10.91 mg/l) ions concentration exceeded the WHO recommended values (2 mg/l and 10 mg/l, respectively). The mean range of all the groundwater physicochemical parameters indicated that some of the wells sampled were not safe for drinking, particularly those having concentrations higher than the WHO recommended values.

From ANOVA analysis, the p-value (0.00) indicated a significant difference between and among the groundwater physicochemical parameters.

PCA indicated that groundwater data is an eight-component system, explaining 61.46 of the overall variation in the dataset. Based on the PCA, eight components loaded onto the groundwater quality, thus, affecting the groundwater quality.

From the Wilcox and Gibbs diagrams, the dominant factors that influence the groundwater in the study regions are precipitation, rock and evaporation. Most of the groundwater sampled from Central Region concentrated under precipitation and rock weathering dominance regions, while few of them were under evaporation dominance regions. The wells sampled from the Western Region concentrated under precipitation, rock weathering and evaporation dominance regions.

The mean concentration of dissolved oxygen in the Western Region was the highest compared to Central and Greater Accra Regions. The mean pH value of the groundwater sampled from the Greater Accra Region was the lowest compared to Central and Western Regions. The mean temperature value of Western Region was the lowest compared to Central and Greater Accra Regions. Greater Accra Regions recorded the highest total hardness and total dissolved substances, followed by Central and Western Regions. Again, Western Region recorded the highest turbidity and then Central and Greater Accra Regions. Central Region recorded the highest electrical conductivity compared to Western Region, making Greater Accra Region the least electrical conductivity recording region.

The sodium ion concentration sampled from the Western Region was the highest, followed by Greater Accra and Central Regions. Greater Accra Region featured the highest potassium ion concentration and then Western and

Central Regions, respectively. The mean magnesium of Western Region was the lowest compared to Central and Greater Accra Regions. Western Region recorded the highest sulphate ion concentration, followed by Central and Greater Accra Regions. The mean phosphate ion concentration sampled from the Western Region was the highest compared to Greater Accra and Central Regions. The mean chlorine ionic concentration of the groundwater sampled from Greater Accra Region was the highest compared to Central and Western Regions. Greater Accra Region featured the highest hydrogen bicarbonate ion concentration, followed by Central and Western Regions. The mean silicate ion concentration of Western Region was the highest compared to Central and Greater Accra Regions. Western Region recorded the highest nitrate ion concentration, followed by Central and Greater Accra Regions.

Key Findings (Based on Model Predicted Data Analysis)

The model's prediction accuracy of calcium ion (0.96) was the highest, followed by sulphate (0.95), potassium (0.92), magnesium (0.90), nitrate (0.85), phosphate (0.84), sodium (0.83), silicate (0.82), hydrogen bicarbonate (0.81), and chloride (0.80) ions. Based on the accuracy of the model (0.96), the supervised algorithm could be used to model and predict groundwater quality.

The predicted cationic sodium (244 mg/l) and calcium (172 mg/l) concentrations were higher than the observed concentrations (242.44 mg/l and 171.76 mg/l, respectively. The observed potassium (124.92 mg/l) and magnesium (52.27 mg/l) concentrations were less than the model predicted values (121 mg/l and 53 mg/l, respectively). Similarly, the predicted anions concentrations such as chlorine (147 mg/l), sulphate (82 mg/l) and silicate

(48mg/l) ions were less than the observed values (147.72 mg/l, 82.14 mg/l and 48.69 mg/l), while predicted phosphate (4 mg/l), hydrogen bicarbonate (12 mg/l) and nitrate (44 mg/l) concentrations were higher than the observed concentration (3.97 mg/l, 10.91 mg/l and 42.83 mg/l, respectively). This satisfied the objective that assessed the model's ability in predicting groundwater chemical parameters of the study regions using easily measured parameters.

From the regression analysis, model determinants such as R^2 were > 95 percent, demonstrating that the model could be used to predict the groundwater chemical parameters based on the easily measured parameters at a very high accuracy. DTR and polynomial algorithms could be used to predict groundwater chemical parameters with high accuracy using easily measured parameters for each of the regions.

Based on the RMSE, the model performance for sulphate and calcium ions was the best, followed by chloride, magnesium, phosphate and hydrogen bicarbonate ions. The model performed poorly for chloride ions compared to the other ions. In general, the model had RMSE of 1.690, showing the model's ability to forecast groundwater quality parameters with great accuracy and precision.

The difference between the observed (242.44 mg/l) and predicted sodium concentration by Visual Minteq (241.85 mg/l) and Wateq4F (241.85 mg/l) compared to Phreeq C (331.05 mg/l) and ML model (244.00 mg/l) was minimal. There was a significant difference between the predicted calcium (172.00 mg/l), magnesium (53.00 mg/l), chloride (147.00 mg/l), sulphate (82.00 mg/l) and silicate ion (48.00 mg/l) concentrations by ML model and observed values (242.44 mg/l, 52.27 mg/l, 147.72 mg/l, 82.14 mg/l, 48.69 mg/l, respectively) compare to that of Visual Minteq, Phreeq C and Wateq4F.

ML model had the best predictive accuracy compared to Visual Minteq, Phreeq C, and Wateq4F, with the lowest RMSE of 1.69, the highest R^2 value of 0.999, and a perfect d-statistic of 1.00. Western Region had the highest WQI at 111.56, followed by Greater Accra (94.75) and Central (87.32) Regions.

Conclusions

ML has a high level of accuracy and performance in predicting groundwater chemical parameters using easily measured parameters. Using ML model for predicting groundwater quality could provide a cost-effective and time-efficient alternative to traditional methods.

Compared to other commonly used models like Wateq4F, Visual Minteq, and Phreeq C, ML model had the highest level of accuracy. This indicates that the ML model could be a valuable tool for predicting groundwater quality on a wide scale.

ANOVA analysis showed significant differences between the observed and predicted groundwater chemical parameters, demonstrating that ML model could be used for identifying areas with potential water quality issues.

Recommendations

With reference to the findings of this study, the following recommendations are made.

1. The study demonstrated that the supervised learning algorithm can be used to predict groundwater quality parameters with high accuracy using easily measured parameters. Artificial intelligence tools, such as

supervised learning as an easy, time and cost-effective way of modelling and predicting water quality should be encouraged in groundwater assessment, especially for regions with limited resources.

2. The study found that 23% of the groundwater in the study areas failed to meet the WHO recommended values for excellent safe drinking water. This highlights the need for government and non-governmental organisations to work towards improving groundwater quality to ensure safe drinking water for the population. Moreover, the Water Resource Commission Act as well as policies on water and sanitation of groundwater ecosystems should be enforced to protect groundwater from anthropogenic activities that deteriorate water quality.
3. The water quality index (WQI) of the Western Region was found to be the highest among the three regions studied. Government and other relevant stakeholders should pay more attention to this region and take measures to improve the groundwater quality. That is, there should be an adequate monitoring programme for sustaining groundwater and sound mining activities in the regions. Hence, sustaining groundwater quality in the regions is as protecting health, sustaining water supply systems and meeting the water demands of the population growth in the regions.
4. The study used several models to predict groundwater quality and the machine learning model developed had the least RMSE and the highest prediction and performance. However, there are other machine learning methods such as unsupervised machine learning, therefore,

further research could be conducted to compare the performance of unsupervised in predicting groundwater quality parameters.

Practical Implications

The developed ML could help predict groundwater chemical parameters using easily measured parameters. This could help save time and cost compared to traditional laboratory methods. The high accuracy of the model could be useful in identifying areas with poor water quality, enabling targeted interventions to improve water quality. The model could help monitor and manage groundwater resources, especially in areas where water quality is a critical issue.

Policy Implications

The use of the developed Machine Learning model can inform policy decisions and management strategies related to groundwater resources. The model's ability to predict groundwater quality accurately and quickly can help in setting regulations and guidelines related to groundwater quality standards. The significant differences between the observed and predicted groundwater chemical parameters could help identify areas with poor water quality, informing policymakers to take necessary actions. The superior performance of the machine learning model compared to traditional models suggests that policymakers could consider adopting this model in groundwater quality prediction applications, providing more accurate and reliable results.

Contribution to Knowledge

Model performance determinants such as RMSE (1.690) and R-square (R-square values were > 95 percent) demonstrated that the model had high performance, and could be used to predict groundwater chemical parameters based on easily measured parameters.

The model predicting determinants such as accuracies (0.96) showed that the model could be used in predicting groundwater chemical parameters with high accuracy using the easily measured parameters. Moreover, the regression analysis showed that the model could be used to predict groundwater quality in an easy, cost and time-effective manner. This is held by the predictive model theory that groundwater quality could be assessed and predicted with reduced cost and time.

Analysis of variance (ANOVA) test demonstrated that there was a significant difference between and among the groundwater physicochemical parameters (p-value < 0.05). Therefore, there are significant differences between and among means of the observed and predicted groundwater chemical parameters of the study regions.

The machine learning model had the highest prediction and performance with an estimated RMSE of 1.69, followed by the Wateq4F (15.33), Visual Minteq (16.97) and Phreeq C (33.16). Based on the RMSE (1.690), d-statistic (0.791) and R^2 (0.999) estimated, ML model has the potential to predict groundwater quality with high accuracy compared to Visual Minteq, Phreeq C and Wateq4F.

Suggestions for Future Research

This study employed a supervised learning algorithm (SLA) such as DTR and PR approaches to predict the groundwater quality of the study regions (Central, Greater Accra and Western). However, groundwater pollution is not only limited to the study regions, therefore, future studies should consider several regions to investigate the spatiotemporal connection between surface and underground water quality environment.

Further studies can focus on elevating the geological underlying materials of the groundwater for proper management of the groundwater resources using artificial intelligence. Future studies can also assess artificial intelligence tools such as unsupervised learning and deep learning to investigate the spatiotemporal connection between surface and underground water quality.

REFERENCES

Abanyie, S. K., Sunkari, E. D., Apea, O. B., Abagale, S., & Korboe, H. M. (2020). Assessment of the quality of water resources in the Upper East Region, Ghana: a review. *Sustainable Water Resources Management*, *6*(4), 52.

Acheampong, R. A., Agyemang, F. S., & Abdul-Fatawu, M. (2017). Quantifying the spatiotemporal patterns of settlement growth in a metropolitan region of Ghana. *GeoJournal*, *82*(4), 823-840.

Adesanya, K. K., Taiwo, A. I., Adedodun, A. F., & Olatayo, T. O. (2018). Modelling continuous non-linear data with lagged fractional polynomial regression. *Asian Journal of Applied Sciences*, *6*(5).

Adeyemi, F. M., Wahab, A. A., Oyelami, C. A., Oyedara, O. O., Titilawo, M. A., Adebunmi, A. A., & Awoniyi, I. O. (2022). Hydrology survey and water quality assessment of water sources in three selected towns in Osun State, Southwest Nigeria. *International Journal of Energy and Water Resources*, 1-14.

Adjei, V., & Adjokatse, E. F. A. I. T. (2022). The Evil Couple: Illegal Mining in Water Bodies and Climate Change: A Case Study of Ghana. *Journal of Atmospheric Science Research| Volume*, *5*(03).

Agbasi, J. C., & Egbueri, J. C. (2022). Assessment of PTEs in water resources by integrating HHRISK code, water quality indices, multivariate statistics and ANNs. Geocarto International, (just-accepted), 1-24.

Aggarwal, K., Mijwil, M. M., Al-Mistarehi, A. H., Alomari, S., Gök, M., Alaabdin, A. M. Z., & Abdulrhman, S. H. (2022). Has the Future Started? The Current Growth of Artificial Intelligence, Machine

Learning, and Deep Learning. *Iraqi Journal For Computer Science and Mathematics*, 3(1), 115-123.

Agrawal, P., Sinha, A., Kumar, S., Agarwal, A., Banerjee, A., Villuri, V. G. K., & Pasupuleti, S. (2021). Exploring artificial intelligence techniques for groundwater quality assessment. *Water*, 13(9), 1172.

Ahamed, K. I., & Akthar, S. (2016). A study on neural network architectures. *Computer Engineering and Intelligent Systems*, *7*(9), 2222-1719.

Aithal, P. S. (2023). Super-Intelligent Machines-analysis of developmental challenges and predicted negative consequences. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *7*(3), 109-141.

Aithal, P. S., & Aithal, S. (2023). The changing role of higher education in the era of AI-based GPTs. *International Journal of Case Studies in Business, IT, and Education (IJCSBE)*, *7*(2), 183-197.

Akhtar, N., Ishak, M. I. S., Ahmad, M. I., Umar, K., Md Yusuff, M. S., Anees, M. T., & Ali Almanasir, Y. K. (2021). Modification of the water quality index (WQI) process for simple calculation using the multi-criteria decision-making (MCDM) method: a review. *Water*, 13(7), 905.

Akhtar, N., Syakir Ishak, M. I., Bhawani, S. A., & Umar, K. (2021). Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water*, *13*(19), 2660.

Alavizadeh, H., Alavizadeh, H., & Jang-Jaccard, J. (2022). Deep Q-Learning based Reinforcement Learning Approach for Network *Intrusion Detection. Computers*, 11(3), 41.

Albert, K. M. (2015). Role of revegetation in restoring fertility of degraded mined soils in Ghana: A review. *International journal of biodiversity and conservation*, 7(2), 57-80.

Aldhyani, T. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics*, pg. 1-20.

Alfarrah, N., & Walraevens, K. (2018). Groundwater overexploitation and seawater intrusion in coastal areas of arid and semi-arid regions. *Water*, 10(2), 143.

Alldritt, B., Urtev, F., Oinonen, N., Aapro, M., Kannala, J., Liljeroth, P., & Foster, A. S. (2022). Automated tip functionalization via machine learning in scanning probe microscopy. *Computer Physics Communications*, 273, 108258.

Alvarez-Bastida, C., Martínez-Miranda, V., Solache-Ríos, M., Linares-Hernández, I., Teutli-Sequeira, A., & Vázquez-Mejía, G. (2018). Drinking water characterization and removal of manganese. Removal of manganese from water. *Journal of Environmental Chemical Engineering*, 6(2), 2119-2125.

Amarappa, S., & Sathyanarayana, S. V. (2014). Data classification using Support vector Machine (SVM), a simplified approach. *Int. J. Electron. Comput. Sci. Eng*, *3*, 435-445.

Amiri, Z., Heidari, A., Navimipour, N. J., Unal, M., & Mousavi, A. (2023). Adventures in data analysis: A systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. *Multimedia Tools and Applications*, 1-65.

Ampim, P. A., Ogbe, M., Obeng, E., Akley, E. K., & MacCarthy, D. S. (2021). Land cover changes in Ghana over the past 24 years. *Sustainability*, *13*(9), 4951.

Amuah, E. E. Y., Boadu, J. A., & Nandomah, S. (2022). Emerging issues and approaches to protecting and sustaining surface and groundwater resources: Emphasis on Ghana. *Groundwater for Sustainable Development*, 16, 100705.

Apley, D. W., & Zhu, J. (2020). Visualising the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 1059-1086.

Arabahmadi, M., Farahbakhsh, R., & Rezazadeh, J. (2022). Deep learning for smart Healthcare—A survey on brain tumor detection from medical imaging. *Sensors*, *22*(5), 1960.

Archontoulis, S. V., & Miguez, F. E. (2015). Nonlinear regression models and applications in agricultural research. *Agronomy Journal*, *107*(2), 786-798.

Asomaning, J., Antwi, E. O., Laar, C., & Saka, D. (2023). Statistical and isotopic analysis of sources and evolution of groundwater. *Physics and Chemistry of the Earth, Parts A/B/C*, *129*, 103337.

Assunção, A. G., Cakmak, I., Clemens, S., González-Guerrero, M., Nawrocki, A., & Thomine, S. (2022). Micronutrient homeostasis in plants for more sustainable agriculture and healthier human nutrition. *Journal of Experimental Botany*, 73(6), 1789-1799.

Austin, J., Drossaert, C. H., van Dijk, J., Sanderman, R., Børøsund, E., Mirkovic, J., & Bohlmeijer, E. T. (2022). Integrating top-down and bottom-up requirements in eHealth development: A case study of a mobile self-compassion intervention for people with newly diagnosed cancer. *JMIR cancer*. 3(6), 89-199.

Azamat, A. (2021). Trans-ai: How to build true AI or real machine intelligence and learning. *Онтология проектирования*, *11*(4 (42)), 402-421.

Bahlol, H. A. E., Kamel, M., Seleem, E. M., Mostafa, A., & Salman, S. A. (2023). Spatial distribution and chemical speciation of some potentially toxic elements in Qena Governorate groundwater, Egypt. *Egyptian Journal of Chemistry*, *66*(11), 137-150.

Balasubramanian, M., Sridhar, S. G. D., Ayyamperumal, R., Karuppannan, S., Gopalakrishnan, G., Chakraborty, M., & Huang, X. (2022). Isotopic signatures, hydrochemical and multivariate statistical analysis of seawater intrusion in the coastal aquifers of Chennai and Tiruvallur District, Tamil Nadu, India. *Marine Pollution Bulletin*, 174, 113232.

Banda, T. D., & Kumarasamy, M. V. (2020). Development of Water Quality Indices (WQIs): A Review. *Polish Journal of Environmental Studies*, 29(3).

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.

Becher, J., Englisch, C., Griebler, C., & Bayer, P. (2022). Groundwater fauna downtown–Drivers, impacts and implications for subsurface ecosystems in urban areas. *Journal of Contaminant Hydrology*, 248, 104021.

Bekhit, A. F. (2022). Computer Vision Applications. In *Computer Vision and Augmented Reality in iOS.* Apress, Berkeley, *CA*. (pp. 49-84).

Bera, D., Chatterjee, N. D., & Bera, S. (2021). Comparative performance of linear regression, polynomial regression and generalized additive model for canopy cover estimation in the dry deciduous forest of West Bengal. *Remote Sensing Applications: Society and Environment*, *22*, 100502.

Bilbao, I., & Bilbao, J. (2017, December). Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. In *2017 eighth international conference on intelligent computing and information systems (ICICIS)* (pp. 173-177). IEEE.

Boakye, A. K., Asante, B. O., Akutinga, S., Agyeman, D., Prah, S., Idan, F., & Asante, I. S. (2023). Farmers' perceptions and willingness to adopt egg and snail shells as alternative calcium sources for poultry production in the Western Region of Ghana. *Journal of Agriculture and Food Research*, *14*, 100818.

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *Journal of Data Analysis and Information Processing*, *8*(4), 341-357.

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: how might the New Zealand government respond?. *Policy Quarterly*, *13*(4).

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, *113*(15), 3932-3937.

Ćalasan, M., Aleem, S. H. A., & Zobaa, A. F. (2020). On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function. *Energy conversion and management*, *210*, 112716.

Calvert, M. B. (2020). *Predicting Concentrations of Selected Ions and Total Hardness in Groundwater Using Artificial Neural Networks and Multiple Linear Regression Models* (Doctoral dissertation, Duke University). *9*(44), e11881119.

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management*, 21(3-4), 193-215.

Cheng, X., Hou, C., Gao, H., Li, P., Zhu, X., Luo, C., & Liang, H. (2022). Synergistic process using calcium peroxide and ferrous iron for enhanced ultrafiltration of Microcystis aeruginosa-laden water. *Water Research*, 211, 118067.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*, *7*, e623.

Chong, B. W., Othman, R., Putra Jaya, R., Mohd Hasan, M. R., Sandu, A. V., Nabiałek, M., & Abdullah, M. M. A. B. (2021). Design of experiment on concrete mechanical properties prediction: a critical review. *Materials*, *14*(8), 1866.

Chycki, J., Zając, T., Maszczyk, A., & Kurylas, A. (2017). The effect of mineral-based alkaline water on hydration status and the metabolic response to short-term anaerobic exercise. *Biology of Sport*, 34(3), 255-261.

Ciccozzi, F., Addazi, L., Asadollah, S. A., Lisper, B., Masud, A. N., & Mubeen, S. (2022). A Comprehensive Exploration of Languages for Parallel Computing. *ACM Computing Surveys* (CSUR), 55(2), 1-39.

Corno, G., Ghaly, T., Sabatino, R., Eckert, E. M., Galafassi, S., Gillings, M. R., & Di Cesare, A. (2023). Class 1 integron and related antimicrobial resistance gene dynamics along a complex freshwater system affected by different anthropogenic pressures. *Environmental Pollution*, *316*, 120601.

Corona, B., Shen, L., Reike, D., Carreón, J. R., & Worrell, E. (2019). Towards sustainable development through the circular economy—A review and critical assessment on current circularity metrics. *Resources, Conservation and Recycling*, *151*, 104498.

Coronado, E., Kiyokawa, T., Ricardez, G. A. G., Ramirez-Alpizar, I. G., Venture, G., & Yamanobe, N. (2022). Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0. *Journal of Manufacturing Systems*, *63*, 392-410.

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*(3), 339-362.

Daniyal, M., Tawiah, K., Qureshi, M., Haseeb, M., Asosega, K. A., Kamal, M., & Rehman, M. U. (2023). An autoregressive distributed lag approach for estimating the nexus between CO2 emissions and economic determinants in Pakistan. *Plos one*, *18*(5), e0285854.

Davand, H., Sepehr, E., Momtaz, H. R., & Ahmadi, F. (2022). Wastewater irrigation: An opportunity for improving soil phosphorus availability; PHREEQC modelling and adsorption studies. *Science of The Total Environment*, 851, 158180.

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760-772.

De Keyser, A., & Kunz, W. H. (2022). Living and working with service robots: a TCCM analysis and considerations for future research. *Journal of Service Management*. 8(5), 80.

Denise, N. F. T., Tiafack, O., & Siméon, T. (2022). Potable Water Supply Deficiency in Yaounde (Centre Cameroon): Challenges and Coping Strategies of the Inhabitants. *Saudi J. Humanities Soc Sci*, 7(9), 398-410.

Dey, S. K., Kumar, S., Rani, D., Maurya, S. K., Banerjee, P., Verma, M., & Senapati, S. (2023). Implications of vitamin D deficiency in systemic inflammation and cardiovascular health. *Critical reviews in food science and nutrition*, 1-18.

Dinesh Kumar, M., Bassi, N., & Kumar, S. (2022). Managing Groundwater Quality for Drinking Water Security in India: Emerging Challenge. In Drinking Water Security in Rural India: Dynamics, Influencing Factors, and Improvement Strategy. Singapore: *Springer Singapore*. (pp. 111-130).

Dippong Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., & Gruszczyńska, K. (2023). What is machine learning, artificial neural networks and deep learning? —Examples of practical applications in medicine. *Diagnostics*, *13*(15), 2582.

Dobrzyński, D. (2005). Silica origin and solubility in groundwater from the weathered zone of sedimentary rocks of the Intra-Sudetic Basin, SW Poland. *Acta Geologica Polonica*, 55(4), 445-462.

Ebrahim, O. A., & Derbew, G. (2023). Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in Afar regional state, Northeastern Ethiopia 2021. *Scientific reports*, *13*(1), 7779.

Egbueri, J. C., & Agbasi, J. C. (2022). Data-driven soft computing modelling of groundwater quality parameters in southeast Nigeria: comparing the performances of different algorithms. *Environmental Science and Pollution Research*, 1-28.

El Bilali, A., Taleb, A., & Brouziyne, Y. (2021). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, 245, 106625.

El-Aziz, A., & Hassanien, S. (2017). Evaluation of groundwater quality for drinking and irrigation purposes in the north-western area of Libya (Aligeelat). *Environmental Earth Sciences*, 76(4), 1-17.

Elbeltagi, A., Pande, C. B., Kouadri, S., & Islam, A. R. M. (2021). Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India. *Environmental Science and Pollution Research*, 1-15.

Eluyode, O. S., & Akomolafe, D. T. (2013). Comparative study of biological and artificial neural networks. *European Journal of Applied Engineering and Scientific Research*, *2*(1), 36-46.

Feuerriegel, G. C., Ritschl, L. M., Sollmann, N., Palla, B., Leonhardt, Y., Maier, L., & Burian, E. (2023). Imaging of traumatic mandibular fractures in young adults using CT-like MRI: a feasibility study. *Clinical Oral Investigations*, *27*(3), 1227-1233.

Foppen, J. W., Lutterodt, G., Rau, G. C., & Minkah, O. (2020). Groundwater flow system analysis in the regolith of Dodowa on the Accra Plains, Ghana. *Journal of Hydrology: Regional Studies*, 28, 100663.

Forest, J. J., Howard, R. D., & Sheehan, A. M. (2013). Weapons of mass destruction and terrorism. *New York: McGraw-Hill*. 10(6), 829.

Frees, E. W., Derrig, R. A., & Meyers, G. (Eds.). (2014). *Predictive modelling applications in actuarial science* (Vol. 1). Cambridge University Press.

Galloway, A. R. (2022). Golden Age of Analog. *Critical Inquiry*, 48(2), 211-232.

Ganesha, H. R., & Aithal, P. S. (2022). Deriving Right Sample Size and Choosing an Appropriate Sampling Technique to Select Samples from the Research Population During Ph. D. Program in India. 211-232.

Gautam, U., Tiwari, V., & Tripathi, V. K. (2022). Evaluation of groundwater quality of Prayagraj city using entropy water quality index (EWQI) and new integrated water quality index (IWQI). *Sustainable Water Resources Management*, 8(2), 1-18.

Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, *5*(1), 1-48.

Goncalves, S., Fong, P. C., & Blokhina, M. (2022). Artificial intelligence for early diagnosis of lung cancer through incidental nodule detection in low-and middle-income countries-acceleration during the COVID-19 pandemic but here to stay. *American Journal of Cancer Research*, 12(1), 1.

Guan, X., Liu, C., Li, Y., Wang, X., Liu, Y., Zou, C., & Zhang, W. (2022). Reducing the environmental risks related to phosphorus surplus resulting from greenhouse cucumber production in China. *Journal of Cleaner Production*, 332, 130076.

Guerard Jr, J. B., Saxena, A., & Gültekin, M. N. (2022). Regression Analysis and Estimating Regression Models. In *Quantitative Corporate Finance* (pp. 291-330). Cham: Springer International Publishing.

Hahad, O., Kröller-Schön, S., Daiber, A., & Münzel, T. (2019). The cardiovascular effects of noise. Deutsches Ärzteblatt International, 116(14), 245.

Hanoon, M. S., Ahmed, A. N., Fai, C. M., Birima, A. H., Razzaq, A., Sherif, M., & El-Shafie, A. (2021). Application of Artificial Intelligence Models for modelling Water Quality in Groundwater: Comprehensive Review, Evaluation and Future Trends. *Water, Air, & Soil Pollution*, 232(10), 1-41.

Haritash, A. K., Gaur, S., & Garg, S. (2016). Assessment of water quality and suitability analysis of River Ganga in Rishikesh, India. *Applied Water Science*, 6(4), 383-392.

Harris, H. (2022). Artificial intelligence and policing of financial crime: a legal analysis of the state of the field. *Financial Technology and the Law: Combating Financial Crime*, 281-299.

Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). Artificial intelligence in surgery: promises and perils. *Annals of surgery*, *268*(1), 70-76.

Hasnaoui, F. M., Bouhelal, A., Smaili, A., & Ramzan, N. (2024). A Neural Network-Based Approach for Predicting Thrust and Power Coefficients in Drone Propeller Performances. In *AIAA SCITECH 2024 Forum* (p. 2688).

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1-10.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.

Hutchins, M. G., Harding, G., Jarvie, H. P., Marsh, T. J., Bowes, M. J., & Loewenthal, M. (2020). Intense summer floods may induce prolonged increases in benthic respiration rates of more than one year leading to low river dissolved oxygen. *Journal of Hydrology X*, 8, 100056.

Ibrahim, I., & Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10-19.

Ibrahim, K. S. M. H., Huang, Y. F., Ahmed, A. N., Koo, C. H., & El-Shafie, A. (2022). A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alexandria Engineering Journal*, 61(1), 279-303.

Ibrahim, M. N. (2019). Assessing groundwater quality for drinking purpose in Jordan: application of water quality index. *Journal of Ecological Engineering*, 20(3).

Islas-Cota, E., Gutierrez-Garcia, J. O., Acosta, C. O., & Rodríguez, L. F. (2022). A systematic review of intelligent assistants. *Future Generation Computer Systems*, 128, 45-62.

Jain, C. K., & Singh, S. (2020). Impact of climate change on the hydrological dynamics of River Ganga, India. *Journal of Water and Climate change*, *11*(1), 274-290.

Jeong, H. Y., Jun, S. C., Cheon, J. Y., & Park, M. (2018). A review on clogging mechanisms and managements in aquifer storage and recovery (ASR) applications. *Geosciences Journal*, *22*, 667-679.

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence? *Discover Artificial Intelligence*, *2*(1), 4.

Jollivet, P., Gin, S., & Schumacher, S. (2012). Forward dissolution rate of silicate glasses of nuclear interest in clay-equilibrated groundwater. *Chemical Geology*, 330, 207-217.

Joshi, N., & Srivastava, S. (2014). Improving classification accuracy using ensemble learning technique (using different decision trees). *Int. J. Comput. Sci. Mob. Comput*, *3*(5), 727-732.

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609-629.

Kasula, B. Y. (2019). Enhancing Classification Precision: Exploring the Power of Support-Vector Networks in Machine Learning. *International Scientific Journal for Research*, *1*(1).

Kelly, S. J., Derrington, S., & Star, S. (2022). Governance challenges in esports: A best practice framework for addressing integrity and wellbeing issues. *International Journal of Sport Policy and Politics*, 14(1), 151-168.

Kerby, M. B. (2015). Toward a new predictive model of student retention in higher education: An application of classical sociological theory. *Journal of College Student Retention: Research, Theory & Practice*, *17*(2), 138-161.

Khalid, S., Shahid, M., Alothman, Z. A., Al-Kahtani, A. A., Murtaza, B., & Dumat, C. (2023). Predicting chemical speciation of metals in soil using Visual Minteq. Soil Ecology Letters, 5(3), 220162.

Khan, M. N., Mobin, M., Abbas, Z. K., & Alamri, S. A. (2018). Fertilizers and their contaminants in soils, surface and groundwater. *Encyclopedia of the Anthropocene*, *5*, 225-240.

Khatibi Bardsiri, M., & Eftekhari, M. (2014). Comparing ensemble learning methods based on decision tree classifiers for protein fold recognition. *International journal of data mining and bioinformatics*, *9*(1), 89-105.

Kortei, N. K., Heymann, M. E., Essuman, E. K., Kpodo, F. M., Akonor, P. T., Lokpo, S. Y., & Tettey, C. (2020). Health risk assessment and levels of toxic metals in fishes (Oreochromis noliticus and Clarias anguillaris) from Ankobrah and Pra basins: Impact of illegal mining activities on food safety. *Toxicology Reports*, 7, 360-369.

Korteling, J. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human-versus artificial intelligence. *Frontiers in artificial intelligence*, *4*, 622364.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24.

Kouadri, S., Elbeltagi, A., Islam, A. R. M., & Kateb, S. (2021). Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Applied Water Science*, 11(12), 1-20.

Kritikos, M., Concepcion Maure, L., Leyva Cespedes, A. A., Delgado Sobrino, D. R., & Hrušecký, R. (2019). A random factorial design of experiments study on the influence of key factors and their interactions on the measurement uncertainty: a case study using the ZEISS CenterMax. *Applied Sciences*, *10*(1), 37.

Krok, B., Mohammadian, S., Noll, H. M., Surau, C., Markwort, S., Fritzsche, A., & Meckenstock, R. U. (2022). Remediation of zinc-contaminated groundwater by iron oxide in situ adsorption barriers–From lab to the field. Science of the Total Environment, 807, 151066.

Kuusi, O., & Heinonen, S. (2022). Scenarios from artificial narrow intelligence to artificial general intelligence—Reviewing the results of the international work/technology 2050 study. *World Futures Review*, *14*(1), 65-79.

Larsen Greiner, T. A., Lie, J. E., Kolbjørnsen, O., Kjelsrud Evensen, A., Harris Nilsen, E., Zhao, H., & Gelius, L. J. (2022). Unsupervised deep learning with higher-order total-variation regularization for multidimensional seismic data reconstruction. *Geophysics*, *87*(2), V59-V73.

Lawniczak, A. E., Zbierska, J., Nowak, B., Achtenberg, K., Grześkowiak, A., & Kanas, K. (2016). Impact of agriculture and land use on nitrate contamination in groundwater and running waters in central-west Poland. *Environmental monitoring and assessment*, 188(3), 1-17.

Li, P., Karunanidhi, D., Subramani, T., & Srinivasamoorthy, K. (2021). Sources and consequences of groundwater contamination. *Archives of environmental contamination and toxicology*, 80(1), 1-10.

Livingston, J. (2021). Water scarcity & health in urban Africa. *Dædalus*, 150(4), 85-102.

Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), 1-29.

Lutterodt, G., Van de Vossenberg, J., Hoiting, Y., Kamara, A. K., Oduro-Kwarteng, S., & Foppen, J. W. A. (2018). Microbial groundwater quality status of hand-dug wells and boreholes in the Dodowa area of Ghana. *International Journal of Environmental Research and Public Health*, 15(4), 730.

Machta, B. B., Chachra, R., Transtrum, M. K., & Sethna, J. P. (2013). Parameter space compression underlies emergent theories and predictive models. *Science*, *342*(6158), 604-607.

Mahapatra, S. K., Das, A., & Narasimhan, R. (2012). A contingent theory of supplier management initiatives: effects of competitive intensity and product life cycle. *Journal of Operations Management*, *30*(5), 406-422.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, *21*(2), 584-595.

Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, *1*(2), 140-147.

Mayowa, D. O., & Olajide, I. A. (2024). Design of a Voice Recognition System Using Artificial Neural Network. *International Journal of Electrical and Computer Engineering Research*, *4*(1), 1-7.

McCaig, M. (2020). Soil properties and subsurface hydrology. In Geomorphology and Soils. *Routledge*. (pp. 121-140).

Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57, 115-129.

Mensah, D. O., Appiah-Adjei, E. K., & Asante, D. (2023). Groundwater pollution vulnerability assessment in the Assin municipalities of Ghana using GIS-based DRASTIC and SINTACS methods. *Modelling Earth Systems and Environment*, *9*(2), 2955-2967.

Messaoud, S., Bradai, A., Bukhari, S. H. R., Quang, P. T. A., Ahmed, O. B., & Atri, M. (2020). A survey on machine learning in Internet of Things: Algorithms, strategies, and applications. *Internet of Things*, *12*, 100314.

Mettu, A., & Latifi, S. (2020, August). Assessment of Groundwater Quality and Quantity Using Supervised Machine Learning. *In International Conference on Systems Engineering. Springer, Cham*. (pp. 160-171).

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, *55*(5), 3503-3568.

Mishra, R. K. (2023). Fresh water availability and its global challenge. *British Journal of Multidisciplinary and Advanced Studies*, *4*(3), 1-78.

Mohammed, K., Salifu, M. G., Batung, E., Amoak, D., Avoka, V. A., Kansanga, M., & Luginaah, I. (2022). Spatial analysis of climatic factors and plasmodium falciparum malaria prevalence among children in Ghana. *Spatial and Spatio-temporal Epidemiology*, *43*, 100537.

Mohapatra, J. B., Jha, P., Jha, M. K., & Biswal, S. (2021). Efficacy of machine learning techniques in predicting groundwater fluctuations in agroecological zones of India. *Science of The Total Environment*, *785*, 147319.

Mokarram, M., Pourghasemi, H. R., Huang, K., & Zhang, H. (2022). Investigation of water quality and its spatial distribution in the Kor River basin, Fars province, *Iran. Environmental Research*, 204, 112294.

Mukherjee, I., & Singh, U. K. (2022). Hydrogeochemical characterisations and quality evaluation of groundwater in the major river basins of a geologically and anthropogenically driven semi-arid tract of India. *Science of The Total Environment*, 805, 150323.

Naderi Peikam, E., & Jalali, M. (2016). Application of inverse geochemical modelling for predicting surface water chemistry in Ekbatan watershed, Hamedan, western Iran. *Hydrological Sciences Journal*, *61*(6), 1124-1134.

Najafzadeh, M., Homaei, F., & Mohamadi, S. (2022). Reliability evaluation of groundwater quality index using data-driven models. *Environmental Science and Pollution Research*, 29(6), 8174-8190.

Namasivayam, V., Senguttuvan, N., Saravanan, V., Palaniappan, S., & Kathiravan, M. K. (2022). Artificial Intelligence and Its Application in Cardiovascular Disease Management. In Machine Learning and Systems Biology in Genomics and Health (pp. 189-236). *Springer, Singapore*.

Newby, G. B. (2001). Cognitive space and information space. *Journal of the American Society for Information Science and Technology*, *52*(12), 1026-1048.

Nordstrom, D. K., & Campbell, K. M. (2014). Modelling low-temperature geochemical processes. *Treatise on geochemistry*, *7*, 27-68.

Nwankwo, C. B., Hoque, M. A., Islam, M. A., & Dewan, A. (2020). Groundwater constituents and trace elements in the basement aquifers of Africa and sedimentary aquifers of Asia: medical hydrogeology of drinking water minerals and toxicants. *Earth Systems and Environment*, *4*(2), 369-384.

Okwu, M. O., Tartibu, L. K., Okwu, M. O., & Tartibu, L. K. (2021). Artificial neural network. *Metaheuristic optimization: nature-inspired algorithms swarm and computational intelligence, theory and applications*, 133-145.

Oludare, F. V. (2017). Effects of weathering and erosion on the geochemistry of rocks and soils. *Int. J. Sci. Res. Sci. Technol*, *3*(6), 74-80.

Oluyisola, O. E., Bhalla, S., Sgarbossa, F., & Strandhagen, J. O. (2022). Designing and developing smart production planning and control systems in the industry 4.0 era: a methodology and case study. *Journal of Intelligent Manufacturing*, 33(1), 311-332.

Omer, N. H. (2019). Water quality parameters. *Water quality-science, assessments and policy*, 18, 1-34.

Onyutha, C. (2020). From R-squared to coefficient of model accuracy for assessing" goodness-of-fits". *Geoscientific Model Development Discussions*, *2020*, 1-25.

Orina, B. G. (2015). Experimental and Modelling Studies on the Interactions of Metals and Natural Organic Matter in Cooling Water at Coal Power-Generation Plants. *University of Johannesburg* (South Africa). 30(2), 39-249.

Panigrahi, B., Kathala, K. C. R., & Sujatha, M. (2023). A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. *Procedia Computer Science*, *218*, 2684-2693.

Parshotam, H. (2014). Modelling Studies of Interactions Between Natural Organic Matter and Metals: Effects on Cooling Water Precipitation Potentials in Power Generation Plants. *University of Johannesburg* (South Africa). 1(4), 1-18.

Pathak, S., Mishra, I., & Swetapadma, A. (2018). An assessment of decision tree-based classification and regression algorithms. In *2018 3rd International Conference on Inventive Computation Technologies (ICICT)* (pp. 92-95). IEEE.

Patil, P. N., Sawant, D. V., & Deshmukh, R. N. (2012). Physico-chemical parameters for testing of water-a review. *International journal of environmental sciences*, 3(3), 1194.

Piatka, D. R., Wild, R., Hartmann, J., Kaule, R., Kaule, L., Gilfedder, B., & Barth, J. A. (2021). Transfer and transformations of oxygen in rivers as catchment reflectors of continental landscapes: A review. *Earth-Science Reviews*, 220, 103729.

Pohl, H. R., Wheeler, J. S., & Murray, H. E. (2013). Sodium and potassium in health and disease. Interrelations between essential metal ions and human diseases, 29-47.

Prommer, H., Sun, J., & Kocar, B. D. (2019). Using reactive transport models to quantify and predict groundwater quality. *Elements: An International Magazine of Mineralogy, Geochemistry, and Petrology*, *15*(2), 87-92.

Raheja, H., Goel, A., & Pal, M. (2022). Prediction of groundwater quality indices using machine learning algorithms. *Water Practice & Technology*, 17(1), 336-351.

Raiber, M., White, P. A., Daughney, C. J., Tschritter, C., Davidson, P., & Bainbridge, S. E. (2012). Three-dimensional geological modelling and multivariate statistical analysis of water chemistry data to analyse and visualise aquifer structure and groundwater composition in the Wairau Plain, Marlborough District, New Zealand. *Journal of Hydrology*, *436*, 13-34.

Rakhimova, N. (2022). Calcium and/or magnesium carbonate and carbonate-bearing rocks in the development of alkali-activated cements–a review. *Construction and Building Materials*, 325, 126742.

Rana, A., Singhal, N., & Badotra, S. (2021). A review paper on artificial intelligence. ACADEMICIA: *An International Multidisciplinary Research Journal*, 11(10), 249-256.

Randall, S. M., Cardasis, H. L., & Muddiman, D. C. (2013). Factorial experimental designs elucidate significant variables affecting data acquisition on a quadrupole orbitrap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *24*(10), 1501-1512.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*.

Rathore, S. S., & Kumar, S. (2016). A decision tree regression based approach for the number of software faults prediction. *ACM SIGSOFT Software Engineering Notes*, *41*(1), 1-6.

Reigl, S., Van Driessche, A. E., Mehringer, J., Koltzenburg, S., Kunz, W., & Kellermeier, M. (2022). Revisiting the roles of salinity, temperature and water activity in phase selection during calcium sulfate precipitation. *CrystEngComm*, 24(8), 1529-1536.

Riedel, T. (2019). Temperature-associated changes in groundwater quality. *Journal of hydrology*, 572, 206-212.

Rights, J. D., & Sterba, S. K. (2023). R-squared measures for multilevel models with three or more levels. *Multivariate Behaviour al Research*, *58*(2), 340-367.

Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modelling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, *36*(5), 355-362.

Roy, B., Roy, S., Mitra, S., & Manna, A. K. (2021). Evaluation of groundwater quality in West Tripura, Northeast India, through combined application of water quality index and multivariate statistical techniques. *Arabian Journal of Geosciences*, 14(19), 1-18.

Rubinstein, R. A. (2022). Disaster and Health. In The SAGE Handbook of Social Studies in Health and Medicine, *SAGE Publications Ltd. Second Edition*. (pp. 562-588).

Sah, B., & Pandey, D. L. (2023). VAT elasticity on Nepalese economy without agriculture. *Nepalese Journal of Management Research*, *3*(1), 16-25.

Sahoo, P. K., Virk, H. S., Powell, M. A., Kumar, R., Pattanaik, J. K., Salomão, G. N., & Tiwari, R. P. (2022). Meta-analysis of uranium contamination in groundwater of the alluvial plains of Punjab, northwest India: Status, health risk, and hydrogeochemical processes. *Science of The Total Environment*, 807, 151753.

Sakizadeh, M. (2016). Artificial intelligence for the prediction of water quality index in groundwater systems. *Modelling Earth Systems and Environment*, 2(1), 1-9.

Sanganyado, E., & Gwenzi, W. (2019). Antibiotic resistance in drinking water systems: Occurrence, removal, and human health risks. *Science of the Total Environment*, 669, 785-797.

Sani, N. S., Abd Rahman, A. H., Adam, A., Shlash, I., & Aliff, M. (2020). Ensemble learning for rainfall prediction. *International Journal of Advanced Computer Science and Applications*, *11*(11).

Sarkar, B., Mitchell, E., Frisbie, S., Grigg, L., Adhikari, S., & Maskey Byanju, R. (2022). Drinking water quality and public health in the Kathmandu Valley, Nepal: coliform bacteria, chemical contaminants, and health status of consumers. *Journal of environmental and public health*, 2*(22), 1-7.*

Sarkar, M., Pal, S. C., & Islam, A. R. M. (2022). Groundwater quality assessment for safe drinking water and irrigation purposes in Malda district, Eastern India. *Environmental Earth Sciences*, 81(2), 1-20.

Scharre, P. (2019). Killer apps: The real dangers of an AI arms race. *Foreign Aff.*, 98, 135.

Schölkopf, B. (2022). Causality for machine learning. In Probabilistic and Causal Inference: *The Works of Judea Pearl* (pp. 765-804).

Şener, Ş., Şener, E., & Davraz, A. (2017). Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Science of the total Environment*, *584*, 131-144.

Serey, J., Alfaro, M., Fuertes, G., Vargas, M., Duran, C., Ternero, R., & Sabattin, J. (2023). Pattern recognition and deep learning technologies, enablers of industry 4.0, and their role in engineering research. *Symmetry*, *15*(2), 535.

Shah, V., Parashar, A., & Scott, A. (2022). Understanding the importance of carbonates on the performance of Portland metakaolin cement. *Construction and Building Materials*, 319, 126155.

Sharma, H., & Kumar, S. (2016). A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*, *5*(4), 2094-2097.

Sharma, M. K., & Kumar, M. (2020). Sulphate contamination in groundwater and its remediation: an overview. *Environmental monitoring and assessment*, 192(2), 1-10.

Sharma, S., Arrawatia, K. S., & Asthana, B. (2019). Transfer Learning: Leveraging Knowledge Across Domains in AI. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *10*(2), 1167-1170.

Sheikh, A. T., Mugera, A., Pandit, R., Burton, M., & Davies, S. (2022). What determines the time to gypsum adoption to remediate irrigated salt-affected agricultural lands? Evidence from Punjab, Pakistan. *Soil and Tillage Research*, 217, 105266.

Shiri, N., Shiri, J., Yaseen, Z. M., Kim, S., Chung, I. M., Nourani, V., & Zounemat-Kermani, M. (2021). Development of artificial intelligence models for well groundwater quality simulation: Different modelling scenarios. *Plos one*, 16(5), e0251510.

Shouval, R., Fein, J. A., Savani, B., Mohty, M., & Nagler, A. (2021). Machine learning and artificial intelligence in haematology. *British journal of haematology*, 192(2), 239-250.

Singh, M. R., & Gupta, A. (2016). Water pollution-sources, effects and control. Centre for Biodiversity, Department of Botany, Nagaland University.

Singh, S. K., Srivastava, P. K., Pandey, A. C., & Gautam, S. K. (2013). Integrated assessment of groundwater influenced by a confluence river system: concurrence with remote sensing and geochemical modelling. *Water resources management*, *27*, 4291-4313.

Singha, S. S., Singha, S., Pasupuleti, S., & Venkatesh, A. S. (2022). Knowledge-driven and machine learning decision tree-based approach for assessment of geospatial variation of groundwater quality around coal mining regions, Korba district, Central India. *Environmental Earth Sciences*, 81(2), 1-13.

Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., & Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, 276, 130265.

Singla, M., & Shukla, K. K. (2020). Robust statistics-based support vector machine and its variants: a survey. *Neural Computing and Applications*, *32*(15), 11173-11194.

Soleimani, H., Nasri, O., Ghoochani, M., Azhdarpoor, A., Dehghani, M., Radfard, M., & Heydari, M. (2022). Groundwater quality evaluation and risk assessment of nitrate using monte carlo simulation and sensitivity analysis in rural areas of Divandarreh County, Kurdistan province, Iran. *International Journal of Environmental Analytical Chemistry*, 102(10), 2213-2231.

Sonone, S. S., Jadhav, S., Sankhla, M. S., & Kumar, R. (2020). Water contamination by heavy metals and their toxic effect on aquaculture and human health through food Chain. Lett. Appl. *NanoBioScience*, 10(2), 2148-2166.

Sorensen, J. P., Carr, A. F., Nayebare, J., Diongue, D. M., Pouye, A., Roffo, R., & Taylor, R. G. (2020). Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators. *Scientific reports,* 10(1), 1-9.

Sultan, A. S., Elgharib, M. A., Tavares, T., Jessri, M., & Basile, J. R. (2020). The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *Journal of Oral Pathology & Medicine*, *49*(9), 849-856.

Sun, A. Y., Green, R., Swenson, S., & Rodell, M. (2012). Toward calibration of regional groundwater models using GRACE data. *Journal of Hydrology*, 422, 1-9.

Teixeira, F. J., & Gomes, C. S. (2021). Mineral Water: Essential to Life, Health, and Wellness. In Minerals latu sensu and Human Health. Springer, Cham. (pp. 517-556).

Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., & Plimpton, S. J. (2022). LAMMPS-a flexible simulation tool for particle-based materials modelling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271, 108171.

Tulbure, A. A., Tulbure, A. A., & Dulf, E. H. (2022). A review on modern defect detection models using DCNNs–Deep convolutional neural networks. *Journal of Advanced Research*, 35, 33-48.

Unigwe, C. O., & Egbueri, J. C. (2022). Drinking water quality assessment based on statistical analysis and three water quality indices (MWQI, IWQI and EWQI): a case study. *Environment, Development and Sustainability*, 1-22.

Van der Veer Martens, B., Illston, B. G., & Fiebrich, C. A. (2017). The Oklahoma Mesonet: A pilot study of environmental sensor data citations. *Data Science Journal*, *16*, 47-47.

Van Herck, S. (2022). Re/constructing Computing Experiences. From" punch girls" in the 1940s to" computer boys" in the 1980s (Doctoral dissertation, University of Luxembourg, *Esch-sur-Alzette*, Luxembourg). *6*(9), 9-236.

Varoquaux, G., & Poldrack, R. A. (2019). Predictive models avoid excessive reductionism in cognitive neuroimaging. *Current opinion in neurobiology*, *55*, 1-6.

Verma, A. (2022). Role of the Procedures of Deep Learning for Higher Profitability in the Agriculture Sector. In Sentimental Analysis and Deep Learning. Springer, Singapore. (pp. 745-755).

Vuurberg, G., Hoorntje, A., Wink, L. M., Van Der Doelen, B. F., Van Den Bekerom, M. P., Dekker, R., & Kerkhoffs, G. M. (2018). Diagnosis, treatment and prevention of ankle sprains: update of an evidence-based clinical guideline. *British journal of sports medicine*, *52*(15), 956-956.

Wang, H., Knabe, D., Engelhardt, I., Droste, B., Rohns, H. P., Stumpp, C., & Griebler, C. (2022). Dynamics of pathogens and fecal indicators during riverbank filtration in times of high and low river levels. *Water Research*, 209, 117961.

Wang, J., Yan, H., Xin, K., & Tao, T. (2019). Iron stability on the inner wall of prepared polyethylene drinking pipe: Effects of multi-water quality factors. *Science of The Total Environment*, *658*, 1006-1012.

Wang, Y., & Li, P. (2022). Appraisal of shallow groundwater quality with human health risk assessment in different seasons in rural areas of the Guanzhong Plain (China). *Environmental Research*, 207, 112210.

Wang, Z., Fok, K. W., & Thing, V. L. (2022). Machine learning for encrypted malicious traffic detection: Approaches, datasets and comparative study. *Computers & Security*, 113, 102542.

Wang, Z., You, X., Yang, C., Li, W., Li, Y., Li, Y., & Jiang, Z. (2022). Ultrathin polyamide nanofiltration membranes with tunable chargeability for multivalent cation removal. *Journal of Membrane Science*, 642, 119971.

Watkins, E. R., & Newbold, A. (2020). Factorial designs help to understand how psychological therapy works. *Frontiers in Psychiatry*, *11*, 477682.

Yadav, B., Gupta, P. K., Patidar, N., & Himanshu, S. K. (2020). Ensemble modelling framework for groundwater level prediction in urban areas of India. *Science of the Total Environment*, *712*, 135539.

Yao, Y., He, K., Li, Y., Zhang, X., Ma, Z., Cui, Z., & Chen, X. (2022). Research and Application of Supersaturated Dissolved Oxygen Technology Combined with Magnetization Technology in the Improvement of Water Quality: Taking the South-to-North Water Diversion Project of China as a Pilot Project. *Sustainability*, 14(5), 2684.

Yeleliere, E., Cobbina, S. J., & Duwiejuah, A. B. (2018). Review of Ghana's water resources: the quality and management with particular focus on freshwater resources. *Applied Water Science*, 8(3), 1-12.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.

Zhou, X., Moinuddin, M., Renaud, F., Barrett, B., Xu, J., Liang, Q., & Hoey, T. (2022). Development of an SDG interlinkages analysis model at the river basin scale: a case study in the Luanhe River Basin, China. *Sustainability Science*, 1-29.

Zhu, D., Chen, J., Shen, X., Li, X., & Elhoseiny, M. (2023). Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.

Zhu, Z., Lin, K., Jain, A. K., & Zhou, J. (2023). Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zuo, R., Wang, J., Xiong, Y., & Wang, Z. (2021). The processing methods of geochemical exploration data: Past, present, and future. *Applied Geochemistry*, *132*, 105072.

APPENDICES

APPENDIX A

Table 1: Descriptive Statistics of Physical Parameters in Groundwater (Each Region)

		Central Region			Greater Accra			Western		
Parameters	N	Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DO	100	7.16	1.84	11.20	7.1	3.39	9.9	7.34	3.88	11.76
Ph	100	6.88	4.80	12.40	6.61	3.4	12	6.97	4.8	10
Temperature	100	27.34	25.60	33.60	28.51	25.9	33.6	26.36	23.6	32.6
TH	100	156.28	13.85	372.00	192.73	16.69	300	129.75	12.85	292
TDS	100	452.23	376.47	555.67	456.95	391.67	553.95	432.14	396.83	553.95
Turbidity	100	7.09	2.84	13.64	8.16	2.43	11.64	9.08	2.89	14.87
EC	100	432.31	269.45	669.73	418.26	245.86	672.56	428.96	277.38	629.28
Na	100	232.29	35.00	675.00	244.64	23	675	249.4	43	397
K	100	118.21	21.00	227.00	126.98	60	206	129.56	36	239
Ca	100	160.05	42.00	284.00	185.76	93	286	169.46	52	421
Mg	100	51.72	11.00	132.00	55.06	4	119	50.02	3	127
SO_4	100	75.09	116.00	161.00	67.25	46.4	131	98.11	54	147
PO_4	100	3.10	0.56	6.30	4.81	3.0	8.32	5.11	1	8.2
Cl	100	141.97	39.00	319.00	162.66	50	586	138.52	47	586
HCO_3	100	10.74	3.50	18.00	11.94	4	22	10.04	3	16
SiO_3	100	47.59	8.00	53.00	35.78	18	83	50.71	12	103
NO_3	100	21.96	7.00	42.00	21.94	5	44	24.6	6	76

APPENDIX B

Table 2: Groundwater Quality Index Estimation of Central Region

Regions in Ghana	Standard Values	1/Sn	Σ1/Sn	K = 1/(Σ1/Sn)	S = K/Sn	Ideal Values	Predicted Values	Vn/Sn	Vn/Sn X 100	WnQn
Na^{+}	200.00	0.01	10.25	0.10	0.00	0.00	134	1.17	117.00	0.06
K^{+}	300.00	0.00	10.25	0.10	0.00	0.00	118	0.55	55.20	0.02
Mg^{2+}	50.00	0.00	10.25	0.10	0.00	0.00	50	0.39	39.33	0.01
Ca^{2+}	100.00	0.02	10.25	0.10	0.00	0.00	160	1.00	100.00	0.20
HCO_{3-}	10.00	0.00	10.25	0.10	0.00	0.00	14	0.71	71.20	0.03
Cl^{-}	250.00	0.01	10.25	0.10	0.00	0.00	138	1.60	160.00	0.16
NO_3^{-}	50.00	0.50	10.25	0.10	0.05	0.00	27	17.50	1750.00	85.33
SiO_3^{-}	100.00	0.10	10.25	0.10	0.01	0.00	55	1.40	140.00	1.37
PO_4^{3-}	2.00	0.01	10.25	0.10	0.00	0.00	3	0.55	55.00	0.05
SO_4^{2-}	250.00	0.02	10.25	0.10	0.00	0.00	78	0.54	54.00	0.11
	Σ1/Sn =	**0.68**			**0.07**		**26.63**			**87.32**

Table 3: Groundwater Quality Index Estimation of Greater Accra Regions

Ions	Standard Values	1/Sn	Σ1/Sn	K = 1/(Σ1/Sn)	S = K/Sn	Ideal Values	Predicted Values	Vn/Sn	Vn/Sn X 100	WnQn
Na^{+}	200.00	0.01	10.25	0.10	0.00	0.00	144	0.72	72.00	0.04
K^{+}	300.00	0.00	10.25	0.10	0.00	0.00	124	0.26	26.00	0.01
Mg^{2+}	50.00	0.00	10.25	0.10	0.00	0.00	51	0.41	41.33	0.01
Ca^{2+}	100.00	0.02	10.25	0.10	0.00	0.00	162	1.02	102.00	0.20
HCO_{3-}	10.00	0.00	10.25	0.10	0.00	0.00	16	0.66	66.40	0.03
Cl^{-}	250.00	0.01	10.25	0.10	0.00	0.00	65	1.62	162.00	0.16
NO_3^{-}	50.00	0.50	10.25	0.10	0.05	0.00	18	19.00	1900.00	92.64
SiO_3^{-}	100.00	0.10	10.25	0.10	0.01	0.00	35	1.60	160.00	1.56
PO_4^{3-}	2.00	0.01	10.25	0.10	0.00	0.00	4	0.35	35.00	0.03
SO_4^{2-}	250.00	0.02	10.25	0.10	0.00	0.00	66	0.36	36.00	0.07
	Σ1/Sn =	**0.68**			**0.07**		**26.63**			**94.75**

Table 4: Groundwater Quality Index Estimation of Greater Accra Regions

Ions	Standard Values	1/Sn	Σ1/Sn	K = 1/(Σ1/Sn)	S = K/Sn	Ideal Values	Predicted Values	Vn/Sn	Vn/Sn X 100	WnQn
Na^{+}	200.00	0.01	10.25	0.10	0.00	0.00	155	0.78	77.50	0.04
K^{+}	250.00	0.00	10.25	0.10	0.00	0.00	128	0.55	55.20	0.02
Mg^{2+}	50.00	0.00	10.25	0.10	0.00	0.00	47	0.43	42.67	0.01
Ca^{2+}	100.00	0.02	10.25	0.10	0.00	0.00	168	0.94	94.00	0.18
HCO_{3-}	10.00	0.00	10.25	0.10	0.00	0.00	13	0.64	64.00	0.02
Cl^{-}	250.00	0.01	10.25	0.10	0.00	0.00	138	1.68	168.00	0.16
SO_4^{2-}	250.00	0.50	10.25	0.10	0.05	0.00	22	22.50	2250.00	109.71
SiO_3^{-}	100.00	0.10	10.25	0.10	0.01	0.00	49	1.30	130.00	1.27
PO_4^{3-}	2.00	0.01	10.25	0.10	0.00	0.00	6	0.49	49.00	0.05
NO_3^{-}	50.00	0.02	10.25	0.10	0.00	0.00	98	0.44	44.00	0.09
	Σ1/Sn =	**0.68**			**0.07**		**26.63**			**111.56**

APPENDIX C

Table 5: ANOVA: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
DO	3	21.6	7.2	0.0156
PH	3	20.46	6.82	0.0351
Temperature	3	82.21	27.40333	1.158633
TH	3	478.76	159.5867	999.8206
TDS	3	1341.32	447.1067	173.5704
Turbidity	3	24.33	8.11	0.9919
EC	3	1279.53	426.51	53.8525
Na	3	726.33	242.11	77.9887
K	3	374.75	124.9167	35.39863
Ca	3	515.27	171.7567	169.207
Mg	3	156.8	52.26667	6.574533
SO_4	3	536.43	178.81	187.6477
PO4	3	29.92	3.70667	20.03203
Cl	3	443.15	147.7167	170.453
HCO_3	3	32.72	10.90667	0.923333
SiO_3	3	134.08	44.69333	62.01923
NO_3	3	68.5	22.83333	2.340933
Central Region	17	2088.16	122.8329	19481.74
Greater Accra Regions	17	2173.14	127.8318	19734.09
Western Region	17	2104.86	123.8153	18777.53