Republic of Iraq
Ministry of Higher Education and Scientific ResearchUniversity of Babylon / College of Engineering
Civil Engineering Department

Hybrid Artificial Intelligence Modeling for Groundwater in the Dammam Confined Aquifer/ Iraq

A Thesis
Submitted to the College of Engineering / Universityof Babylon in Partial Fulfillment of the Requirements for Degree of Doctor of Philosophy in Civil Engineering/ Water Resources Engineering

By:

Marwah Abdullah Shlash

(**M.Sc., 2012**)

Supervised by:

Asst. Prof. Dr. Imad Habeeb Obead

May, 2023 A.D. Dhul Qi'da, 1444 A.H.

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

أَلَمْ تَرَ أَنَّ اللَّهَ أَنزَلَ مِنَ السَّمَاءِ مَاءً فَسَلَكَهُ يَنَابِيعَ فِي الْأَرْضِ ثُمَّ يُخْرِجُ بِهِ زَرْعًا مُّخْتَلِفًا أَلْوَانُهُ ثُمَّ يَهِيجُ فَتَرَاهُ مُصْفَرًّا ثُمَّ يَجْعَلُهُ حُطَامًا ۚ إِنَّ فِي ذَٰلِكَ لَذِكْرَىٰ لِأُولِي الْأَلْبَابِ ﴿٢١﴾

صدق الله العلي العظيم

[الزمر: 21]

Supervisors' Certificate

I certify that this thesis entitled **" Hybrid Artificial Intelligence Modeling for Groundwater in the Dammam Confined Aquifer/ Iraq "** was prepared by **" Marwah Abdullah Maery Shlash"** under my supervision as a partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering / Civil Engineering / Water Recourses at University of Babylon.

Signature:
Name: Asst. Prof. Dr. Imad Habeeb Obead
Date: / /2023

Dedication

To my father, who taught me how to stand firm. He has left us, but his advice still guides me.

To my mother, who taught me to give and overwhelmed me with her tenderness and generosity.

To my brothers and sister, who taught me that life without bonding, love, and cooperation is worth nothing.

To my husband, who filled my life with a challenge and overcame the odds.

To my soul and the beat of my heart. My kids, Noor and Mohammed.

I dedicate this scientific effort with love and respect.

Marwah Abdullah Shlash

Acknowledgments

First of all, I thank God almighty, who provided me with the strength to do this task.

I would like to express my deepest appreciation to my supervisor, Asst. Prof. Dr. Imad Habeeb Obead, for his guidance, encouragement, ideas, helpful suggestions, and continuous effort.

Special thanks are extended to Prof. Dr. Thair J. Mizhir Alfatlawi, Head of the Civil Engineering Department at the Faculty of Engineering-University of Babylon, as well as Asst. Prof. Dr. Zaid Hameed Majeed and Asst. Prof. Dr. Haider M. Allrubaie for their assistance and ongoing support.

I would like to express my sincere gratitude to every faculty member in the Civil Engineering Department-University of Babylon who has provided invaluable assistance during my academic journey.

I would also like to take this opportunity to extend my deepest appreciation and thanks to everyone at the Company of Geological Survey and Mining (GEOSURV) in Baghdad for the support they offered and the valuable data they provided.

Finally, many thanks to everyone who contributed to showing this effort for existence.

Marwah Abdullah Shlash

ABSTRACT

Groundwater is a valuable resource in arid and semi-arid environments due to the dependable nature of deep aquifers. However, as climate change progresses, there is expected to be a reduction in precipitation and an increase in droughts. Iraq has been severely impacted by an extended period of drought, leading to a notable increase in the use of groundwater resources over the past decade.

This study examines the Dammam confined aquifer in the Najaf and Muthanna governorates. The work is divided into two parts. The first one presents a numerical simulation using Groundwater Modelling System (GMS v10.4) software to analyze the groundwater flow under steady and transient conditions. Additionally, two realistic operational scenarios were introduced. The findings indicate that operating 89 wells at full capacity as the first scenario caused a groundwater head drawdown of 0.4-5.8 m. While the operating 228 wells in the second scenario caused a drawdown of 1-22 m due to the increased pumping rates from 582.3 ℓps to 1534.1 ℓps. The study area's hydraulic conductivity was found to vary from 1.47 to 20.0 m/day, and the recharging rate was determined to be 1.66×10^{-06} m/day. The satisfactory match between the measured and predicted groundwater heads during the operation period from 1 Jul 2021 until 1 Jun 2022 obtained from transient-state calibration demonstrated that the storage coefficient (storativity) for the study region ranged from 5×10^{-5} to 6.5×10^{-4}.

Furthermore, a groundwater potentiality map (GPM) and stressed depletion zones were created to aid in groundwater management using a multicriteria decision-making (MCDM) approach through the analytical hierarchy process (AHP) with seven thematic maps representing various contributing factors. These factors include hydrogeological aquifer unit types, transmissivity, lineaments density, slope, soil, land use and land cover, and drainage density that significantly influence groundwater availability. The generated GPM classifies the Dammam

confined aquifer of the study area into five zones, indicating very good, good, moderate, poor and very poor groundwater potential zones representing (15.91%, 39.02%, 33.01%, 6.49%, and 5.57%) of the total area, respectively. The AHP method achieves an accuracy of (82.5%) for accurately identifying GWPZs.

The second phase of this study utilizes artificial intelligence (AI) methods to produce hybrid models involving five supervised machine-learning algorithms (MLAs) integrated with GIS-based techniques to assess the groundwater potentiality that tackles the global water depletion issue. The MLAs used were Ensemble Boosted Trees (EBT), Naive Bayes (NB), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), and k-Nearest Neighbours (kNN) to generate the GPM. The results show that the weighted kNN classifier indicates better performance with a validation accuracy of 90.7 %. Out of 349 observation wells, the kNN model accurately classified 322 wells with the area under the ROC curve (AUC) equal to 1, indicating accurate predictions. Consequently, a five-classes GPM is created representing five GWPZ accounting for 12.37% (very good), 42.12% (good), 35.69% (moderate), 4.24% (poor), and 5.58% (very poor) of the total area, respectively. While in the binary classification, a comparative analysis was conducted to assess the effectiveness of predictive classifiers capabilities. The kNN classifier with a better accuracy of 93.3% is used to create the final binary GPM, indicating two classes of low-yield and high-yield GWPZs covering about 45.51% and 54.49% of the total area, respectively. Moreover, a graphical user interface (GUI) was developed in MATLAB-based applications to identify significant groundwater potential zones in both multiclass and binary formats. The process involves the insertion of the geographical coordinates of a well using the UTM Projection, including its latitude and longitude, which facilitates the identification of its classification and location on the study area map.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Symbols	Definition	Dimension
b	Thickness of the aquifer	L
h	Head	L
K	Hydraulic conductivity	L/T
K_x, K_y, K_z	Components of the hydraulic conductivity in the x, y, and z axes	L/T
q	Flux through the representative elementary volume REV	L/T
q_x , q_y, q_z	Flux components along the x, y, and z axes	L/T
Q	Discharge	L^3/T
R^2	Coefficient of Determination	---
S	Storativity, alternatively referred to as the storage coefficient	---
Ss	Specific storage (the volume of water released from storage per unit change in the head (h) per aquifer unit volume)	L^{-1}
t	Time	T
T	The transmissivity of an aquifer that is related to its hydraulic conductivity as T=Kb)	L^2/T
ΔV	Rate of change in storage	L^3
Δx, Δy, Δz	Finite difference grid intervals	L

LIST OF ABBREVIATIONS

Abbreviations	Description
AHP	Analytical Hierarchy Process
ANNs	Artificial Neural Networks
AUC	Area Under Curve
Comp.	Computed
DEM	Digital Elevation Model
FDR	False Discovery Rates

Abbreviations	Description
FNR	False Negative Rates
GCGW	General Committee for Groundwater
GIS	Geographic Information System
GMS	Groundwater Modeling System
GPM	Groundwater potential map
GWPZ	Groundwater potential zone
IMOS	Iraqi Meteorological Organization and Seismology
kNN	k-Nearest Neighbours
LDA	Linear Discriminant Analysis
LULC	Land Use and Land Cover
MAE	Mean Absolute Error
ME	Mean Error
MLAs	Machine Learning Algorithms
MLP	Multi-Layer Perceptron
NB	Naive Bayes
NE	Northeast
Obs.	Observation
PPV	Positive Predictive Values
QDA	Quadratic Discriminant Analysis
RH	Recharge Rate
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristics
RS	Remote Sensing
SLP	Single-Layer Perceptron
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SW	Southwest
TPR	True Positive Rates
Trans.	Transient

Chapter One

Introduction

1.1 General

Throughout history, the lack of attention to regional water diplomacy has exposed Iraq's transboundary water resources to vulnerability. Iraq is currently facing a crisis with its water resources and services. This is mainly due to various issues affecting surface water supply, such as excessive use of the source for all life-sustaining purposes, including agriculture, industry, domestic usage, and drinking, as well as high temperatures leading to rapid evaporation from surface water bodies. Additionally, the completion and operation of Turkey's Southeastern Anatolia Project could potentially lead to a shortage of surface water sources, as it involves constructing many dams on the Euphrates and Tigris rivers to expand the agricultural area and generate electricity. Moreover, rapid population growth and the consequent increase in water consumption have aggravated the problem of water scarcity, particularly around large urban centers. Many countries worldwide have adopted a proactive approach to establishing a water-efficient society that emphasizes water conservation and optimal utilization of alternative water resources to combat the mounting pressure on water resources (Dai et al., 2022).

According to (UN-Water, 2021), around 2.3 billion people now live in countries with water scarcity, also known as water-stressed countries. About 733 million people are in countries with high and critical water scarcity. Water consumption has grown more than twice as fast as population growth over the past century. By 2025, an estimated 1.8 billion people will live in water-stressed areas, with two-thirds of the world's population living in water-stressed regions due to land use, growth, and climate change. The main concern is developing strategies

to efficiently conserve, manage, and allocate water resources. Therefore, there is growing interest in using groundwater to alleviate this crisis. According to UNESCO, groundwater is a vital freshwater resource, providing about 50% of available drinking water (Ghasemian, 2016; Karim, 2022). These resources are generally high quality, do not require chemical treatment, and are generally free of pathogenic factors, making groundwater an essential and reliable resource for various users (Carrard et al., 2019).

To predict groundwater availability, hydrogeological data is necessary for quantitative and qualitative numerical studies that require data from hydrogeological surveys and man-made involvements. The hydrogeological information helps to develop models analogous to field conditions. However, obtaining accurate field data representing hydrological processes and simulating them in numerical models is becoming increasingly complex, as they require a large amount of detailed hydrological data that suffers from accuracy during its inaccurately acquired and pre-processing (Yadav et al., 2017). As a result, data-driven machine learning models are becoming more attractive as they require less data and improve groundwater prediction accuracy. These models can recognize patterns and capture the time-dependent behaviour of hydrological systems without detailed knowledge of the physical properties required by groundwater flow models (Zhou et al., 2017; Kanyama et al., 2020).

1.2 Statement of Problem

The research problem is to investigate the general movement and availability of groundwater in the Dammam confined aquifer within the study area by recognizing patterns and capturing the time-dependent behaviour of hydrological systems without relying on resource-intensive and time-consuming numerical models that necessitate detailed knowledge and accurate field data to

facilitate the implementation of effective management strategies and ensure an adequate water supply to address future scarcity concerns. It is worth noting that the Dammam confined aquifer is considered the main aquifer in the southern desert of Iraq and holds significant potential (Al-Jiburi and Al-Basrawi, 2021).

Developing an accurate and reliable data-driven machine-learning model that can predict groundwater availability by recognizing patterns and capturing the time-dependent behaviour of hydrological systems without relying on detailed knowledge of the physical properties required by groundwater flow models, thereby overcoming the challenges associated with obtaining accurate field data.

1.3 Significance and Potential Contributions

Several key contributions motivate this work, which can be summarized as follows:

(1) Predicting groundwater availability is essential for ensuring water sustainability and minimizing the impact of droughts.

(2) Remote Sensing (RS) and Geographic Information Systems (GIS) can be utilized to collect, analyze, present, and monitor geospatial data. RS data can be combined with field observations to conduct resource inventory and assess the current state. When integrated into a GIS environment, this information can be utilized to evaluate and analyze factors that impact decision-making concerning groundwater exploration and mapping.

(3) Machine learning algorithms can improve groundwater prediction by identifying patterns and learninging the nonlinear relationship between input variables and outputs during training phase. This allows resource planners to develop targeted monitoring programs, establish protective measures for groundwater, and evaluate the sustainability of drinking water sources derived from groundwater.

(4) The methodology developed in this study can also be applied to other regions of Iraq where significant groundwater aquifers have not yet been explored or examined.

1.4 Aim and Objectives

This study aims to innovate a hybrid model to narrow down uncertainties affecting the complete model response and derive a calibrated model concerning the limited knowledge available using data-driven techniques, including machine learning and GIS-based methodology. The proposed model aims to accurately predict the state of groundwater potential zones and provide insights into the behaviour of a groundwater source by including the effect of the set of influencing factors. Furthermore, it can be easily updated with new information for effective groundwater management to meet the future demand in the study area. The major objectives are:

(1) Analyzing the groundwater flow under steady and transient conditions using the Groundwater Modeling System (GMS) software to determine the hydraulic conductivity and recharge rate of the Dammam confined aquifer in the study area, simulate groundwater movement, and generate a contour map for the simulated heads.

(2) Mapping the groundwater piezometric head drawdowns for the investigated area under transient conditions by applying different operating scenarios to explore the probable head changes caused by pumping from the aquifer between July 1st, 2021, and June 1st, 2022.

(3) Creating a groundwater potentiality map (GPM) and stressed depletion zones by integrating the Geographical Information Systems (GIS) and Multicriteria Decision-Making (MCDM) through the Analytical Hierarchy Process (AHP) to facilitate decision-making by evaluating multiple factors to generate seven

thematic maps of seven groundwater potentiality influencing factors using ArcGIS 10.7.1.

(4) Construction of hybrid machine-learning (ML) models through establishing a binary and multiclass classification database. This database will be utilized to train five different ML models employing five different learning strategies. The predictions made by these models will be evaluated using different statistical measures in addition to assessing classifier accuracy.

(5) Conduct a comparative analysis to evaluate the outcomes produced by hybrid ML approaches to identify the most accurate model that could be employed to create the final groundwater potential map.

(6) Developing a graphical user interface (GUI) in MATLAB-based applications that can be shared with individuals who do not possess proficiency in MATLAB using the Standalone Desktop App. To significantly detect the groundwater potential zones (GWPZs) by identifying the features of seven conditioning influencing factors related to groundwater potentiality. The identified features will be trained and tested using one of the five classifiers provided. The generated output will be in the form of multiclass and binary. This is achieved by inputting the latitude and longitude of a well, which enables the detection of the well's classification and location on the study area map.

1.5 Assumptions and Limitations

A numerical and hybrid machine learning model is developed for the water flow system of the Dammam confined aquifer, taking into consideration its hydrogeologic features. The developed model is dependent on the following assumptions:

1. The numerical model used in this study was created using the conceptual model approach within the MODFLOW-2000 package of the

Groundwater Modeling System (GMS) program, which employs a 3D, cell-centered, finite difference approach.

2. The Layer-Property Flow (LPF) package within MODFLOW-2000 defines the horizontal (K_h) and vertical hydraulic conductivity (K_v) for each layer. The MODFLOW then computes the cell-by-cell conductances using the K values and the layer geometry. This study specifies the vertical anisotropy factor rather than vertical hydraulic conductivity values. This option is particularly useful when performing automated parameter estimation (PEST) since it ties the K_v to K_h and eliminates the need to define K_v as an independent parameter.
3. The process of approximating the governing equations involves discretizing the modelled domain into rectangular cells in the x-y direction, with variable cell thickness in the z-direction.
4. The model consisted of two layers (K=2); the first is unconfined, while the second is confined.
5. The horizontal hydraulic conductivity is heterogeneous (spatial variation), in which the study area is segmented into five different zones to facilitate the input of the initial hydraulic conductivity data.
6. The recharge rate was assumed to be uniform. The initial guess of recharge is 1.4×10^{-5} calculated according to meteorological records provided by the General Commission of Iraqi Meteorological Organization and Seismology (IMOS, 2021) for Najaf and Muthanna governorates monitoring stations for the (47) years from (1974 – 2021).

1.6 Thesis Organization

The present thesis consists of eight chapters. The first chapter discusses the context of the groundwater issue, focusing on Iraq. It also highlights the research objectives, problem statement, and motivation for investigating the Dammam

confined aquifer in southern Iraq's desert zone, specifically in Najaf and Muthana governorates.

Chapter two introduced the groundwater model, encompassing its various types and the prevalent groundwater codes and limitations. Also, reviewed the investigations conducted by other academics. These studies are divided into two categories: a review of prior numerical models related to groundwater and earlier machine-learning models concerning groundwater potential mapping associated with the current study.

Chapter three defines the governing equations for groundwater flow. The data necessary for constructing the numerical model are also discussed, along with the hydrogeological and geological characterization of the study area. It also illustrates the aquifers' description and the conceptualization of the model using the MODFLOW package integrated with the GMS program, including the model's input parameters, boundary conditions, and initial conditions. Moreover, it discusses the groundwater model's calibration, sensitivity analysis, validation procedures, and implementation of a calibrated groundwater model to simulate groundwater flow behaviour in both steady and transient states.

Chapter four provides a framework to identify potential groundwater zones using RS and GIS techniques, such as the Analytical Hierarchy Process (AHP), in collecting, analyzing, presenting, and monitoring geospatial data. A set of influencing factors (seven thematic maps) was prepared by giving weights to each influencing factor according to the degree to which that component contributes to groundwater availability to generate the final GPM for the study area. Additionally, a comparison was performed between the groundwater yields in each well and the corresponding groundwater potential zones to calculate the used techniques' overall accuracy for accurately delineating a GPM.

Chapter five reviews the concepts behind some of the most popular and standard supervised classification machine-learning methods and best practice strategies for its implementation. Also, this chapter presents a generalized methodology for supervised machine learning classification implementation, including data pre-processing to correct the inconsistencies in raw data before feeding the database to train the machine learning algorithms. In addition, statistical measures of classifier performance were applied as a direct and statistically accurate indicator of the predicted performance of an achieved trained classifier.

Chapter six provides a detailed methodology to construct a novel hybrid model for groundwater potentiality mapping (GPM). The supervised machine learning (ML) of data-driven based techniques integrated with knowledge-driven GIS-based techniques will be used to evaluate the GWPZ of the Dammam confined aquifer. Also, it compares five machine learning algorithms' responses to variations in training data spatial distribution based on their accuracy matrices to generate multiclass and binary GPM maps. Additionally, this chapter presents the methodology employed in this study to create a graphical user interface (GUI) within MATLAB-based applications, which is capable of identifying significant groundwater potential zones. This is achieved by inputting the latitude and longitude of a well, which enables the detection of the well's classification and location on a map.

Finally, chapter seven presents the findings based on the data given in the preceding chapters. Furthermore, recommendations for additional research are offered.

Chapter Two

Review of Literature

2.1 Introduction

Groundwater flow modeling has received considerable attention from researchers. There are multiple aquifers collecting groundwater within the study area, including the Hartha, Tayarat, Umm Er Radhuma, Rus/Jill, Dammam, Ghar/Euphrates, Nfayil, Zahra, Dibdibba, and Quaternary sediment formations from oldest to newest (Al-Jiburi and Al-Basrawi, 2009). However, a limited number of national-level investigations have been conducted on groundwater in Iraq. Several factors contribute to this, including limited experience in groundwater analysis, challenges in assigning accurate boundary conditions for investigation areas, and difficulty accessing data. Additionally, some studies require financial support to conduct experiments that enhance data needed for developing groundwater models.

A wide range of models has been created and used to model groundwater dynamics. These models may be divided into numerical models (physical descriptive models) and empirical models. The inability of empirical models to adequately account for the dynamical behaviour of the groundwater system changes is one of the most significant drawbacks of these models. Over the last three decades, many physically based numerical models for modeling groundwater systems have been developed (Markstrom et al., 2008, Neitsch et al., 2009).

In general, three key goals are typically followed when creating models, as Anderson and Woessner summarised in 1992. These objectives involve anticipating how the modelled system will respond to specific events, gaining a

comprehensive understanding of the controlling parameters through analysis of the system's dynamics, and establishing regulatory criteria for the relevant study areas by first developing suitable geological settings for flow analysis.

The traditional method of groundwater analysis, which relies mainly on mathematical modeling, has certain limitations, such as the need for a considerable number of accurate data, which can never be established with exact certainty. Other drawbacks include the physical properties of the aquifer. In addition, the computational resources are insufficient to meet the ever-increasing demands of more accurate and complicated numerical models (Luo et al., 2003). Since the groundwater balance is affected by various circumstances, it is challenging to estimate the groundwater fluctuation rate accurately. Hence, artificial intelligence (AI) methods can help to solve such problems. In hydrology and water management, artificial intelligence (AI) methods are increasingly used for analyzing large-scale and long-term data due to their ability to make accurate predictions with minimal inputs, even when data is limited. These AI models are designed to mimic the human brain's ability to rationalize and make decisions that are most likely to achieve a desired outcome and to predict outcomes with high accuracy despite limited data (Wang et al., 2009).

Recently, hydrologic studies have employed data-driven models like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) to predict future results. The Applications of advanced (AI) models are expected to provide more precise results and serve as a dependable alternative for complex geoscience problems (Vapnik, 2013; LeCun et al., 2015).

2.2 Models of Groundwater

The structure provided by a model helps the modeler identify mistakes in assumptions and unconsidered processes. In other words, the model application is

an exercise about how a system operates. The information about the hydrogeology state is simplified and made easier to understand by using a conceptual model of a groundwater system. A conceptual model represents the system's history and current condition based on field information from the site and the knowledge provided by other similar sites. Groundwater flow models solve for what cannot be fully observed or measured, such as the head distribution in space and time. A groundwater model that quantitatively describes heads in space and time as a simplified representation of complicated hydrogeologic circumstances is considered a more powerful groundwater model. To explain and assess these models comprehensively, it is important to categorize them into two main categories, physical (laboratory) models and mathematical models (Anderson and Woessner, 2015).

2.2.1 Physical Models

Laboratory tanks and columns filled with porous material (typically sand) are used to directly monitor groundwater heads and flows. Darcy (1856), for example, measured the head in sand-packed columns of different diameters and lengths to demonstrate that flow in a porous medium is linearly related to the head gradient. Oswald and Kinzelbach (2004) conducted a series of laboratory experiments to investigate the phenomena of variable-density subsurface flow in a saturated media, intending to use the results to evaluate the validity of some numerical codes. Furthermore, these models may be utilized to investigate groundwater pollution movement under various site field circumstances (Hoopes and Harleman, 1967; Ishaq and Ajward, 1993).

However, due to the laboratory model's limited scale compared to the actual dimensions of the field location, notable differences exist between the phenomena experienced in the sandbox and those observed in the field. Therefore,

it is necessary to reevaluate the results of such models when they are implemented and translated into a real-world situation (Loudyi, 2005).

2.2.2 Mathematical Models

The functioning of mathematical models relies on mathematical equations to represent the physical processes that govern the movement of groundwater, such as flow through porous media, transport of contaminants, and interactions with surface water. By inputting data on factors such as geology, climate, and human activities, groundwater models can provide valuable insights into the behavior of aquifers and assist in managing and protecting these important water resources. They are used in various applications, including water resource management, contaminant remediation, and environmental impact assessment (Todd, 2007). These models use mathematical expressions, such as linear algebraic equations or partial differential equations, to represent the fundamental concept of the behavior of the aquifer system (Loudyi, 2005).

According to Anderson and Woessner (2015), it is possible to subdivide the mathematical models into two types:

i- Process-based models, also known as physically-based models, are based on mathematical formalizations of physical laws and assumptions to represent groundwater flow within the problem domain. A boundary condition that specifies heads or flows along the problem domain's boundaries, initial conditions that specify heads within the problem domain at the beginning of the simulation for time-dependent problems, and a governing equation that describes the physical processes occurring within the problem domain create a process-based mathematical groundwater flow model.

ii- Data-driven models use empirical or statistical equations constructed from the available data to compute an unknown variable, for example, computing the head at the water table from knowledge about another variable that can be measured reliably, such as precipitation. Such methodology will be relied upon in this study.

Mathematical models may be solved through analytical or numerical methods. The solution of groundwater flow mathematical models involves the determination of head distribution in both space and time, particularly for transient scenarios. In contrast to other modeling approaches, analytical models require a significant level of simplification of natural phenomena to formulate a mathematically applicable problem that can be solved through closed-form solutions.

2.3 Modeling Software for Groundwater Systems

Groundwater modeling software has become an essential tool for understanding and managing groundwater resources. Various types of groundwater modeling software are available, each with strengths and limitations. While modeling software has the potential to be very useful, it is important to recognize the limitations and potential errors associated with these models.

One limitation of groundwater modeling software is that they are based on simplifications of complex hydrogeological systems. The accuracy of the models depends on the quality and quantity of data inputted and the assumptions made about the hydrogeological system. Additionally, it is essential to conduct model calibration and validation to ensure accurate predictions and understanding of results. A possible limitation is related to the utilization of software, as the necessary level of expertise required to operate such programs may present a difficulty for certain users. In addition, it has been noted that some modeling

software may require expensive licensing fees, making them inaccessible to smaller entities (Zheng and Wang, 1999; Anderson & Woessner, 2015). Table (2.1) summarizes some of the most powerful groundwater modeling software available, according to Kumar (2019).

Table (2.1): The most efficient and available groundwater modeling software (Kumar, 2019).

Software	Function	Source
MODFLOW	simulates groundwater flow and transport through porous media.	United States Geological Survey (USGS).
FEFLOW	A software designed to simulate complex groundwater flow and transport phenomena, including heat and mass transport.	DHI-WASY GmbH, a German company that specializes in developing software solutions for water resources and environmental engineering.
Visual MODFLOW	A user-friendly groundwater modeling software that combines the power of MODFLOW with the simplicity of a graphical user interface.	Waterloo Hydrogeologic, Inc.
Groundwater Vistas	simulating and analyzing groundwater flow and contaminant transport.	Environmental Simulations Inc. (ESI).
HYDRUS	simulate water level and solute movement in the unsaturated zone and can be used to model soil water dynamics, evapotranspiration, and groundwater recharge.	Dr. Jirka Šimůnek and his colleagues, the University of California, Riverside. It is now maintained and distributed by PC-Progress GmbH.

In 1990, the Groundwater Modeling System (GMS) was developed by Aquaveo Limited Liability Company in Provo, Utah. GMS allows different applications, such as GIS, EXCEL, and AutoCAD, to be integrated and work together. It provides multiple numerical models, including MODFLOW, MODPATH, and PEST, which simplifies the management of the extensive data required for groundwater modeling and simulation. MODFLOW is a widely used numerical model developed by the USGS that can simulate groundwater flow in a three-dimensional heterogeneous and anisotropic medium. The GMS package of MODFLOW is suitable for both steady-state and transient simulations. Estimating groundwater levels and movement directions and predicting groundwater level changes connected to long-term climatic conditions. This software provides excellent capabilities for accurately representing and reflecting field reality. GMS presents the grid and conceptual approaches to build the MODFLOW numerical model. The grid approach works directly with the 3D grid and applies sources/sinks and other model parameters on a cell-by-cell basis.

In contrast, the conceptual model approach uses the GIS tools in the map module to develop a conceptual model of the modeled site. Once the conceptual model is complete, the grid is generated, the conceptual model is converted to the grid model, and all cell-by-cell assignments are performed automatically. This allows the location of sources and sinks, layer parameters (such as hydraulic conductivity), and other data necessary for the simulation to be defined at the conceptual model level (GMS user manual, 2018). Hence the Groundwater Modeling System GMS is the most advanced and comprehensive groundwater modeling software since it is a complete package that includes tools for every step of a groundwater simulation, including site characterization, model creation, calibration, and visualization.

2.4 Previous Studies Upon Groundwater Numerical Models

The numerical approach provides the hydraulic values, and the approximate solution reflects the numerical magnitudes at defined points in the space and time domains. The finite difference and finite element approaches are widely used and regarded as crucial tools in solving groundwater flow problems (Kareem, 2018). There have been many studies conducted on this topic. In this section, the focus is on the research that is most relevant to the research problem.

Al-kharabsheh (2000) utilized MODFLOW to simulate groundwater flow and assess the impact of five pumping scenarios on the upper aquifer in Jordan's Azraq basin. The objective was to determine the safe yield of future abstraction based on a groundwater balance analysis. The basin's safe yield was determined to be approximately 25×10^6 m^3. The division of the basin into three distinct zones, namely the well field, northern part, and eastern part, demonstrates that a decrease in pumping activities within a given sub-basin or zone, particularly in the well field, has a positive impact on the remaining two zones.

Al-Taee and Al-Sadiq (2003) conducted a research on the movement, direction, and expected quantity of groundwater in the Sulaivaney plain of Mosul governorate, Iraq. Their study involved the creation of a 3D numerical groundwater flow model using the MODFLOW package in GMS software. By mapping the water flow in the study area, the researchers were able to estimate the volume and direction of groundwater traveling through unconfined pathways. The study revealed that approximately 6708.97 ℓps of groundwater flowed within the Sulaivaney plain, while the volume of groundwater feeding the Mosul dam reservoir from the studied area was 6177.03 ℓps.

Jacek and Maciek (2004) developed a numerical model to simulate groundwater flow in a complex, multi-aquifer system in the Swidnica area of

southwestern Poland using the MODFLOW program integrated with GMS software to construct and calibrate the conceptual model. The model was based on data obtained from numerous boreholes and was calibrated under steady-state conditions. This study aimed to assess the influence of well abstractions of 613.43 ℓps on the groundwater level. The findings indicate that the abstractions had a negligible effect on the multi-aquifer system, as evidenced by the minimal decline in groundwater levels since it received recharges from various sources, including rainfall and infiltrations from several lakes. Furthermore, the aquifer system showed the potential to facilitate more well abstractions.

Hasan and Al-Taiee (2006) developed a numerical model of Mosul City using the groundwater modeling system's (GMS) built-in finite-difference flow code (MODFLOW). After a model calibration with the automated parameter estimation (PEST), the hydraulic conductivity values were subsequently determined. Maps of elevations, paths, and flow vectors were generated using the simulated data. The findings of the simulation process suggest that the rise in groundwater elevations in Mosul city can be attributed to an imbalance in the study area. This imbalance is caused by an increase in the sources of groundwater feeding, such as water supply, sanitary drainage, and rainfall, coupled with a decrease in hydraulic conductivity values (k) due to variations in soil properties. As a result, the aquifers' capacity to store and transmit excess water to the river will decrease, leading to consistently high groundwater elevations.

Rejani et al. (2008) created a 2-dimensional groundwater flow and transport model of the Balasore basin (India) using the Visual MODFLOW package to study how an aquifer reacts to different pumping strategies. The simulation model was successfully calibrated and verified. Five pumping scenarios were simulated using the validated model to determine the impact on groundwater levels under

the existing cropping conditions. According to the sensitivity analysis, the hydraulic conductivities and specific storage are less important to the Balasore aquifer system than river seepage, recharge from rainfall, and interflow.

Al-Hassoun and Mohammad (2011) conducted a study on an alluvial aquifer in Wadi Hada Al Sham, close to Makkah, Saudi Arabia. The researchers analyzed the annual fluctuations in the water table and formulated prognostications regarding future variations. A numerical groundwater model was developed utilizing MODFLOW. The calibration of the model was conducted to forecast the water table elevations resulting from pumping activities over a period of 5 years. The consistency between the model's predictions and historical data was established. The simulation findings demonstrate a reasonable decline in the elevations of the water table within the examined region.

Al-Mussawy and Khalaf (2013) applied the MODFLOW 2000 package integrated with the GMS 7.1 program to develop a three-dimensional model for effectively managing groundwater in the Karbala desert region of Iraq. The model was designed to simulate the flow of groundwater in the Dammam confined aquifer. According to the calibrated model, the hydraulic properties of the aquifer have been determined. The hydraulic conductivity values in the study region were found to be within the range of 0.65 to 50.3 m/day, while the recharge rate exhibited variability between 0.001 to 0.031 m/year.

Rahnama (2014) employed a three-dimensional finite-difference groundwater flow model to examine the diverse hydrogeological conditions and simulate the volumetric flow budget in the aquifer of Agartala and Khowai City, located within the State of Tripura, India. The modeling package MODFLOW in the Groundwater Modelling System (GMS) was used for steady-state analysis.

The findings indicate that the hydraulic conductivity values are 20.9425 and 25.9425 m/day, respectively.

Seeyan and Merkel (2015) presented a three-dimensional groundwater flow model for the Shaqlawa basin in Iraqi Kurdistan using Visual MODFLOW. This model was used to compare hydrogeological conditions and simulate the flow system's features under various stress scenarios for the unconfined Harrir Plain and semi-confined Mirawa basins. The top layer of the Mirawa Valley is Pliocene and Miocene formation. In contrast, the Eocene formation served as the lower layer for modeling. Hydraulic characteristics from observation and pumping wells calibrated the two-layer model under steady-state settings. The calibrated model generates a steady-state groundwater head distribution that matches observed data. The estimated standard error for Harrir and Mirawa were 1.06 m and 2.24 m, respectively. The normalized root mean square error (NRMSE) values for Harrir and Mirawa were calculated as 2.6% and 2.46%, respectively. The research indicated that raising the wells' pumping rate by 200% and 400% lowered the water table by 6 m and 18 m in Harrir Plain and 1 m and 2 m in Mirawa Valley. The average predicted rate of recharge was 0.258 mm/day.

Al-Areedhi and Khayyun (2019) developed a numerical model using the MODFLOW 2000 package integrated with the GMS software to simulate the flow of groundwater and the heads for the upper zone of Iraq's aquifers. The calibration results indicated that the hydraulic conductivities varied from 0.559 to 31.44 m/day. The calculated range for groundwater recharge rates ranged from 0 to 0.000488 m/day. Sensitivity analysis for the modeled aquifers demonstrated that the model is highly sensitive to changes in recharge amount. The water budget analysis shows a daily average groundwater recharge rate of 0.00014 m/day. Calculations of the water budget show that regional rainfall is the main

source of water for the upper zone of the Iraqi aquifer system, providing an average of 47.12 % of the total annual recharge, followed by surface water bodies and adjacent boundary areas, which together account for 37.46% and 15.41%.

Hadi and Alwan (2020) developed a conceptual model in GMS software to calculate how different feeding sources interact. Hydraulic conductivity and recharge were evaluated using PEST by applying the observed heads in the field. The model predicted a hydraulic conductivity of (0.958 - 24.932) m/day and a recharge rate of (5.3928×10^{-5}) m/day. Levels of groundwater measured in the field were consistent with those predicted by the model. Based on the results, Shatt Al-Diwaniya uses around (3466.37) m^3/day of groundwater while recharging about (2576.20) m^3/day. Therefore performing as a drain.

Jalut (2021) used the GMS mathematical model to simulate the southwest area of the Khanaqin province in Iraq, specifically to study the impact of increasing aquifer depletion rates on groundwater quality and the effect of a declining recharge rate on the groundwater in the region of study. Good agreement was found throughout the steady-state calibration phase between the model and measured groundwater heads. Five different cases were utilized to represent the unsteady situation. The study area's left side of the modeled region reported the highest drawdown value, while the area around Al-Wand River reported the lowest.

Ali (2021) used MODFLOW version 7.0.13 and updated MODFLOW 2005-NWT version 1.0.4, modified by US Geological Survey Techniques and Methods, to create a mathematical model for the area studied in the Western Iraqi desert. This area is roughly (49764) $Km^2$and is split into three sub-basins: Qasir Al-Ukhaider, Shbicha, and Al-Salman. As determined by applying a numerical model to the sub-basins, the Qasir Al-Ukhaider sub-basin has an average

horizontal hydraulic conductivity of (0.391) m/day and an average transmissivity of (64.3) m^2/day. The specific yield average is (0.0062). The drawdown is less than (0.31) m. The model recharge calculation result is (7.567×10^6) Mm^3/year and the static water level contour varied from (30 - 230) m. The average horizontal hydraulic conductivity in Shbicha is 0.761 m/day, whereas the average transmissivity is 60.9 m^2/day. The specific yield average is (0.0083). The drawdown difference is less than (0.25) m. Recharge is (17.812×10^6) Mm^3/year, the static water level contour was between 40 and 270 m. The average horizontal hydraulic conductivity in Al-Salman is (2.539) m/day, while the average transmissivity is (2158.4) m^2/day. The specific yield average is (0.012). The drawdown difference is less than (0.29) m. The recharge is (29.2×10^6) Mm^3/year, and the static water level contour was between 40 and 310 meters.

Ibrahim (2022) compiles a report on groundwater's current amount and quality for human consumption using groundwater modeling system (GMS) software for the Al-Mahaweel district in Babil Governorate, Iraq. The steady-state groundwater hydraulic modeling results were calibrated for hydraulic conductivity and recharge rates. One scenario simulated a drought condition with zero recharge and five operating wells pumping at a rate of 0.56 ℓps, resulting in a significant drop in the water table throughout the area. No well was determined to be drinkable throughout the research period according to WHO and Iraqi criteria, as measured by the water quality index.

2.5 Data-Driven and Machine Learning Techniques

Data-driven and machine-learning techniques have emerged as promising tools for modeling and predicting groundwater behavior. These techniques offer significant advantages over traditional modeling methods, including handling complex and high-dimensional data, identifying nonlinear relationships, and making accurate predictions. In this context, this topic has attracted increasing attention from researchers and experts who are exploring the potential of data-driven and machine-learning techniques in groundwater modeling. This section aims to review the recent state of research on this topic and provide insights into the benefits and challenges of using these techniques in groundwater modeling (Ye, 2015; Mohr et al., 2018; Marinósdóttir, 2019).

The machine learning algorithms may be split into two categories: supervised and unsupervised algorithms. During unsupervised learning, the algorithm derives patterns from data that has not been labeled or otherwise categorized. The objective is to get an understanding of the fundamental structure of the data because the model does not have a reference for the expected output data format. The model investigates the structure of the data so that valuable information may be extracted from it (Da Silva et al., 2017).

When both the input and output variables are clearly labeled, supervised learning algorithms are the preferred method. The idea is to learn patterns and correlations between variables based on prior experience (training data) and then apply that information to generate predictions on data that has not been seen or is unknown (test data). In supervised learning, two subcategories of data can be modeled: classification, in which a model aims to predict categorical or class labels, and regression, in which the models attempt to predict a continuous output.

Both classification and regression are examples of predictive modeling (Agatonovic-Kustrin and Beresford, 2000).

Machine learning provides various methods for predictive analysis, and artificial neural networks (ANNs) have become one of the most popular techniques. ANNs are used to model how the human brain processes data by simulating the brain's neuronal systems. The architecture of an ANN's comprises three primary layers: the input layer, which is responsible for receiving data from the external environment; the hidden layers, also known as the "black box," which perform the majority of the ANN's internal processing; and the output layer, which generates the network outputs. Nodes in an ANN are connected by weighted connections at each layer, and each node receives a weighted input that is passed through an activation function before being transmitted to the next layer. The activation function enables the artificial neural network to form complicated connections between the inputs and outputs of the network. Adjusting the activation function and the number of hidden layers can modify the degree of non-linearity in the data (Da Silva et al., 2017; Shahin et al., 2008).

2.6 Previous Studies on Machine Learning for Groundwater Potential Mapping

The impacts of climate change are expected to result in decreased rainfall and increased severity of droughts. Therefore, the importance of groundwater resources will increase, making groundwater potential mapping (GPM) a vital technique for supporting groundwater resource planning and exploration (Elbeih, 2015). GPM involves estimating groundwater storage in a particular region and measuring the probability of discovering groundwater or predicting the location of the highest borehole yields (Díaz-Alcaide and Martínez-Santos, 2019).

Two basic techniques used in GPM are expert-based decision systems and machine learning. The multi-influencing factor method, which belongs to the expert-based techniques, has been utilized with the analytical hierarchy processes (Mohammadi-Behzad et al., 2019; Al-Djazouli et al., 2021). Expert-based decision approaches have been in use for a long time, while machine learning is a newer technique. Machine learning classification uses artificial intelligence capabilities to discover complex associations between explanatory factors that may otherwise go unnoticed, which is a crucial difference between machine learning and expert techniques. Therefore, machine learning is well suited for mapping complicated spatially distributed variables, such as the occurrence of groundwater (Gómez-Escalonilla et al., 2022).

Ozdemir (2011) used a binary logistic regression combined with GIS to create a groundwater spring potential map GPM of the Sultan Mountains in central Turkey. Field investigations identified 440 springs in the study area. This study examined 17 spring-related factors: geology, relative permeability, land use/land cover, precipitation, elevation, slope, aspect, total curvature, plan curvature, profile curvature, wetness index, stream power index, sediment transport capacity index, distance to drainage, fault, drainage density, and fault density map. The validation of the spring potential map was conducted through the utilization of directly observed springs. The receiver operating characteristics ROC was used to calculate model accuracy. The model ROC was determined to have an area value of 0.82. These results indicate that the model effectively forecasts the potential of spring in the examined area. The map illustrating the potential of groundwater springs in various classes, namely extremely low, low, moderate, and high, indicates that the respective areas covered by these classes are 105.586, 74.271, 101.203, and 90.05 Km^2, respectively.

Naghibi et al. (2016) used supervised classification methods to create groundwater spring potential maps for the Koohrang Watershed, Iran, using three machine learning models: boosted regression tree (BRT), classification and regression tree (CART), and random forest (RF). A total of thirteen hydrological-geological parameters were considered because they potentially affect spring distribution. Factors include altitude, topographic wetness index, slope length, plan curvature, profile curvature, distance to rivers, distance to faults, land use, lithology, drainage density, and fault density. Then, CART, RF, and BRT algorithms were used to simulate and map groundwater spring potential. A total of 864 springs were identified, of which 605 about 70% were utilized for training spring potential mapping, while the remaining 259 about 30% were used for model validation. The calculated area under the curve (AUC) using ROC values to measure the accuracy of the three models' predictions shows that the BRT model has an AUC of 0.8103. In contrast, the CART and RF models have AUC values of 0.7870 and 0.7119, respectively. Consequently, the BRT model was shown to be superior to the CART and RF models for predicting the positions of springs. Geospatially integrated BRT, CART, and RF techniques proved to help generate the spring potential map (SPM) with reasonable accuracy.

Naghibi et al. (2017) aimed to assess the potential of groundwater through the analysis of spring sites; this research proposes to use support vector machine (SVM), random forest (RF), and genetic algorithm-optimized random forest (RFGA) techniques to generate the GPM for Ardebil Province, Iran. Fourteen parameters were provided, including variables derived from DEM, rivers, faults, land-use, and lithology. Of the total 842 spring sites, 589 of them, which accounts for 70%, were utilized for model training. The results suggest that using RFGA and RF techniques yielded better results with an AUC of 0.856 and 0.846, respectively, compared to the alternative SVM model with an area under the ROC

curve of 0.786. Furthermore, The results suggest that altitude, topographic wetness index (TWI), and slope angle are important factors to consider when assessing groundwater.

Falah and Zeinivand (2019) employed the use of Frequency Ratio (FR) and Weights of Evidence (WoE) methodologies within a Geographic Information System (GIS) framework to develop the (GPM) of Khorramabad, located in the Lorestan Province of Iran. A geographic database was established, which incorporated altitude, slope angle, slope aspect, plan curvature, topographic moisture index, land use, drainage density, fault density, geology, and soil maps. Subsequently, the generated maps were classified into four distinct categories: low, moderate, high, and very high. The created maps were evaluated in the final phase using the area under the receiver operating characteristic (ROC) approach referred to as AUC. The study found that the produced FR and WoE maps had areas under the ROC curve (AUC) of 0.891 and 0.882, respectively. These results suggest that both models can accurately simulate groundwater potentiality in the research region.

Al-Fugara et al. (2020) evaluated the predictive performance of five distinct machine-learning models in the context of potential groundwater mapping in the Wadi az-Zarq watershed of Jordan. The study employed five models, namely Random Forest (RF), Boosted Regression Tree (BRT), Support Vector Machine (SVM), Mixed Discriminant Analysis (MDA), and Multivariate Adaptive Regression Spline (MARS). The present study employed algorithms to examine the spatial distributions of twelve hydrological-geological conditioning factors that play a crucial role in determining the precise location of groundwater springs. The assessment of prediction accuracy for the five distinct models was conducted

using the area under the curve (AUC) of the ROC curve. The results indicated that the MDA of (83.2%) accuracy performs better than other trained models.

Moghaddam et al. (2020) examined the effect of sample size on the accuracy of various individual and hybrid models, adaptive neuro-fuzzy inference system (ANFIS), ANFIS-imperial competitive algorithm (ANFIS-ICA), alternating decision tree (ADT), and random forest (RF) to model groundwater potential for the Hableh-Roud basin extends from Tehran Province to Semnan Province in Iran. Fifteen unique geo-environmental elements were used as independent variables. The AUC of the ROC curve and the true skill statistic (TSS), which is a metric that views a model's true success, were used to evaluate the performance of models. Results indicated that sample size affected the performance of four machine learning algorithms, while RF was less sensitive to sample size. Since the validation results revealed that RF with AUC ranged from 0.9074 -0.9632 due to the change of sample size and TSS = 0.79–0.85 had the highest efficiency for each of the four datasets, followed by ANFIS-ICA, ADT, and ANFIS. Also, the slope, lithology, and distance from nearby faults were the most critical determinants of the spring's impact on the groundwater potential model.

Nguyen et al. (2020) used an advanced ensemble machine learning model (RABANN) that combines Artificial Neural Networks (ANN) with RealAdaBoost (RAB) ensemble approach; the main objective of this research is to evaluate the groundwater potential of the DakNong province, Vietnam. This work utilized twelve conditioning parameters and well-yield data to generate the training and testing datasets for the ensemble RABANN model's creation and validation. Using the area under the Receiver Operating Characteristics (ROC) curve (AUC) and numerous statistical performance indicators, the ensemble RABANN model was

validated and compared to the single ANN model. Both models performed well in the training phase of evaluating groundwater potential (AUC 0.7). However, in the validation phase, the ensemble model outperformed the single ANN model (AUC = 0.776 vs. 0.699). This indicated that the RAB ensemble approach improved the performance of the ANN model. By making minor modifications to the input data, the ensemble-developed model may be applied to map the groundwater potential of other regions and countries to improve water resource management.

Phong et al. (2021) proposed three unique ML hybrid models. These models were the modified RealAdaBoost MRAB-FT, bagging BA-FT, and rotation forest RF-FT. All models are functional tree (FT) base classifiers that can deal with binary and multi-class target variables for the GPM modeling in the DakLak region of Vietnam. The following variables serve as the model's input data: topographic parameters, including slope angle, slope aspect, elevation, and curvature, are crucial for GPM because they impact how much groundwater infiltrates, in addition to the hydrological aspects of GPM, such as flow accumulation, rainfall, topographic wetting index (TWI), and closeness to rivers. The weathering crust, distance to fault, and hydrogeologic formation thematic maps for the geology-related factors were obtained and created. This study employed a threshold value greater than 2 ℓps for selecting training and testing datasets for groundwater classification as groundwater and non-groundwater. Standard statistical methods assessed how well the created models performed (PPV, NPV, SST, SPC, Kappa, RMSE, and ROC). The analysis' findings demonstrated that all the unique hybrid models created had high predictive abilities, but the model MRAB-FT performed the best in recognizing groundwater potential zones.

Saha et al. (2021) assessed the groundwater potential maps (GPMs) in the Bazar Block of Birbhum District, India, using four GIS-based machine-learning algorithms (MLA), including predictive neural network (PNN), decision tree (DT), Nave Bayes classifier (NBC), and random forest (RF). Twelve influencing factors: elevation, slope, drainage density (DD), topographical wetness index, geomorphology, lineament density, rainfall, geology, pond density, land use/land cover (LULC), geology, and soil texture were employed using GIS analysis. ROC curves determined each GPM's prediction accuracy by calculating the AUC. The PNN, RF, DT, and NBC models achieved accuracies with an AUC of 0.801, 0.783, 0.752, and 0.692, respectively. The ROC values indicate that the four applied models offer satisfactory results for GWP mapping in this location, but the PNN model performs the best.

Sarkar et al. (2022) have devised ensemble machine learning (EML) algorithms to analyze the (GPM) in the Teesta River basin of Bangladesh. The algorithms under consideration are random forests (RF) and random subspaces (RSS). This study utilized twelve distinct parameters included soil types, topographic wetness index (TWI), topographic roughness index (TRI), aspects (directions), stream power index (SPI), slope, curvature, elevation, distance to the road, land use land cover (LULC), rainfall, sediment transport index (STI). The groundwater inventories were comprised of 220 well points collected from various sources. The RSS method, with an AUC value of 0.892, was found to be the most effective representation model for modeling groundwater potentiality. This was closely followed by the RF method, which achieved an AUC value of 0.86. In the context of groundwater potentiality models, it has been observed that the RSS model has superior performance compared to the RF model because a higher AUC value indicates a greater accuracy in predicting the model's output.

Tamiru et al. (2022) assessed the effectiveness of artificial intelligence (AI) in the Fincha catchment, Abay, Ethiopia, using geospatial analysis and GIS platforms to prospect possible groundwater zones. In this work, a multi-layer perceptron (MLP) structure was used. The eight input parameters, including rainfall, LULC, lineament density, drainage density, geology, slope, soil, and geomorphologic units, were sorted and weighted in an Artificial Neural Networks (ANN) training model in the input nodes. Groundwater potential zones of five classes (Very High, High, Moderate, Low, and Very Low) were identified using AI techniques. The AI models revealed an agreement with 96% of the identified groundwater potential zones. Ultimately, it is stated that the use of the ANN model serves as a valuable means of identifying potential groundwater zones in instances where the expenses associated with direct field investigation are not feasible.

2.7 Significance of the Research Gap

The introduced literature review highlights the importance of groundwater as a crucial source of freshwater that sustains various life aspects. However, the growing international concern for the sustainability of groundwater resources necessitates effective planning and management of groundwater resources, which can be achieved through groundwater flow modeling. While many studies have been conducted on groundwater flow for limited regions within the study area, no one has attempted to simulate the Dammam groundwater confined aquifer within the studied area, the main source of stored water in Iraq's western and southern deserts. The identified research gap highlights the need for a study that models this area to develop an overall perspective of the controlling parameters, such as hydraulic conductivity and recharge characteristics.

The accurate estimation of groundwater potential zones over large-scale regions is a challenging task, mainly due to the absence of accurate long-term

geotechnical and hydrogeological data. The utilization of physically-based and numerical models is restricted by numerous factors that impact the groundwater balance as it requires a significant amount of data and difficulty accessing accurate data. Furthermore, in order to enhance the required data for the development of groundwater models, certain research necessitates financial support to conduct experiments. This research gap highlights the significance of developing alternative strategies for building groundwater models. This study aims to construct a novel hybrid model for groundwater potentiality mapping using knowledge-driven technologies integrated with data-driven modeling methodologies to accurately estimate groundwater potential maps (GPM) with a minimum amount of field data. The present research investigated the spatial distributions of seven conditioning factors that determine the location of groundwater potential zones. These factors were sorted and weighted in machine learning models and used as the input layer, while the pumping rate was evaluated as the output layer.

This study is significant because it uses a new hybrid-supervised classification model by applying five ML algorithms represent five different learning strategies employed by MLAs: Ensemble Boosted Trees -logic-based learners; Naive Bayes (NB) -statistical learning algorithms; Support Vector Machines (SVM); Multi-Layer Perceptron (MLP)- Artificial Neural Networks (ANN); and k-Nearest Neighbours (kNN)– instance-based learners to construct GPM that contribute in the development of strategies for the sustainable management of water resources. The developed hybrid model can analyze the dynamic relationship between groundwater potentiality and influencing factors, providing a more accurate prediction of groundwater potentiality. This study is the first of its kind in Iraq and can inform decision-makers and interested organizations on the sustainable management of water resources in the study area.

Chapter Three

Groundwater Flow Simulation Using Numerical Modeling

3.1 Introduction

In recent years, the acceleration and improvement of mathematical models of groundwater have been achieved through the application of various software. The numerical groundwater model is a mathematical simulation tool founded on the conceptual model generalization of groundwater systems. Its purpose is to describe the quantitative relationship between the parameters and metrics of the groundwater system. According to the verification and identification of the mathematical model, further validation of groundwater system adaptability can be carried out, and a deeper understanding will be achieved (Doherty et al., 2010).

Therefore, numerical models are considered a powerful tool in evaluating water resource management scenarios because they allow decision-makers to compare alternative measures and make management decisions to achieve efficiency goals without violating certain constraints (Olayinka, 2013). An essential step in modeling is selecting appropriate software that considers all possible options. Researchers have often adopted the MODFLOW program developed by the US Geological Survey or MODFLOW package integrated with Groundwater Modeling System (GMS) software. The Groundwater Modeling System (GMS) is a comprehensive graphical user environment for groundwater simulations developed by Aquaveo, LLC in Provo, Utah (GMS User Manual, 2018). The initial step for groundwater modeling involves constructing a conceptual model that accurately represents the groundwater system, including its geological and hydro-geological units, system boundaries and boundary types, and aquifer parameters. Spatial analysis techniques are used to prepare data for the conceptual model's input. Once the conceptual model is constructed, the next

step is to map the conceptual model into the MODEFLOW simulation and check for errors applied to run the model correctly. Figure (3.1) illustrates the current study methodology of groundwater modeling.

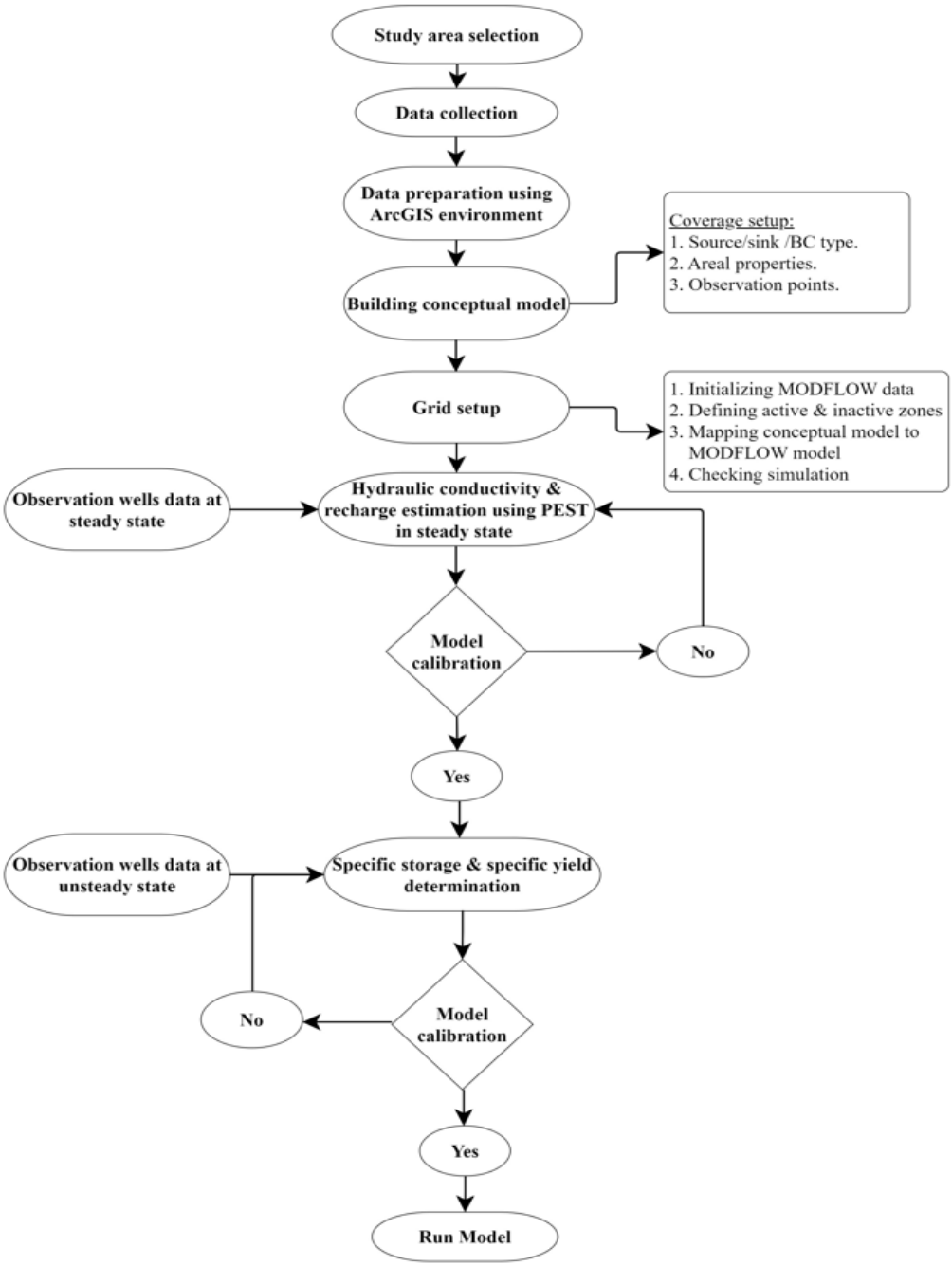


Figure (3.1): Methodology for simulating groundwater flow in Dammam Confined Aquifer.

3.2 Study Area Location

The study area of about 76732.71 Km^2 is located in Najaf and Muthanna governorates in Iraq's southern desert zone, as shown in Figure (3.2). The geographic coordinates for the study area are between longitudes 44° 19 '07.68" to 46° 33' 26.55" East and between latitudes 29° 06 '15.74" to 32° 19'55.97" North. The Southern Desert is related to the geographical region located in the southern part of Iraq. It is located west of the Euphrates River and Shatt Al-Arab, extending further to Saudi Arabia and Kuwait. Wadi Al-Khair serves as a natural boundary separating it from the Western Desert. The southern region of the Mesopotamian Zone covers the eastern and, to some extent, the northern areas. The remaining areas are attributed to the Salman zone of the stable shelf.

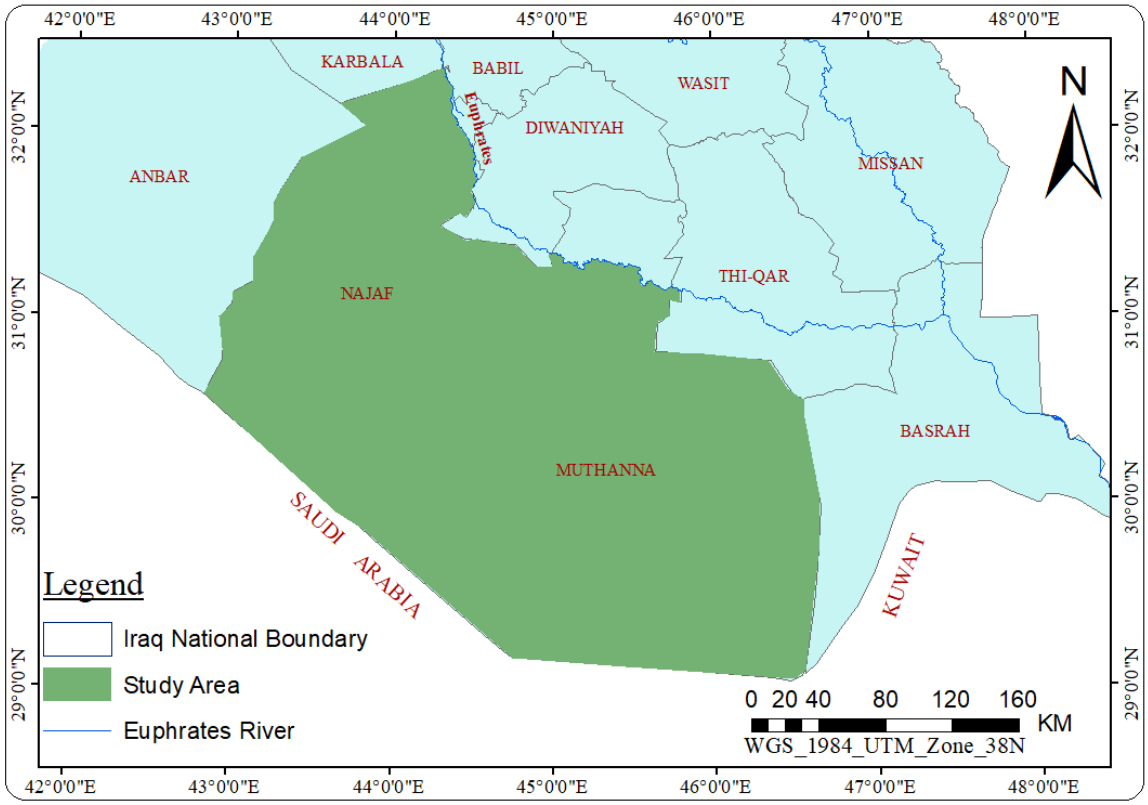


Figure (3.2): Location of the study area.

3.3 Governing Equation for the Flow of Groundwater

The governing equation for flow in porous media is traditionally derived by referring to the flux of water through a porous material cube. This cube should be large enough to be properties representative of the medium yet small enough that the change in the head within the volume is relatively small (Anderson et al., 2015). The porous material cube, also known as a representative elementary volume or REV, has a volume of ΔxΔyΔz, as shown in Figure (3.3).

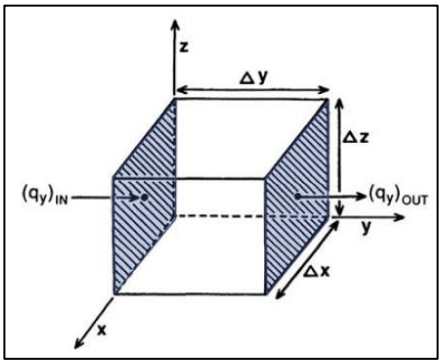


Figure (3.3): Flow components in the (REV) (Anderson et al., 2015).

The governing equation can be derived by combining the equation of continuity with the constitutive relation of Darcy's law (Anderson and Woessner, 2015). The continuity equation can be obtained by applying the elemental Cartesian fixed control volume concept, as shown in Figure (3.3). The flow is assumed to be uniform in time and space. The flux (q) through the REV, which is a vector that three components can express, q_x , q_y, q_z

$$q= q_x \ i_x + q_y \ i_y + q_z i_z \quad (3.1)$$

Where $i_x , i_y , and \ i_z$ are unit vectors along the x, y, and z axes. A representative elementary volume water balance is necessary to maintain mass conservation.:

$$\text{outflow - inflow = storage change} \quad (3.2)$$

The inflow occurs through the REV face ΔxΔz and is equal to $(q_y)_{IN}$, while the outflow is equal to $(q_y)_{OUT}$. The difference between outflow and inflow along the y-axis is written as:

$$outflow - inflow = \frac{(q_y)_{OUT} - (q_y)_{IN}}{\Delta y} (\Delta x \Delta y \Delta z) \quad (3.3)$$

The conversion of Equation (3.3) from difference notation to derivative notation results in the following equation:

$$outflow - inflow = \frac{\partial q_y}{\partial y} (\Delta x \Delta y \Delta z) \quad (3.4)$$

Similarly, the same expression in equation (3.4) is written for the change in flow rate along the x- and z-axes. Since the total change in flow rate is equal to the change in storage:

$$\left(\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z}\right)(\Delta x \Delta y \Delta z) = \Delta\ storage \quad (3.5)$$

It's essential to allow for the possibility of a sink or water source within the REV. The expression denoting the volume of flow rate that results from sources and sinks is expressed as (W^* ΔxΔyΔz), where W^* is positive if it is a water source. As a source of water, W^* is subtracted from the left-hand side of Equation (3.5), considering the negative sign preceding the inflow term in Equation (3.2). The resulting expression is as follows:

$$\left(\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} - W^*\right)(\Delta x \Delta y \Delta z) = \Delta\ storage \quad (3.6)$$

By considering the right-hand side of equation (3.6), the change in storage is defined by the specific storage (Ss, [L^{-1}]), which is the volume of water released from storage per unit change in the head (h) per aquifer unit volume:

$$S_s = -\frac{\Delta V}{\Delta h\, \Delta x \Delta y \Delta z} \tag{3.7}$$

The rate of change in the amount of stored water within REV is:

$$\frac{\Delta V}{\Delta t} = -S_s \frac{\Delta h}{\Delta t} \Delta x \Delta y \Delta z \tag{3.8}$$

By combining equations (3.6) and (3.8) and dividing by $\Delta x \Delta y \Delta z$ the final form of the water balance equation will be yielded:

$$\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} - W^* = -S_s \frac{\partial h}{\partial t} \tag{3.9}$$

Expressing the governing equation in terms of the head variable is desirable as it is a measurable quantity that can be observed in wells. According to Darcy's law, the variables q and grad h are vector quantities, while K represents the hydraulic conductivity. The components of the specific discharge vector, denoted as q, can be expressed as:

$$\begin{aligned} q_x &= -K_x \frac{\partial h}{\partial x} \\ q_y &= -K_y \frac{\partial h}{\partial y} \\ q_z &= -K_z \frac{\partial h}{\partial z} \end{aligned} \tag{3.10}$$

where K_x, K_y, and K_z are the components of the hydraulic conductivity K and $\frac{\partial h}{\partial x}\ \frac{\partial h}{\partial y}\ \frac{\partial h}{\partial z}$ are the components of the gradient of head h.

Equation (3.10) is substituted into Equation (3.9) to give the differential general governing equation for (3D) transient groundwater flow for heterogeneous and anisotropic conditions as:

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) = S_s \frac{\partial h}{\partial t} - W^* \quad (3.11)$$

The variable h, is the dependent variable, while x, y, z, and t are the independent variables and K_x, K_y, K_z Ss, and W^* are parameters. It can be noticed from equation (3.11) that the hydraulic conductivity can vary with direction, x, y, and z. The placement of K within the differential signs allows the heterogeneity (spatial variation) in hydraulic conductivity (Anderson et al., 2015). Analytical solutions of the equation (3.11) are rarely possible, so various numerical methods could be used to obtain approximate solutions (Doherty 1998, 2010). One such approach is the finite difference method, where a finite set of discrete points in space and time replaces the equation (3.11). The partial derivatives are replaced by terms computed from the differences in head values at those points. The process leads to simultaneous systems of linear algebraic difference equations; their solution provides values of the head at specific points and times (Dohrty, 2004). These values represent an approximation of the time-varying head distribution that would result from an analytical solution of the partial differential equation of the flow.

3.4 Data Collection

To accurately model groundwater flow, it is essential to acquire multiple types of data, such as topographical, geological, hydrological, and climatic information, as Rapantova et al. (2017) outlined. This data is crucial for constructing groundwater models. The limited availability of such data may pose a significant challenge to groundwater modeling and investigations, as Hogeboom et al. (2015) noted. The study area's data were obtained from the General Committee for Groundwater (GCGW)/ Ministry of Water Resources of Iraq, utilizing the most recent survey, especially in the case of calibrating the generated

numerical model. The necessary data for modeling any given area using GMS software can be briefly outlined as follows:

i- Geographical information involves the specific location of the area under investigation.

ii- Topographic data: The present investigation employs topographic data obtained through the processing and analysis of a Digital Elevation Model (DEM) of STRM classification. This model can provide information regarding terrain elevations and contour lines with an accuracy of 30 meters and was processed using ARC-GIS software.

iii- The hydrology and hydrogeological data include parameters such as hydraulic conductivity, recharge rate, storage coefficient, and groundwater elevations.

iv- The geological data required includes the distribution and type of aquifers, their thickness, soil type, and a stratigraphic column that displays the formations in order of occurrence from the oldest to the most recent.

v- Historical records of meteorological datasets: This study utilizes historical meteorological data sets obtained from the Iraqi Meteorological Organization and Seismology (IMOS, 2021) to provide the climate parameters, mainly monthly rainfall values for a period of 47 years (1974-2021) in the monitoring stations of Najaf and Muthanna governorates.

vi- The dataset related to wells comprises data regarding the geographical coordinates of wells (latitude and longitude), groundwater heads, rate of pumping, and depth of wells.

3.5 Topography of the Study Area

The digital elevation model (DEM) generated by the United States Geological Survey (USGS) provides an extensive dataset with detailed information on topographical elevation and contour lines. Moreover, the analysis and manipulation of this image have the potential to provide a substantial quantity of supplementary data in which the data files consist of a rasterized representation of topographic data and are comprised of a set of ground locations that have been surveyed at regular horizontal intervals, as stated by Al-Mussawy and Khalaf (2013). The (DEM) used in this study was acquired from the website (http://earthexplorer.usgs.gov). It has an accuracy of 30 meters and falls under the Shuttle Radar Topography Mission (SRTM) category. The topographic elevation map was determined using ArcMap 10.7.1 software by employing the raster processing tools of the surface spatial analyst tool to produce the investigation site's topographic elevation map, shown in Figure (3.4).

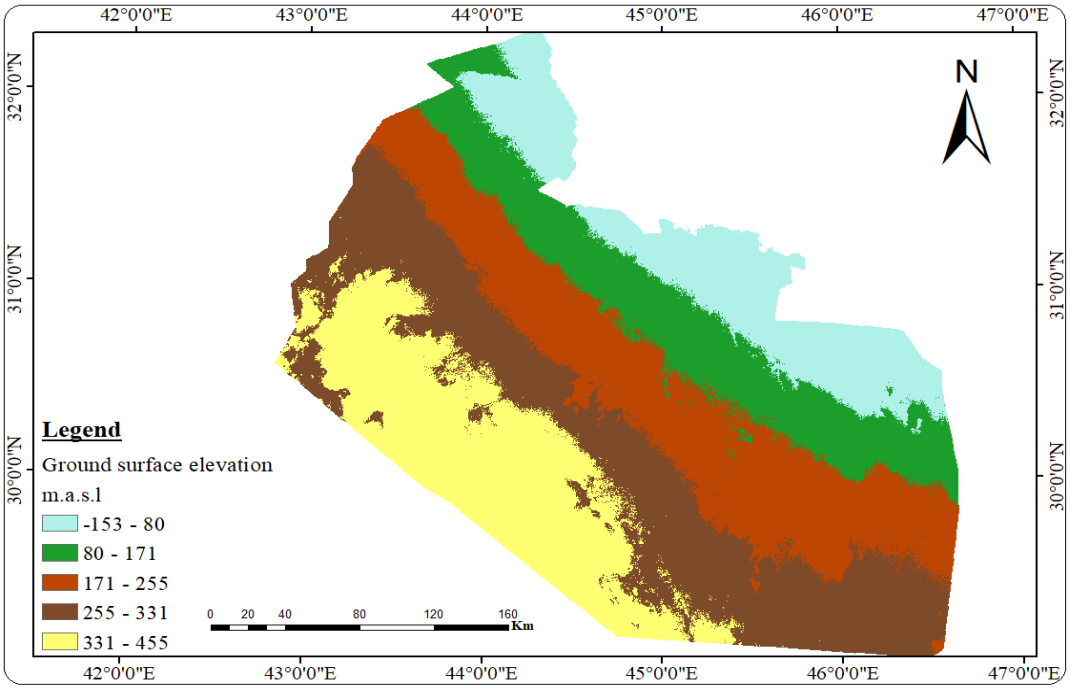


Figure (3.4): Topographic map of the study area.

The ground surface in the southern desert generally rises from the Euphrates River in the east towards the southwest in the direction of Saudi Arabia by about 50 m every 10-15 km (Sissakian et al., 1994). The minimum ground surface elevation is (-153) m towards the Euphrates River since the ground elevation started from sea level in Faw district South Basra, and the maximum elevation is (455) m in the direction of Saudi Arabia (southwest).

From a topographical perspective, most of the Southern Desert, particularly its southern and southwestern regions, can be characterized as a rolling plain entirely covered in residual soil containing angular fragments of limestone and chert. This terrain is further marked by numerous shallow, basin-like depressions and sharp escarpments, particularly along its southwestern boundary. The topography of the northern and northeastern regions of the Southern Desert is characterized by flat to slightly rolling terrain, which is intersected by shallow valleys that flow into marshes and depressions that bordering the Euphrates River. The topography of the eastern and southeastern regions is characterized by gentle hills and valleys, with mainly sandy to gravelly soils that the Dibdibba Formation overlies.

The valleys in this region are characterized by their shallow and wide nature, draining to the zones adjacent to the Euphrates River. A belt of sand dunes of varying widths can be found from Samawa, extending along the northeastern side of the Southern Desert up to the borders of Iraq and Kuwait. The Southern Desert comprises sedimentary strata that exhibit a gentle dip and are primarily composed of carbonate sequences. The geological layers exhibit a slight inclination towards the Mesopotamian Plain. The northwest-southeast direction of these strata in the area indicates that they are sub-parallel to the path of the Euphrates River (Al-Jiburi and Al-Basrawi, 2009).

3.6 Hydrogeological Characteristics of the Study Area

The field of groundwater hydrogeology focuses on the occurrence, distribution, movement, and geological interactions of groundwater. To effectively recognize the importance of relevant flow processes, it is essential to characterize the hydrogeological conditions of the area under study adequately. Without addressing proper site characterization, it becomes unfeasible to select an appropriate model or develop a reliably calibrated one. Aquifer parameters are only available at specific locations where the survey work was carried out; therefore, the Kriging interpolation method was used to generate the required data, such as groundwater contour map, aquifer thickness, and hydraulic conductivity. The aquifer hydrogeological characterization will be discussed in the following sections:

3.6.1 Groundwater Flow Mapping in the Study Area's Dammam Aquifer

To create a groundwater flow map and correctly interpret groundwater flow lines, it is necessary to make the implicit assumption that flow is perpendicular to the lines of equal hydraulic heads. That flow moves in the direction of decreasing static water levels (Keith et al., 2002). However, it is important to note that the direction of groundwater flow can vary depending on the specific conditions of the aquifer and the surrounding geology. It is important to note that groundwater flow is not always in the direction of decreasing static water levels. In some cases, such as where there is a pumping well, the groundwater flow may be towards the well. Therefore, while the assumption that groundwater flow is perpendicular to the lines of the equal hydraulic head may be a useful simplification in some situations, it is important to carefully consider the specific conditions of the aquifer and the surrounding geology to interpret the direction of groundwater flow accurately. Consequently, a groundwater flow map is constructed using the Arc-

GIS tools and raster interpolation based on available groundwater head data for the numbers of wells distributed within the study area using Kriging interpolation. The technique of Kriging interpolation gives the best results if the data is normally distributed (Al-Mussawy and Khalaf, 2013). Therefore, the normality test was conducted using different testing procedures to determine whether groundwater piezometric heads followed a normal distribution. First, a descriptive statistics summary of the groundwater heads is presented in Table (3.1).

Table (3.1): Summary of descriptive statistics for the groundwater wells located in the study region.

Variable	Descriptive	Statistic	Units
Groundwater Heads (m)	Mean	137.752	m
	Maximum	231.50	m
	Minimum	43.00	m
	Median	142.600	m
	Standard Deviation	55.930	m
	Variance	1328.160	m^2
	Kurtosis	-0.998	---
	Skewness	-0.162	---

A normal probability plot, also known as a quantile-quantile (Q-Q) plot, shows the data distribution against the expected normal distribution, used to determine whether or not a set of data follows a normal distribution. In this type of plot, the observation data points should lie approximately on a 45-degree straight line for normally distributed data. If the data is non-normal, the points form a curve that deviates significantly from a straight line. The observations of groundwater piezometric heads were tested using a Q-Q plot. The results indicated that these observations are normally distributed since the points lie close to the

straight diagonal line with some minor deviations, as shown in Figure (3.5). Based on this plot, we could safely assume that this data set is normally distributed.

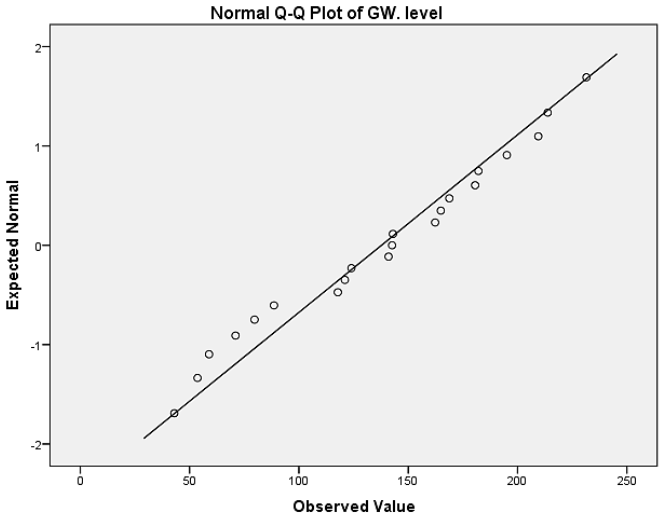


Figure (3.5): Normal Q-Q Plot of groundwater piezometric head data.

In addition, a graphical check for normality is performed using the histogram as presented in Figure (3.6).

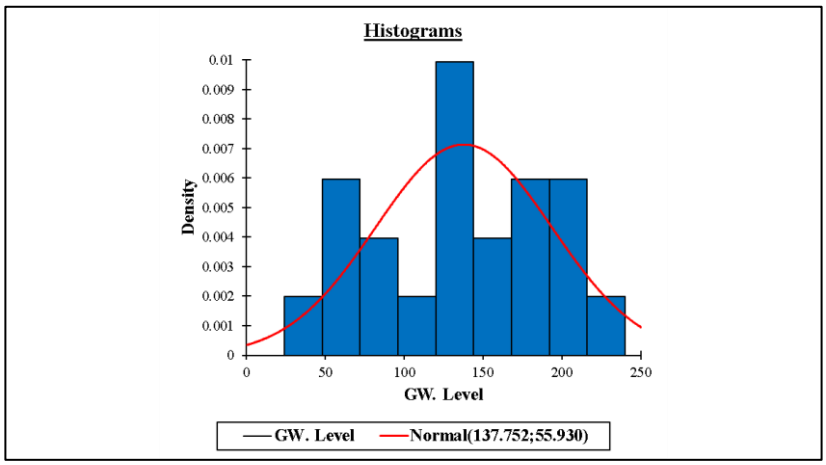


Figure (3.6): Testing for normality using a histogram.

Furthermore, the test for normality can be achieved by applying a goodness-of-fit method (i.e., chi-square test or Kolmogorov-Smirnov test). The results of these two tests are shown in Table (3.2).

Table (3.2): Normality test using chi-square and Kolmogorov-Smirnov tests.

Variable	Kolmogorov-Smirnov			Chi-square				
	D_{Max}	p-value	α	Observed Value $\chi^2_{cal.}$	Critical value $\chi^2_{tab.}$	Degree of freedom	p-value	α
GW. Heads	0.098	0.975	0.05	5.810	14.067	7	0.562	0.05

The groundwater piezometric head follows a normal distribution since the computed p-value is greater than the significance level (α=0.05). Accordingly, the Kriging interpolation method can be used to generate the groundwater piezometric heads contour map for the study area, as shown in Figure (3.7).

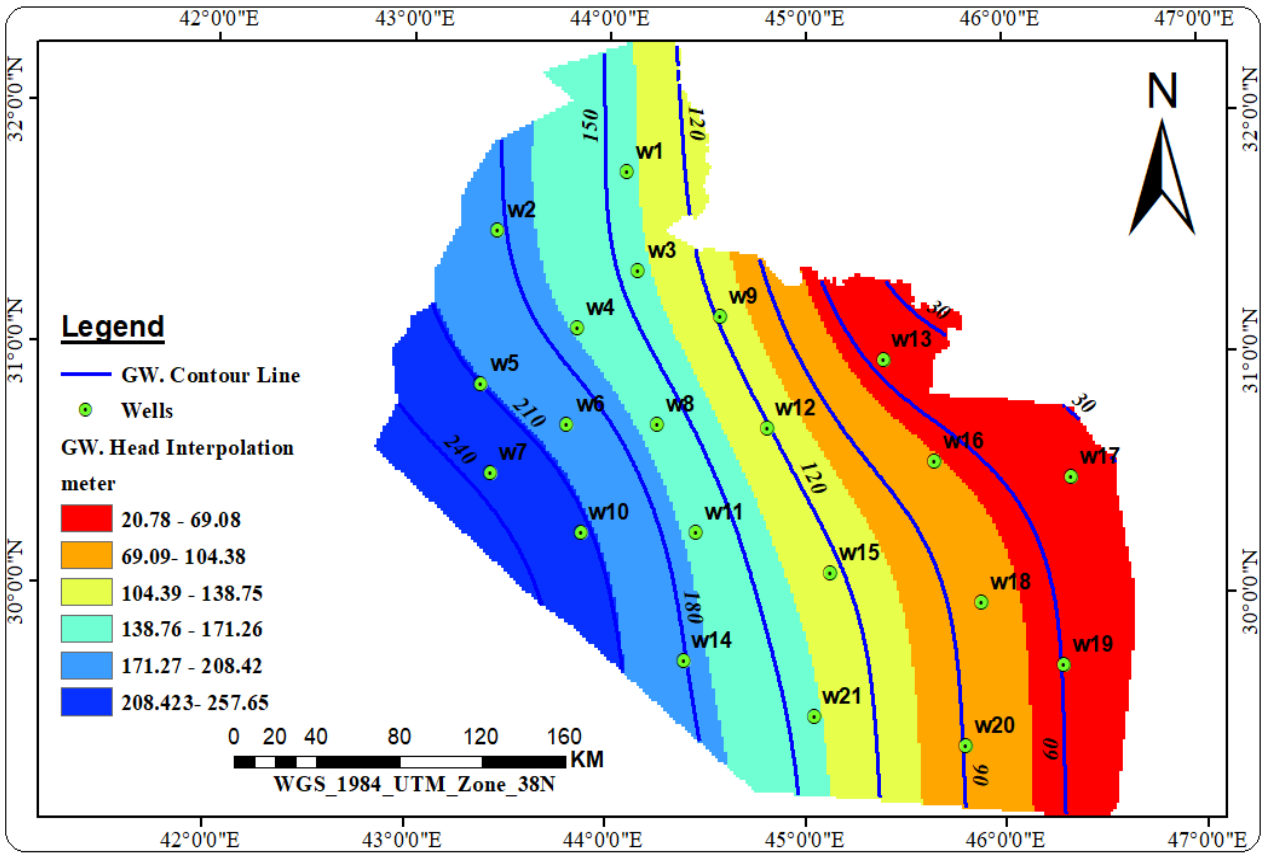


Figure (3.7): Dammam groundwater piezometric head contour map.

3.6.2 Hydraulic Conductivity Distribution

The hydraulic conductivity parameter has significant importance in the process of modeling the layers of an aquifer. In the context of GMS MODFLOW 2000 Simulation, the utilization of the Layer-Property Flow (LPF) package involves the specification of hydraulic conductivity as horizontal hydraulic conductivity along model rows. Subsequently, MODFLOW relies on the hydraulic conductivity (K) data of the aquifer and the thickness (b) of the layer to derive the conductance (transmissivity) of each cell, which is computed as T=Kb. The hydraulic conductivity values obtained from pumping tests conducted by the General Committee for Groundwater in Baghdad, Iraq, were extrapolated throughout the groundwater system using Kriging interpolation with the spatial analyst tools of ArcMap 10.7.1 software. This process was illustrated in Figure (3.8) and facilitated the assignment of appropriate values to each zone within the study area, which served as the initial values for the parameter estimation process.

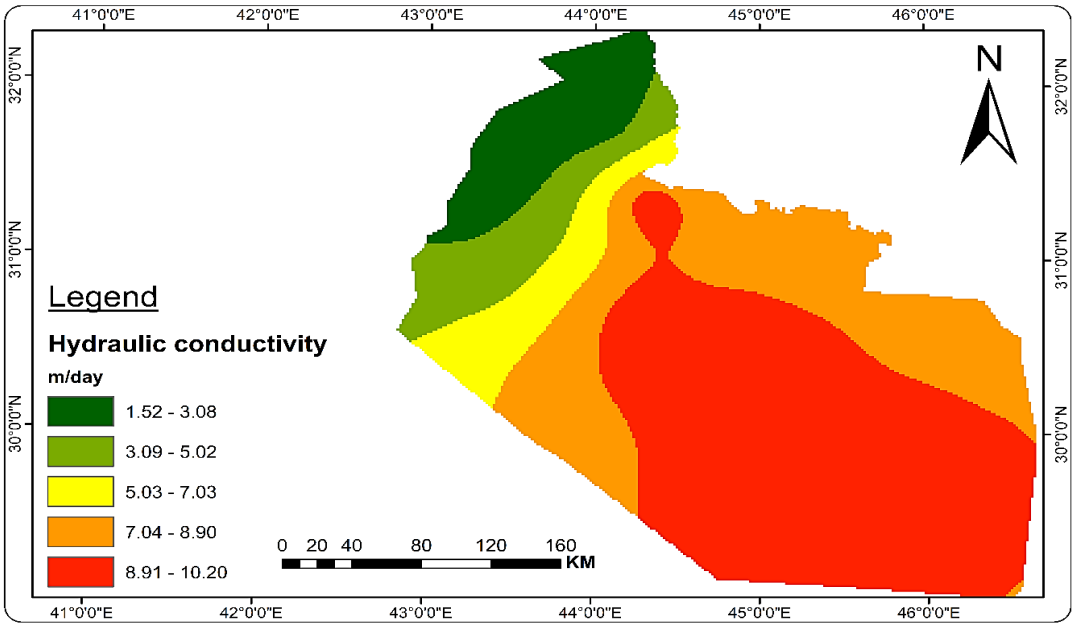


Figure (3.8): Hydraulic Conductivity Spatial distribution of the study area's Dammam aquifer.

3.7 Geological Characterization of the Study Area

The studied area is composed of different geological formations belonging to different ages. The geology of the southern desert, extending from Najaf to Nasiriyah and passing through the cities of Samawa along the Euphrates River with a width of about (100-150) km, has been significantly altered. Much of the mapped upper member of the Dammam formation was covered by the basal conglomerate of the Euphrates formation (Sissakian and Fouad, 2015)**.** The geological characterization of the studied area can be summarized as follows:

3.7.1 Type and Extent of Aquifer in the Study Area

Table (3.3) presents a list of geological formations that can be observed in the southern desert arranged in chronological order from oldest to youngest with corresponding ages according to (Jassim and Al-Jiburi, 2009).

Table (3.3): Geological formations in the southern desert and their ages within the study area (Jassim and Al-Jiburi, 2009).

Formation	Age
Umm Er Radhuma	Middle-Late Paleocene
Dammam	Early – Late Eocene
Euphrates	Early Miocene
Ghar	Early Miocene
Nfayil	Middle Miocene
Dibdibba	Pliocene – Pleistocene
Zahra	Pliocene – Pleistocene

The main Iraqi geological map of Scale 1:1000000 created by the Geological Survey and Mining Company (GEOSURV, 1986) was reprocessed and digitalized using Arc-GIS tools to create the geological map for the study area, as shown in Figure (3.9), which illustrates the distribution of geological units within the study area. The southern desert comprises sedimentary rocks, mainly carbonate sequences and friable clastics. From the oldest onwards, the main investigated water-bearing aquifers underlying the southern Iraqi desert are Hartha, Tayarat, Umm Er Radhuma, Dammam, Ghar-Euphrates, and Dibdibba formations alongside Quaternary sediments, and the main recharge of groundwater occurs through direct rainfall and seepage from intermittent valleys (Ali and Hamamin, 2012).

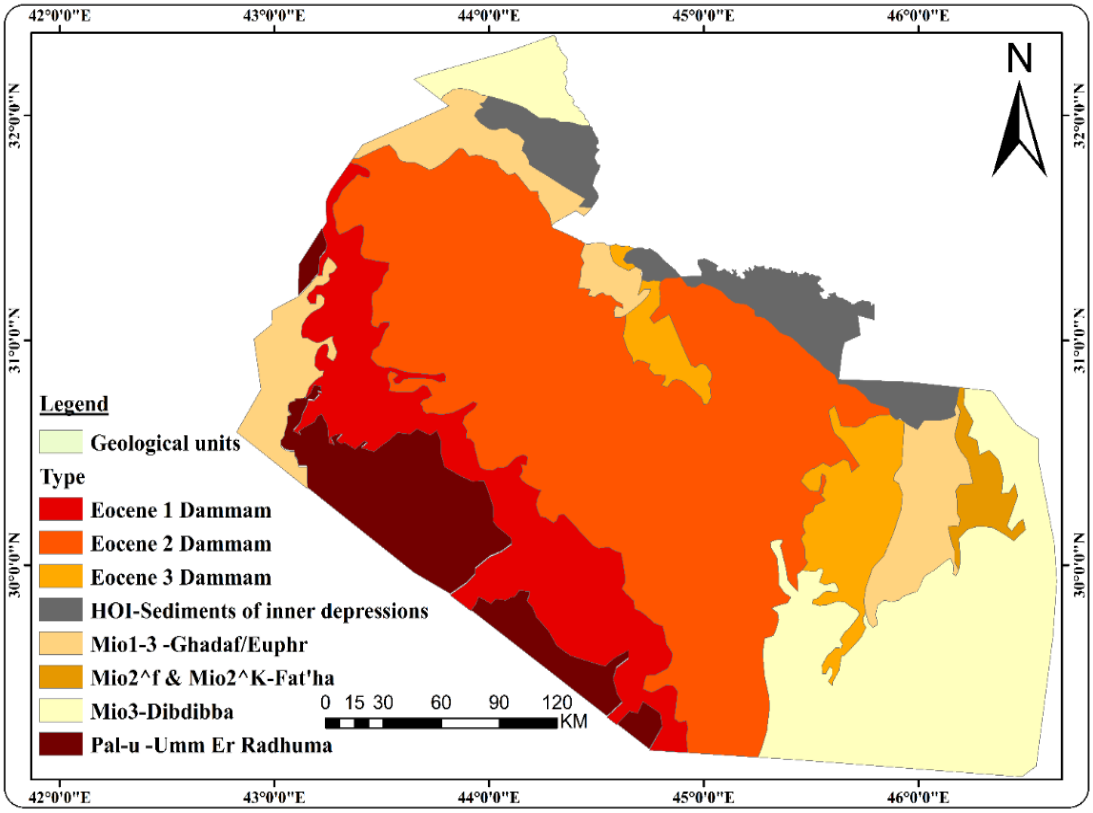


Figure (3.9): Geological map of the studied area modified from Geological Map of Geological Survey and Mining,1986.

Al-Jiburi and Al-Basrawi (2009) state that the Dammam carbonate is a significant regional aquifer in the Southern desert. It is valued for its extensive coverage and large reserves of high-quality water. As a confined aquifer, it is safeguarded against contamination. Dammam aquifer is divided into two aquifer types: unconfined and confined. Its karst porosity is associated with cracks, increasing the aquifer storage properties. The highest permeability occurs in depressions where strong water circulation occurs (Jassim and Goff, 2006).

3.7.2 Thickness of the Dammam Aquifer of the Study Area

The simulation of aquifer systems relies on the spatial distribution of permeability and thickness that can be categorized into simple or complex aquifer systems. If the above two parameters do not diverge spatially, the aquifer is considered simple; Otherwise, it is complex, so the response of aquifer systems depends on the variation in these two limitations. Due to these factors, the efficient models depend on the accuracy of the aquifer parameters (Al-Siba′ai, 2005). Consequently, The variation in the thickness and hydraulic conductivity was considered, and the spatial distribution maps were generated using 3D analysis tools powered by ArcMap, the hydraulic conductivity spatial distribution map was previously presented in Figure (3.8). In contrast, the thickness distribution is presented in Figures (3.10 and 3.11) for the first (unconfined) and second (confined) layers of the Dammam aquifer.

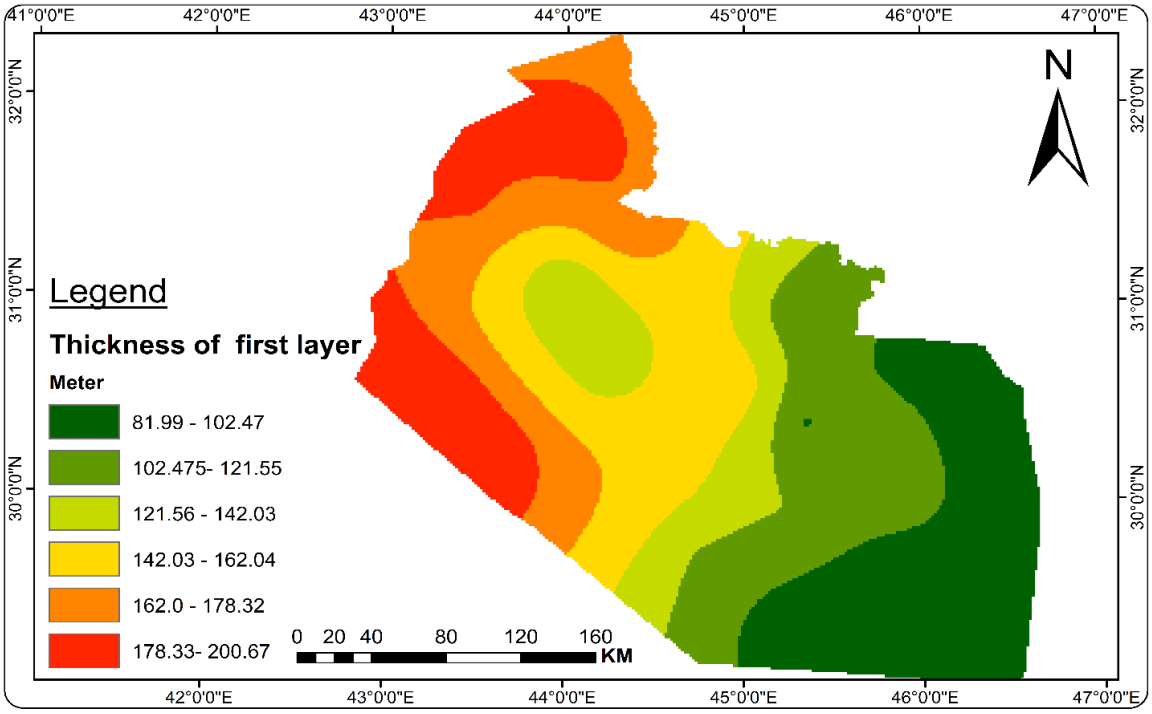


Figure (3.10): Thickness of the unconfined layer of Dammam aquifer.

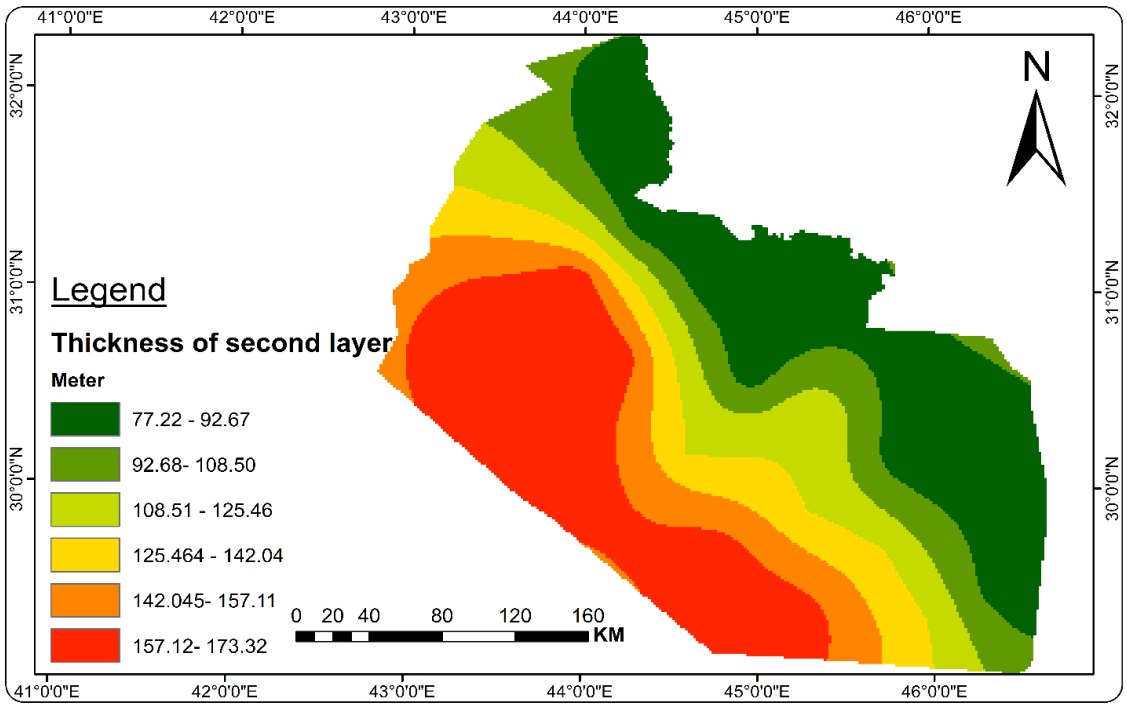


Figure (3.11): Thickness of the confined layer of Dammam aquifer.

3.8 Conceptualization of the Groundwater Model

In GMS, a MODFLOW model can be constructed in one of two ways: by manually assigning and editing values to the cells of a grid (the grid approach) or by building a high-level representation (the conceptual model approach) of the model using feature objects and allowing GMS automatically assign the values to the cells based on the feature objects' properties. Points, arcs, and polygons are all examples of feature objects. Layers or coverages are sets of feature objects. A GMS project can comprise multiple coverages as the groundwater modeling problem can be conceptually represented by a sequence of coverages to ascertain the position of sources and sinks, as well as other layer characteristics such as hydraulic conductivity, model boundaries, and other data that are critical for the simulation. Each coverage has a unique function. Single coverage is a commonly employed technique for various applications such as wells, defined head borders, recharge, hydraulic conductivity, etc. However, it is not required to use coverages this way, and the characteristics associated with coverage are user-defined. Thus, coverages may be built with any combination of attributes (GMS user manual, 2018).

The conceptual model approach is widely regarded as the most effective methodology for representing field circumstances by systematically acquiring data to define groundwater flow dynamics within the modeled region. The user must develop several conceptual models in many cases before a suitable calibration is achieved. This is particularly true in groundwater modeling, where appropriate boundary conditions and stratigraphic representations are often difficult to determine. A grid that fits the conceptual model can be created automatically after the conceptual model item is specified. The model parameters and boundary conditions are calculated and allocated to the proper cells. The conceptual model's MODFLOW data is translated into the grid's cells using the

feature objects and by selecting the Map to 3D grid command to automatically generate a grid. The package dialogs and interactive editing tools available in the MODFLOW menu facilitate the editing and evaluation of data subsequent to the creation and conversion of the conceptual model into a grid. The following sections will detail the basic steps involved in setting up the MODFLOW conceptual model and converting it to a numerical model.

3.8.1 Modeling Input Parameters

In this step, model interfaces such as (FEMWATER, MODAEM, MODFLOW, MODPATH, etc.), the display projection, and unit type for length, mass, time, force, and concentration should be selected. Since the difference in the input parameter projections leads to an error with the operating model, the model will not terminate effectively (Al-Areedhi and Khayyun, 2019). This study employs MODFLOW-2000 within the GMS software, that can conduct steady-state and transient analyses, offering diverse boundary conditions and input options. The projection for the modeled area was the Universal Transverse Mercator (UTM) Projection; Zone 38 (42°E - 48°E – Northern Hemisphere); Meters Planar Units, and the unit type is (meter) for length and (day) for time. Using the Arc-Map 10.7.1 program, the base map for the modeled study region was created. It was then imported into the GMS program and prepared to establish boundary conditions.

3.8.2 Design of Grid

A finite-difference cell-centered 3D grid was developed using the MODFLOW package integrated with the GMS program. The locations of cells are labeled in terms of rows (**I**), columns (**J**), and numbers of layers (**K**). The study area was discretized into rectangular cells in x-y directions with variable cell thickness in the z-direction. The modeled domain comprised 200 rows and 200

columns, resulting in a total of 80000 cells. Among these cells, 41530 were active, while the remaining 38470 were inactive. Each cell had a dimension of (1953.5*1879) m. The model consisted of two layers (**K**=2); the first layer is unconfined, while the second is confined, as shown in Figure (3.12). The elevation of the first layer bottom is considered the top elevation for the second layer.

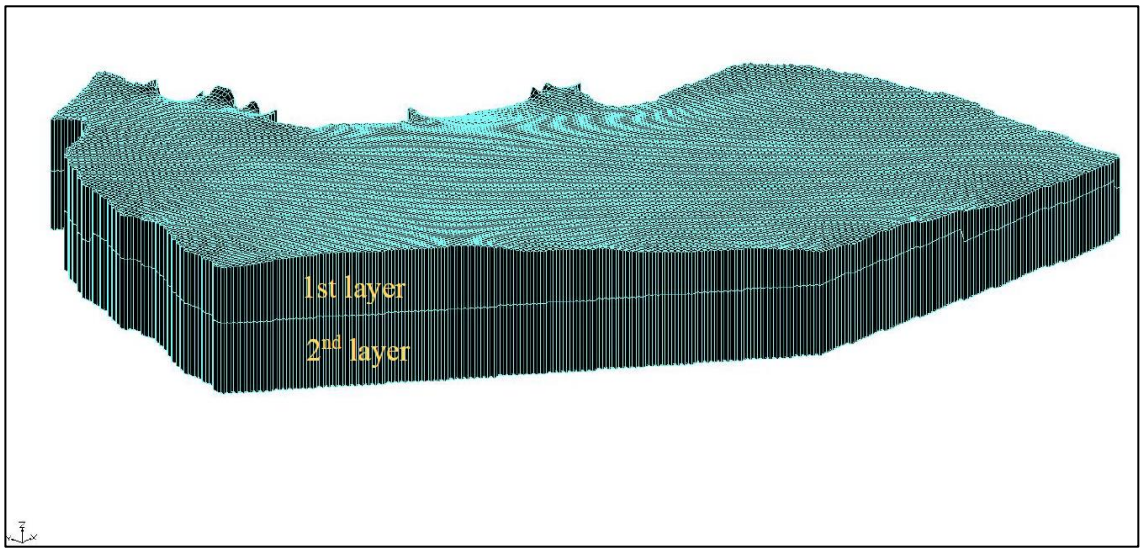


Figure (3.12): Three-Dimensional grid of the modeled study area.

3.8.3 Boundary Conditions

The boundary conditions may be classified as either physical or hydraulic. Anderson and Woessner (1992) have identified impermeable rocks, lakes, and waterways as common forms of physical boundaries, while hydraulic boundaries are often characterized by water divides and flow lines. The term "boundary conditions" refers to the hydraulic conditions that exist along the perimeter of the problem domain. These conditions can be categorized mathematically into three distinct types, as outlined by Anderson et al. (2015).

i- Specified Head Boundary: This boundary condition is also known as the Dirichlet condition. The head along this boundary is set to a known value. Depending on the space, the heads of this type can vary. The constant head

boundary is a special case of this boundary, where the heads along the boundary have a constant value.

ii- Specified flow boundary: This type of boundary condition is also known as a Neumann condition, which gives a specified derivative of the head along with the boundary. For example, in Figure (3.3), the flux boundary condition imposed at the ΔxΔz face can be written as:

$$\frac{\partial h}{\partial y} = -\frac{q_y}{K_y} \tag{3.12}$$

Where; (∂h/∂y) is the y-component of the head gradient and q_y: is the y-component of the specific discharge vector. The no-flow boundary is considered a special case of this boundary, where the flow across the boundary is set to zero.

iii- Head-dependent boundary: Also known as Cauchy conditions, in this type of boundary, the flow across the boundary is calculated according to Darcy's law using a gradient given as the difference between a specified head outside the boundary and that computed by the model at the node that located on the boundary or close to it, which can be expressed as:

$$q_y = -K_y \frac{h_b - h_{i,j,k}}{(\Delta y/2)} \tag{3.13}$$

Where; $h_{i,j,k}$ is the computed head at the cell center, h_b is the specified head along the face of (ΔxΔz) located at a distance of (Δy/2) from the cell center.

The selection of the constant head boundary for the study area was based on the regional groundwater flow pattern of the Dammam aquifer. This boundary was positioned at a considerable distance from the well-field to limit any potential impact on the simulated heads within the model domain, as noted by Al-Basrawi (1996). The boundary was defined in sources/sinks boundary conditions as a

specified head (CHD). Additionally, the Euphrates River, which flows through the eastern side of Najaf and Muthanna governorates, was modeled as a River in sources/sinks boundary conditions. The formation of the river was achieved through the utilization of an arc that possessed defined endpoints, also known as nodes. Each of these nodes required a specific river head stage and bottom elevation to be established. The head stage measurements for the Euphrates River were recorded in the Najaf and Muthanna governorates. In Najaf, the head stage was found to be 21 m for both the start and end nodes of the river arc, with corresponding bottom elevations of 15 m and 14.5 m. In Muthanna, the head stage measurements were 14 m and 13.5 m for the start and end nodes, respectively, with relevant bottom elevations of 8 m and 7 m. Figure (3.13) illustrates the boundary conditions associated with the first layer of the Dammam unconfined aquifer in the assigned study region. On the other hand, the second confined layer was exposed to a boundary condition of constant head.

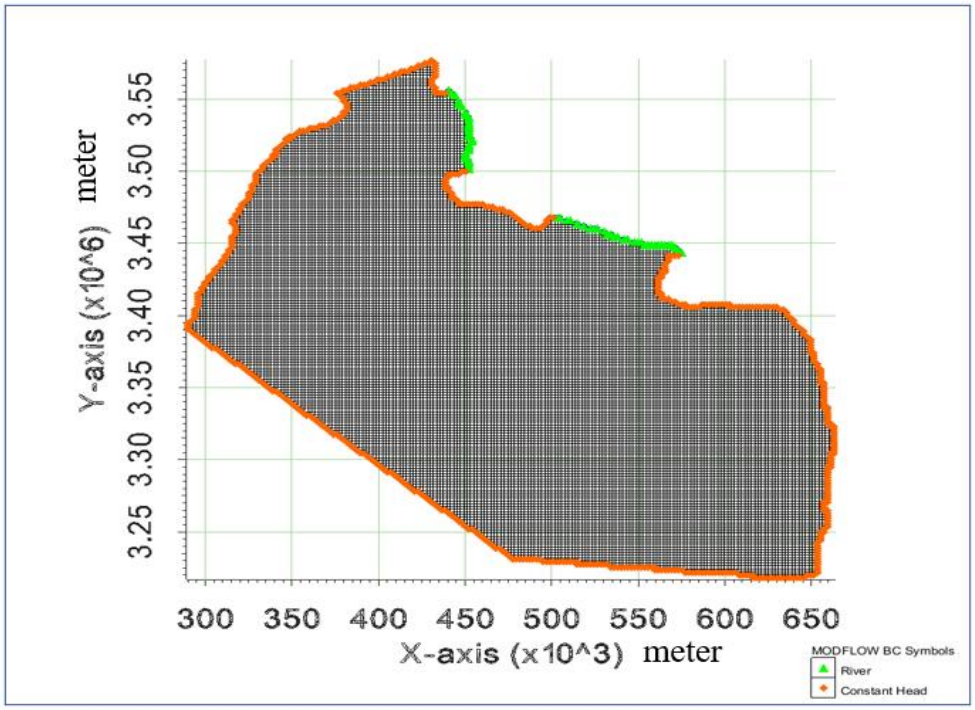


Figure (3.13): Boundary conditions of the modeled study area.

3.8.4 The Initial Hydraulic Parameters

The process of parameter estimation in steady-state necessitates the establishment of initial values for hydraulic parameters, including recharge and hydraulic conductivity. The initial parameter values were obtained from pumping tests carried out by the General Committee for Groundwater (GCGW) in Baghdad, Iraq, and subsequently extrapolated across the groundwater system through Kriging interpolation. As presented in Figure (3.14), the study region was segmented into five distinct zones to facilitate the initial hydraulic conductivity data input. Additionally, one zone was designated for inputting the initial recharge rate.

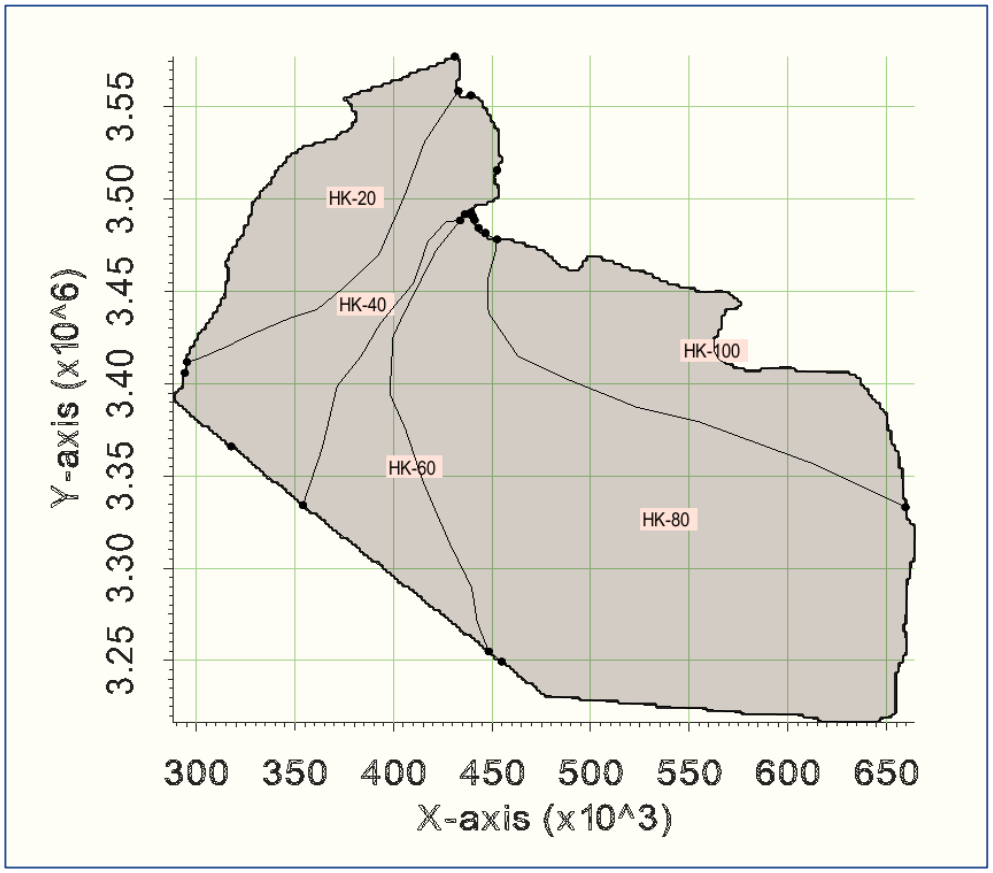


Figure (3.14): Hydraulic conductivity zones for the study area.

An arid climate represents the climate of the region, according to Iraqi Meteorological Organization and Seismology (IMOS, 2021) for Najaf and Muthanna governorates monitoring stations from (1974 – 2021). The climate averages parameters in the Najaf climate monitoring station show the annual average precipitation is (98.8 mm), the average annual temperature is (25C^{o}), the total yearly evaporation is (3448.9 mm), the relative humidity is (42 %), the average annual sunshine is (7.6 hr/d.), and the average annual wind speed is (1.9 m/s). In contrast, the Muthanna climate monitoring station shows the annual average precipitation is (103.3) mm, the average annual temperature is (25C^{o}), the total yearly evaporation is (3476.0 mm), the relative humidity is (40 %), the average annual sunshine is (8.7 hr/d), and the average annual wind speed is (3.2 m /s).

According to (Al-Jiburi and Al-Basrawi, 2009), the majority of the groundwater in the Southern Desert is recharged by direct rainfall. The recharging zone extends beyond Iraq's borders and into Saudi Arabia's territory. Furthermore, Al-Areedhi and Khayyun (2019) reported that Iraq's arid and semi-arid regions mostly rely on rainfall. Consequently, the initial guess of recharge is calculated according to the data in Table (3.4). The annual average precipitation in the Najaf climate monitoring station is (98.8) mm. In contrast, the Muthanna climate monitoring station shows that the annual average precipitation is (103.3) mm.

Table (3.4): The average monthly rainfall records at Najaf and Muthanna monitoring stations from 1974 to 2021 (IMOS, 2021).

Months	Rainfall (mm)	
	Najaf Station	**Muthanna Station**
Oct.	5.01	4.25
Nov.	16.2	17.33
Dec.	15.4	14.50
Jan.	18.03	20.39
Feb.	14.83	15.73
Mar.	13.05	17.29
Apr.	11.9	8.75
May	4.35	4.86
Jun.	0.04	0.05
Jul.	0	0
Aug.	0	0
Sep.	0	0.13
Total	98.8	103.3

The recharge was assumed to be uniform in the study area. In GMS, the recharge rate has been set to be calculated from the following equation (GMS user manual, 2018; Sule, 2017; Al-Areedhi and Khayyun, 2019 and Yehia, 2020):

$$\text{Recharge} = \frac{\text{annual rainfall (m) x } 0.05}{365} \tag{3.14}$$

The governorate of Najaf had 98.8 mm of precipitation on average annually. As a result, it was determined that the recharge rate was 1.35×10^{-5} m/day, whereas the annual average precipitation for the Muthanna governorate was 103.3 mm. As

a result, 1.4×10^{-5} m/day was determined to be the recharging rate. The examined governorates' determined recharge values differed very little from one another. Therefore, the research region was treated as a single zone, and the initial recharge of 0.000014 m/day was input. The model will be run using these initial values during the steady state, and the solution will be examined.

3.9 Model Calibration

The calibration of the model is crucial in groundwater modeling. If a groundwater modeling system will be utilized for any predictive features, it has to prove that it can accurately simulate aquifer behavior. The calibration process can be defined as using the groundwater flow model to calculate the aquifer hydraulic parameters. This method continuously modifies one or more aquifer parameters, such as recharge and hydraulic conductivity. The numerical model is iteratively executed until the obtained solution aligns with the observed values in the field with a satisfactory degree of accuracy. (Anderson and Woessner, 1992). It is crucial to carry out the steady-state calibration for the groundwater model of the study area because the steady-state simulation output is used as the initial condition to execute the unsteady-state condition. Hence, in addition to (25) well-distributed observation head points, two adjustable aquifer parameters, hydraulic conductivity and recharge rate, were adopted.

The steady-state calibration was achieved by reducing the divergence between the observed and simulated groundwater head piezometric heads. In steady-state calibration, initial aquifer parameters, such as recharge and hydraulic conductivity, are systematically changed until the calculated head values match the observed or the differences between them are reasonably small and within an acceptable interval of accuracy.

The parameters estimation procedure can be carried out using either an inverse modeling technique or a trial-and-error method. The trial-and-error strategy can be applied iteratively to adjust model parameters until the model computed values match the field observed values to an acceptable degree of agreement. This method was used when slight adjustment was required. However, more accurate calibration can be accomplished if inverse modeling is used. The process of estimating parameters may be made easier with the use of an inverse model. To reduce the difference between the calculated and observed groundwater head values, the inverse model is regularly updated depending on the user-defined value, which includes the initial value with high confidence and a range of minimum and maximum values.

GMS includes an interface to an inverse model called PEST, developed by Doherty (1998) and used to calibrate the model. During calibration, PEST runs the MODFLOW model thousands of times and compares the results predicted by the model to observations. Each observation point is accompanied by a calibration target representing the residual error's magnitude. The residual is determined by subtracting the observed value from the simulated value. The confidence interval or standard deviation is used to identify a desired target size. A user can choose to show a number of statistical plots in addition to the calibration target next to the observation points. After the parameter estimation run, the adjustable parameters are finally calculated. New values for hydraulic conductivity and recharge will be entered, and a new MODFLOW solution will be computed and plotted.

3.10 Evaluation of Steady-State Calibration

The first step in determining the groundwater system's behavior under typical conditions is a steady-state flow simulation to understand the behavior of the groundwater system initially. The obtained results were then used as initial inputs of an unsteady state simulation, which acts as the baseline for the long-term aquifer behavior with various pumping operations. The model includes observed data for heads at several scattered observation wells. The calibration results were derived from the optimal HK/RH values that minimize the calibration error. Each zone value was determined based on the minimum, maximum, and initial values, as depicted in Figure (3.15).

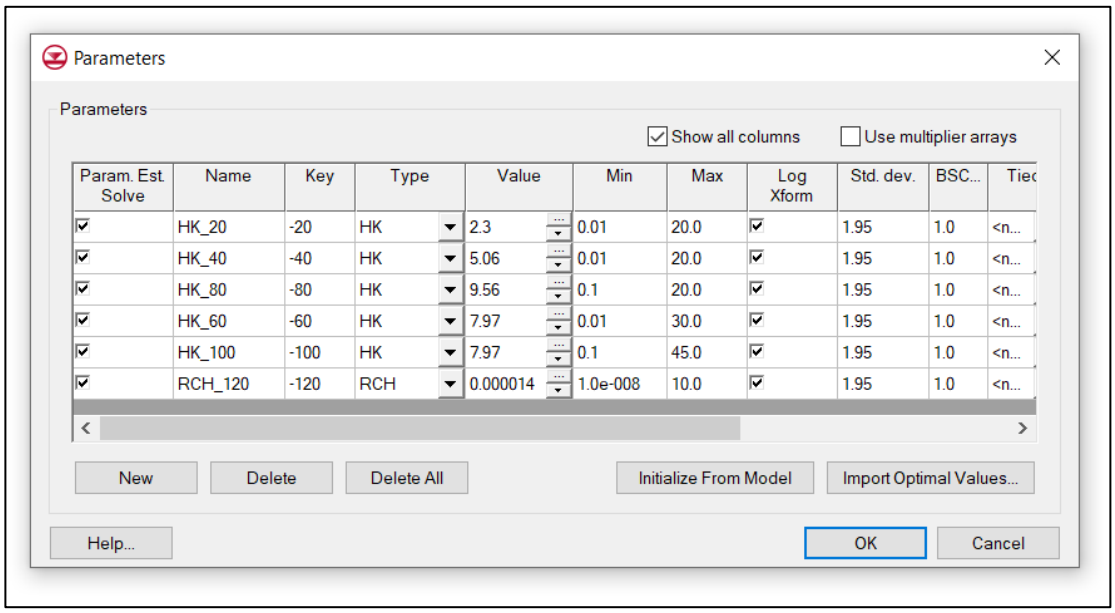


Figure (3.15): Parameters estimation via PEST tools for the study area.

The field observed values and the estimated values generated by the model would be automatically compared throughout the calibration process. The residual values were computed for a (95%) confidence and head interval of (1.5 m). The calibration targets illustrated the validity of the calibration process. If an observed value has been assigned to a feature object, the calibration error at each object can

be plotted using a "calibration target." A set of calibration targets provides useful feedback on the calibration error's magnitude. Green, yellow, and red are the three degrees of color that may be found on the calibration target; these colors are used as markers for the accuracy of the calibration results (Christensen and Cooley, 2003). The components of a calibration target are illustrated in Figure (3.16).

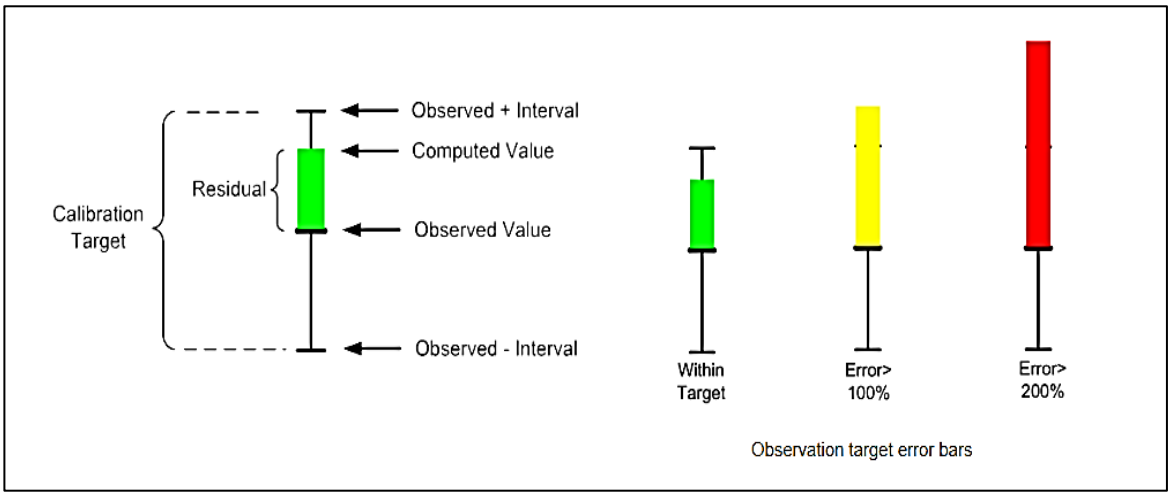


Figure (3.16): Calibration target.

The observed value is represented by the center of the target. The upper limit of the target is equivalent to the observed value added to the interval, while the lower limit corresponds to the observed value reduced by the interval. The colored bar denotes the presence of an error. When the bar is completely contained within the target, it is represented by the color green in the visualization. If the error in the observation target exceeds the permissible limits by a small margin, it will be indicated by a yellow highlight on the error bar. On the other hand, when the error exceeds 200%, it is denoted by the color red, as stated by Aquaveo (2017). The findings of the calibrated model during steady-state conditions can be stated as follows:

1. The calculated and observed groundwater piezometric head measurements exhibited a high degree of agreement in the modeled region. The contour map showing the simulated heads of the aquifer has been presented in Figure (3.17).

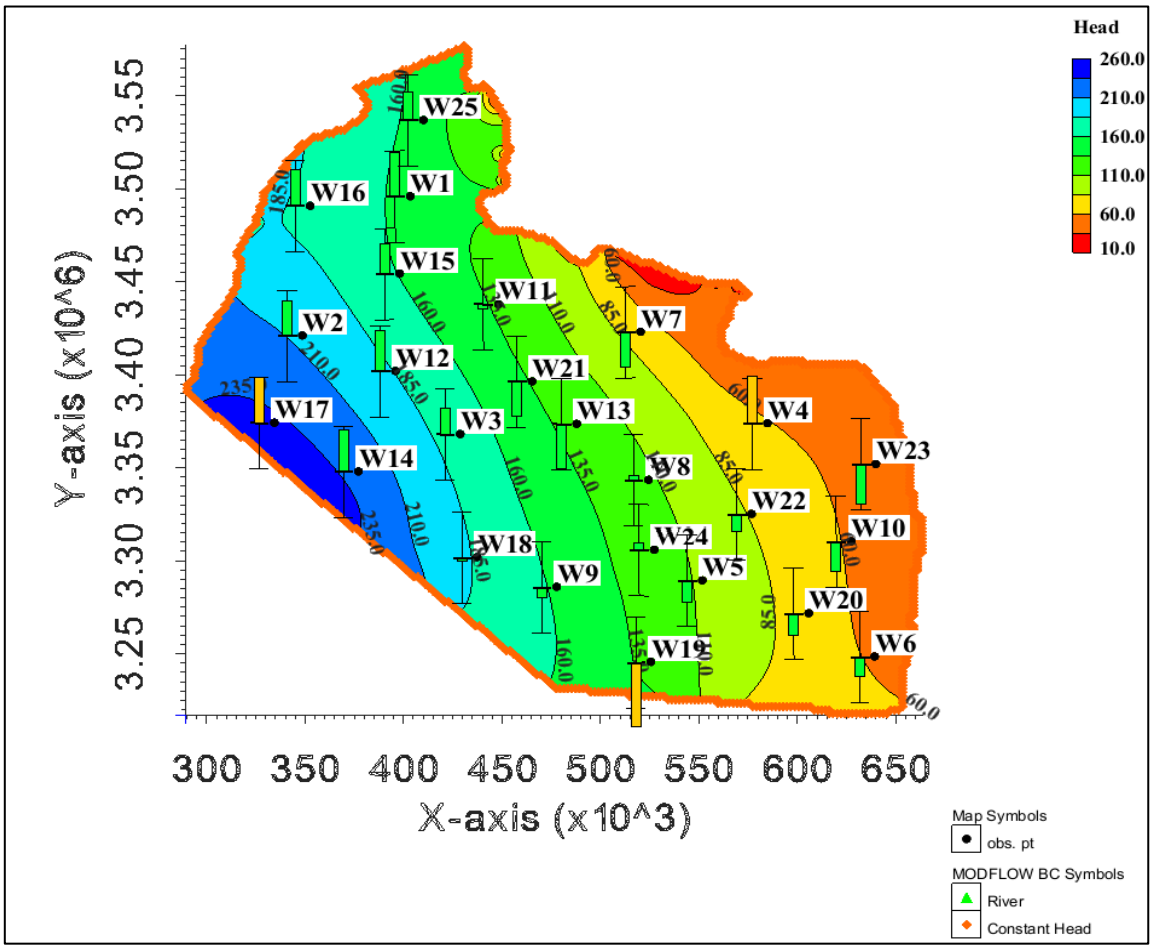


Figure (3.17): steady-state groundwater heads distribution in the study area.

Out of the 25 observation targets in the steady state simulation, 21 of them have been denoted with green error bars. Additionally, one of the observation target wells has been found to be in perfect agreement with the observed values. Merely three error bars were colored in yellow; however, they were found to be within the permissible limits of the target, thereby signifying a substantial enhancement over the initial solution.

2. The study conducted an error analysis, as reported by Anderson and Woessner (1992), in which the Mean Error (ME), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were employed as metrics to assess the model's reliability. These metrics can be mathematically represented as:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(h_m - h_s)_i \tag{3.15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(h_m - h_s)_i| \tag{3.16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(h_m - h_s)^2{}_i} \tag{3.17}$$

Where h_m is the measured head and h_s represent the simulated head.

The ME, MAE and RMSE for the steady-state calibration were found to be very low, specifically (-0.0655, 0.9665 and 1.0921) meters, respectively. According to Anderson et al. (2015), the calibration of the model is deemed successful if the RMSE is lower than certain values of the calibration target range. However, there are no guidelines regarding the acceptable magnitude of the ME and MAE other than it is desirable to minimize these values. Although the utility of standard criteria is recognized, the modeling community has not adopted uniform calibration standards. This partially reflects the awareness that all modeling requires subjective judgment (Fienen, 2010). The findings of the study suggest that the conceptual model, boundary conditions, and final hydrological parameters are reliable and can be further used. This is supported by the fact that the RMSE is below the calibration target value of 1.5 m, while both ME and MAE exhibit minimal values.

3. The calibration fit was demonstrated using a scatter plot comparing computed head values with observed head values. This presentation can be observed in Figure (3.18) and Table (3.5). The value of the coefficient of determination (R^2), commonly known as the goodness of fit, was determined to be 0.9997.

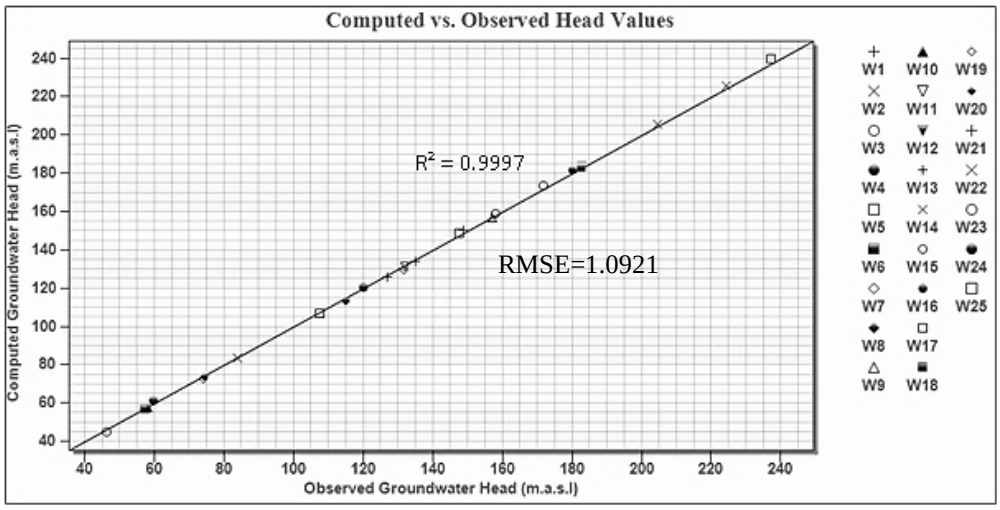


Figure (3.18): Scatter plot of computed vs. observed head

Table (3.5): Observation wells data of the calibration process.

ID	Longitude (E)	Latitude (N)	Obs. Head (m)	Comp. Head (m)	Residual Head (m)
W1	43° 58' 56.2754"	31° 35' 41.1816"	148.750	150.218	-1.468
W2	43° 24' 57.7358"	30° 54' 48.9036"	204.500	205.665	-1.165
W3	44° 15' 30.1231"	30° 26' 34.1984"	172.900	173.758	-0.858
W4	45° 53' 04.2182"	30° 29' 40.1829"	59.900	61.467	-1.567
W5	45° 32' 06.8596"	29° 44' 02.6956"	107.600	106.891	0.708
W6	46° 26' 06.2870"	29° 21' 28.6247"	57.340	56.691	0.648
W7	45° 12' 48.8061"	30° 56' 28.1578"	74.020	72.869	1.150
W8	45° 15' 12.7971"	30° 13' 22.0477"	112.980	113.147	-0.167
W9	44° 46' 13.8069"	29° 42' 14.1101"	157.230	156.889	0.340
W10	46° 19' 11.2395"	29° 55' 07.0393"	58.200	57.236	0.963
W11	44° 27' 25.1437"	31° 04' 25.4341"	131.400	131.279	0.120
W12	43° 54' 44.8533"	30° 44' 46.9633"	182.750	184.091	-1.341

ID	Longitude (E)	Latitude (N)	Obs. Head (m)	Comp. Head (m)	Residual Head (m)
W13	44° 52' 29.8399"	30° 29' 40.5681"	127.110	125.662	1.448
W14	43° 43' 20.9153"	30° 15' 26.6924"	224.330	225.729	-1.399
W15	43° 55' 48.6398"	31° 13' 09.4988"	157.960	158.969	-1.009
W16	43° 26' 49.9841"	31° 32' 33.9049"	180.500	181.705	-1.205
W17	43° 16' 28.6251"	30° 29' 19.1704"	238.030	239.552	-1.522
W18	44° 21' 07.0150"	29° 50' 45.6878"	182.750	182.655	0.094
W19	45° 15' 49.3164"	29° 20' 21.2537"	131.880	129.776	2.103
W20	46° 05' 32.1760"	29° 34' 17.2911"	74.440	73.738	0.701
W21	44° 38' 15.5435"	30° 41' 59.6365"	134.970	133.839	1.130
W22	45° 47' 48.1820"	30° 03' 19.7133"	83.850	83.288	0.561
W23	46° 27' 17.0022"	30° 17' 29.9661"	46.440	45.148	1.291
W24	45° 16' 57.8083"	29° 53' 00.2789"	119.960	120.216	-0.256
W25	44° 02' 55.0063"	31° 57' 52.3314"	147.580	148.517	-0.937

Table (3.6) displays the calibrated hydraulic conductivity and recharge rate values for each zone of the study area.

Table (3.6): Parameter estimation outputs.

Parameter keys	Initial value m/day	Estimated value m/day
HK_20	2.30	1.46569
HK_40	5.06	2.67806
HK_60	7.97	4.12694
HK_80	9.56	20.0
HK_100	7.97	13.3331
RH_120	1.4×10^{-5}	1.66×10^{-6}
* HK is the horizontal hydraulic conductivity of the aquifer. * RH is the recharge rate for the aquifer.		

The PEST is capable of determining the optimal parameter values and calculating the sensitivities of each parameter during each iteration. The Sensitivity Process is used to calculate the sensitivity of model results to variations in parameter values by computing the partial derivative of the objective function with respect to each parameter of interest. The data presented in Figure (3.19) can be employed to differentiate between the parameters that significantly affect the model's results and those that have minimal impact. The parameter identified as code (HK_100) significantly impacts the model's outcomes relative to other employed parameters, indicating that the model exhibits sensitivity to variations in hydraulic conductivity within the HK_100 zone.

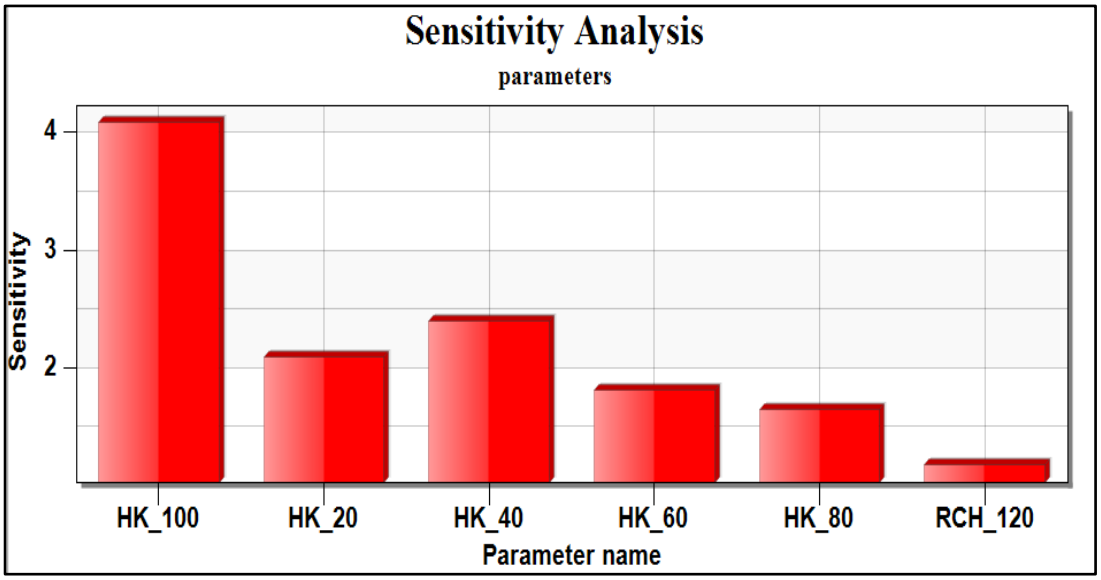


Figure (3.19): Parameters sensitivity analysis.

3.11 Steady-State Validation

The validation technique necessitates the use of the remaining recorded data to verify the model's response through a comparison of the model outputs with the field observations. The validity of the MODELOW- Model was assessed using a total of seven observation wells, as shown in Table (3.7) and Figure (3.20).

Table (3.7): Observations wells data of the validation process.

ID	Longitude (E)	Latitude (N)	Obs. Head (m)	Comp. Head (m)	Residual Head (m)
W1	44° 05' 34.8613"	31° 45' 10.2818"	142.6	142.6349	-0.0349
W2	43° 48' 00.7140"	30° 42' 04.2410"	195.1	194.0856	1.0144
W3	44° 34' 07.6289"	31° 09' 16.0495"	117.8	118.3564	-0.5564
W4	45° 23' 52.5506"	30° 58' 20.2802"	53.6	52.43447	1.1655
W5	44° 23' 38.7233"	29° 43' 27.8103"	180.6	180.2708	0.3292
W6	45° 07' 43.2404"	30° 05' 08.8888"	124	125.1247	-1.1247
W7	45° 02' 54.5451"	29° 29' 21.5182"	141	141.2021	-0.2021

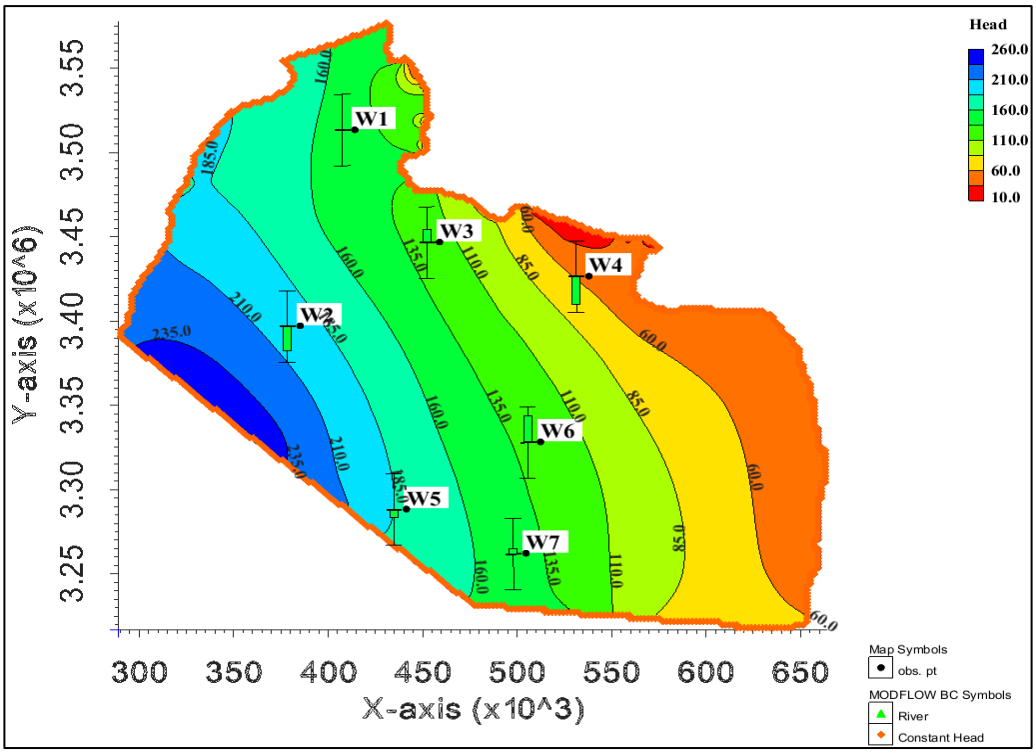


Figure (3.20): Validation target of steady-state.

The scatter plot between the computed and observed head for the validation wells; Figure (3.21) shows that R^2 equal to 0.9997.

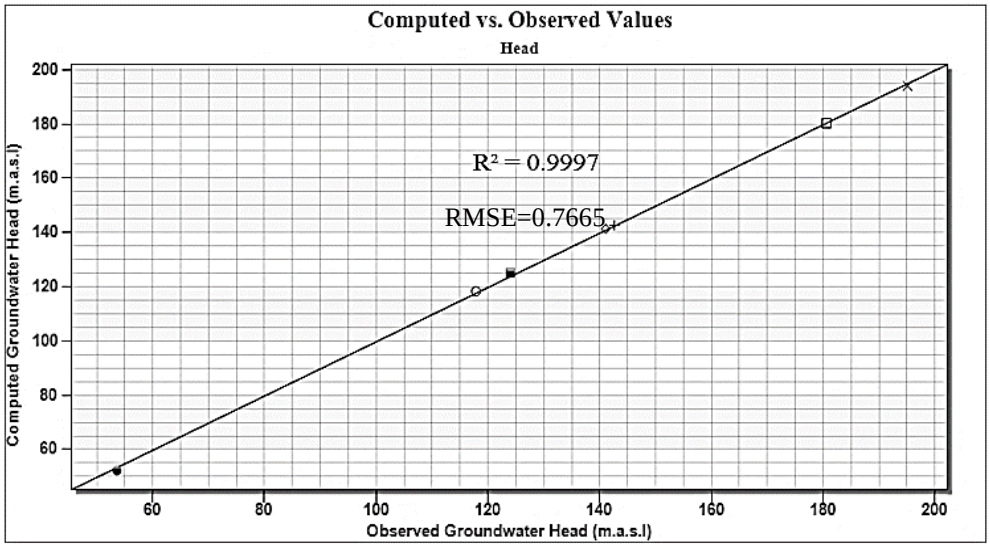


Figure (3.21): Scatter plot of computed vs. observed head of steady-state validation.

The validation of the groundwater flow model was accomplished based on the RMSE in addition to the MAE and ME. The ME, MAE, and RMSE for the steady-state validation were found to be (0.0844, 0.6325 and 0.7665) meters, which validates the reliability of the simulation.

3.12 Transient Model Calibration (with Pumping Conditions)

In the case of a transient flow simulation, the groundwater heads varies from one-time step to another. The hydraulic heads resulting from the steady-state simulation are adopted as the starting heads for the transient analysis. In this condition, it is important to assign new parameters to the model, such as storage coefficient and pumping wells. Effective transient calibration relies chiefly on the good assessment of driven hydraulic conductivities and recharges from the steady state. Generally, it is required to specify storage terms for the confined aquifer,

which account for the quantity of water stored or released from the aquifer due to variations in hydraulic heads.

The storage coefficient (storativity) is the main parameter that is changed during the process of the transient calibration, where its value is modified based on a trial-and-error basis until a good match is obtained between the observed and calculated groundwater heads from July 2021 to Jun 2022. The storage coefficient was initially chosen to be 6.27×10^{-4} based on GCGW pumping test analysis reports.

3.12.1 Time Stepping in Transient Simulation

"Stress periods" are the computational intervals for a MODFLOW simulation. If needed, stress intervals can be subdivided into smaller time increments. A transient flow simulation is performed for a stress period of 365 days, starting on 1st Jul 2021, while the end time was set to 1st Jun 2022, and the number of stress periods is equal to 12.

The selection of the simulation time step is critical in a transient model design because the space and time discretization value strongly impacts the numerical results (Anderson and Woessner, 1992). The time step must not be too large, causing important data to be missed or too small, which requires a long time for the system to complete the calculations. MODFLOW requires that the time values for transient data be entered relative to a time value at the beginning of the simulation. Also, the times must be compatible with the time unit that was chosen for the model. The strategy used in GMS for managing transient data makes it possible to enter all time values using a simple date/time format. The steady state, or reference time, is the beginning of the first MODFLOW stress period.

3.12.2 Evaluation of Transient State Calibration

The unsteady state is a significant step in expressing the behavior of the aquifer as a consequence of the influence of pumping operations since it gives a long-term visualization of this behavior. The evaluation of the transient model's performance was conducted through the utilization of three observation wells, wherein the monthly recorded head measurements from 1 July 2021 to 1 June 2022 are employed. The calibration targets for the three calibration wells are shown in Figure (3.22).

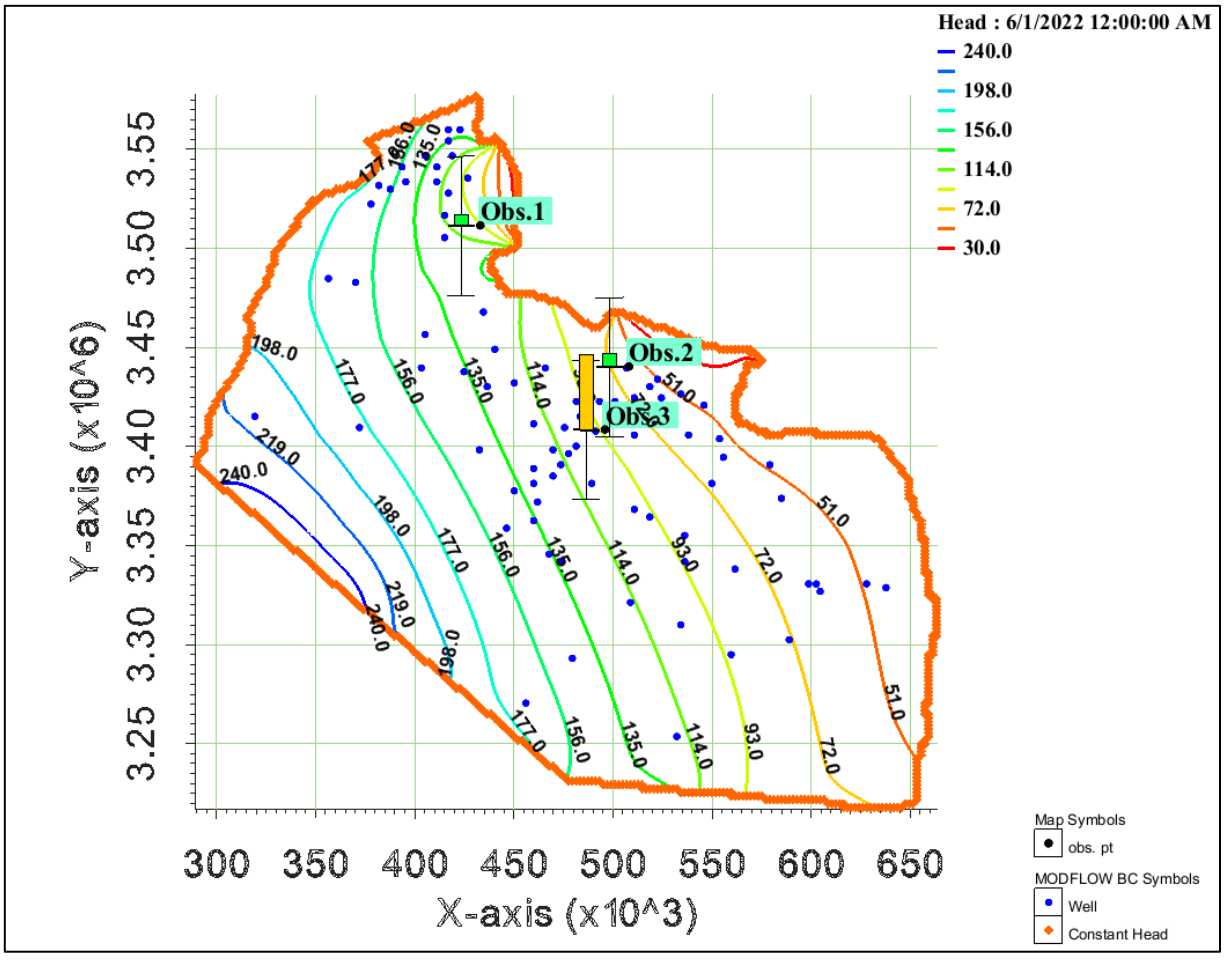


Figure (3.22): Calibration target for the simulated transient groundwater piezometric head of Jun. 2022.

It can be noticed that two of the calibration targets demonstrate a high level of agreement. Furthermore, it is worth noting that the third observation well is equipped with a yellow calibration target, which exhibits a deviation within the permissible error threshold of 1.5 meters at the end of the designated time period. The range of the storage coefficient (storativity) that resulted after the transient state's final calibration was found to vary from 5×10^{-5} to 6.5×10^{-4}. A transient scatters plot was used to evaluate the transient model, as shown in Figure (3.23). The R^2 was calculated to be 0.9884 based on the scatter plot between the computed transient heads and observed values.

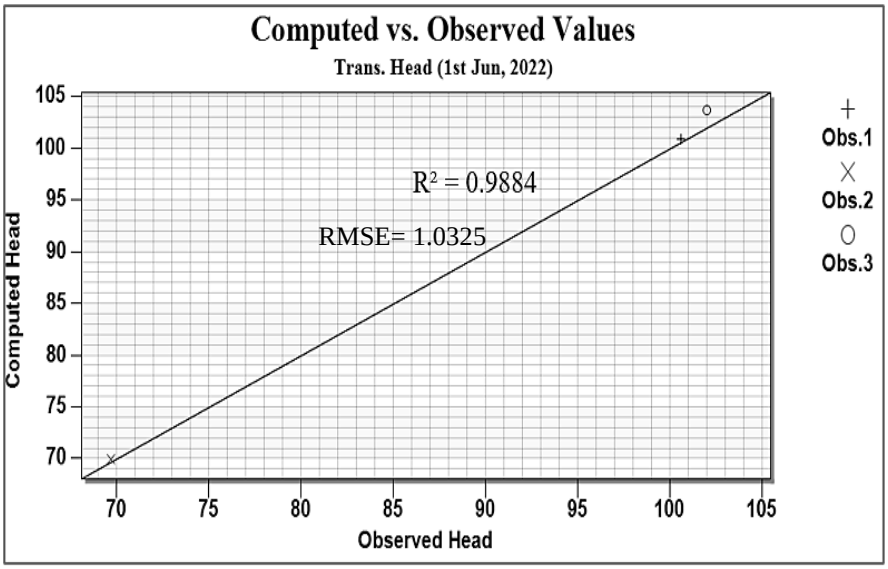


Figure (3.23): Scatter plot of the computed head vs. observed head for transient simulation for Jun 2022.

A time series plot for transient observation wells data was presented in Figure (3.24) to show the relationship between the observed and computed periodic head values. A good performance of the model was observed through the transient simulation.

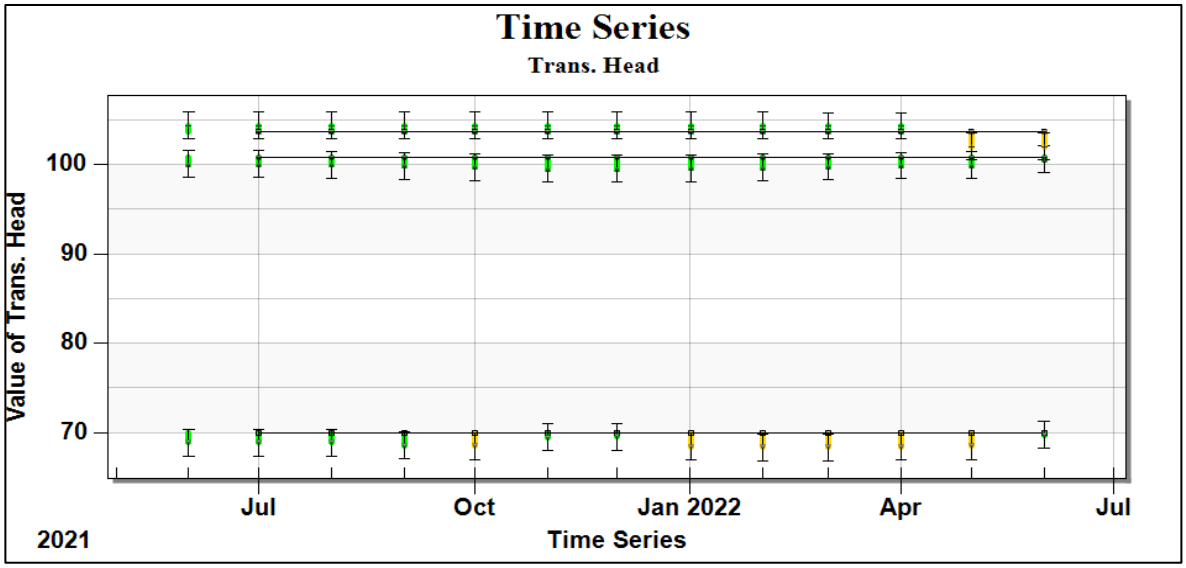


Figure (3.24): Time series plot for transient observation wells.

The transient calibration error for the month of June 2022 yielded groundwater flow model ME, MAE, and RMS errors of (-0.4872, 0.9572, and 1.0325) respectively. The observed errors throughout the simulated periods were found to be minimal, hence confirming the reliability of the simulation.

3.13 Operating Scenarios

Under transient/pumping conditions, two realistic operational scenarios were introduced to provide a better understanding of potential groundwater stress locations within the study region. The first scenario was adopted based on the fact that the study area is facing a decline in surface water inflow and a significant increase in water demand. As a result, 89 wells were pumped at their full capacity from 1st July 2021 to 1st Jun 2022. In contrast, the second operation scenario uses 228 wells in the study area due to the increased water demand. The model was developed to be run for each month. The two operational scenarios will be further discussed in the following sections.

3.13.1 First Scenario

The first scenario includes simultaneously operating 89 pumping wells at their total capacity. Wells have an actual pumping rate ranging from 3 ℓps to 20 ℓps, depending on the conditions. The wells averaged a pumping rate of 6.54ℓps adopted for each of the 89 wells. For a whole year, the model was subjected to these conditions. It was calculated that the total daily discharge from all 89 wells is (582.3 ℓps). Figure (3.25) shows the results of the transient state of operation of the groundwater flow model, which indicated that the drawdown in the groundwater piezometric head ranged from 0.4 m to its maximum value of 5.8 meters during the operation period.

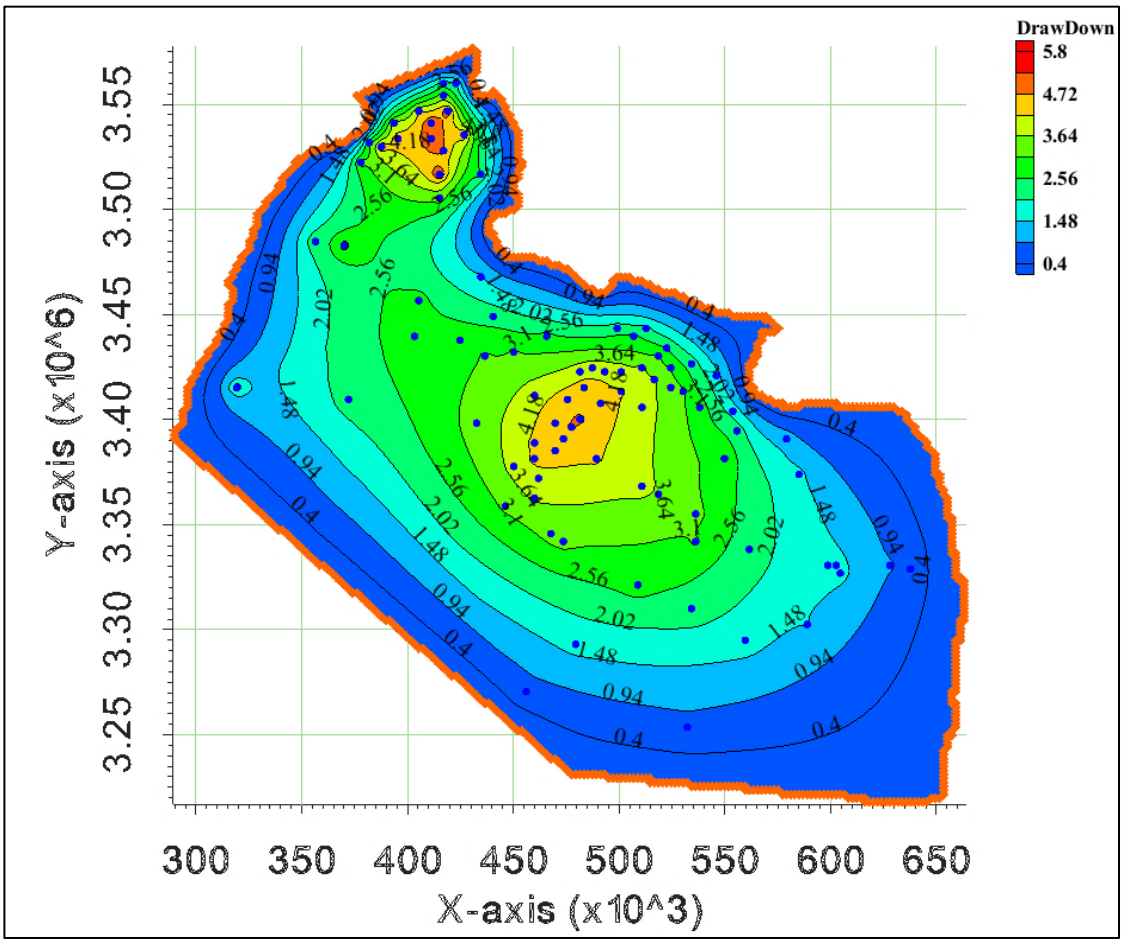
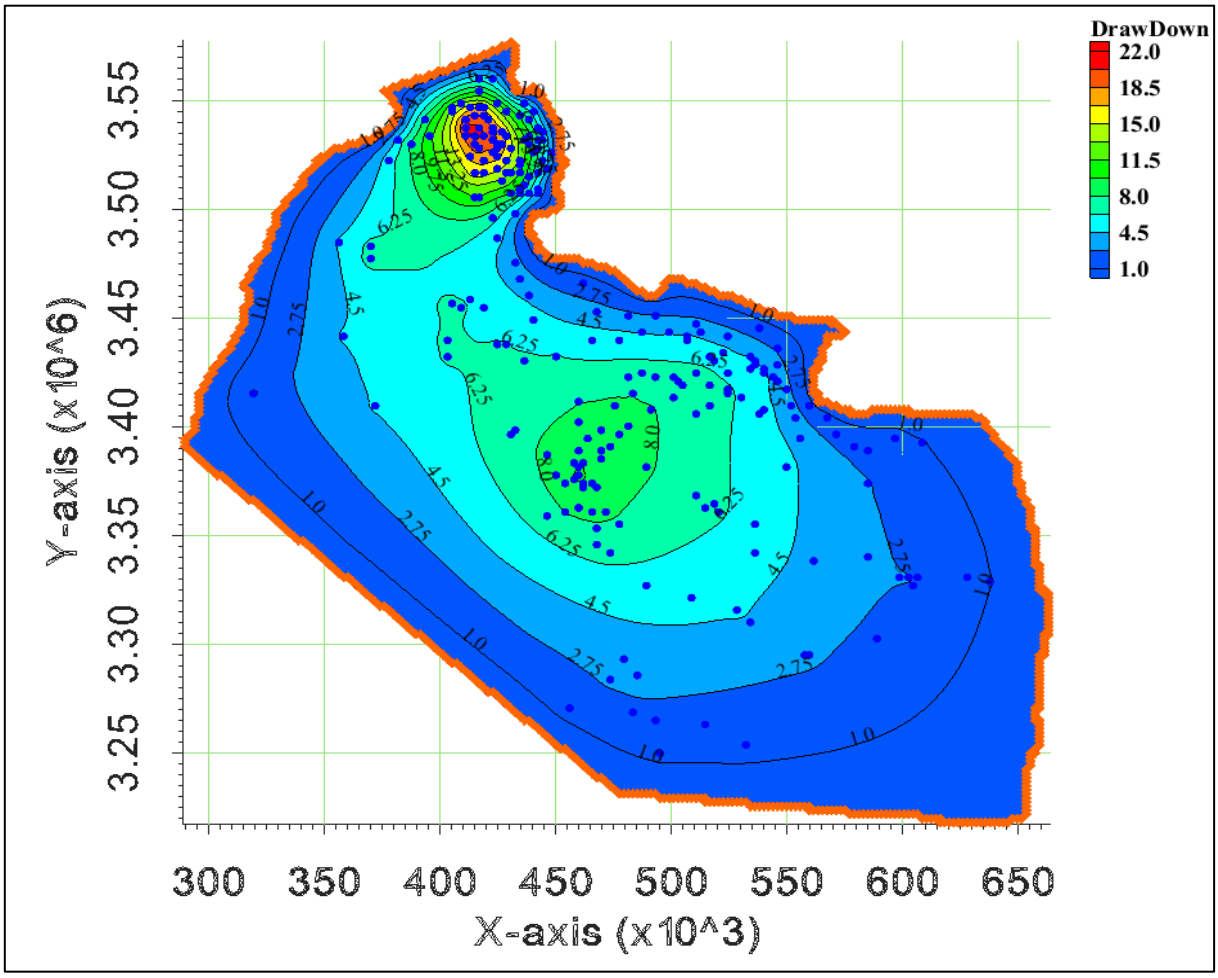


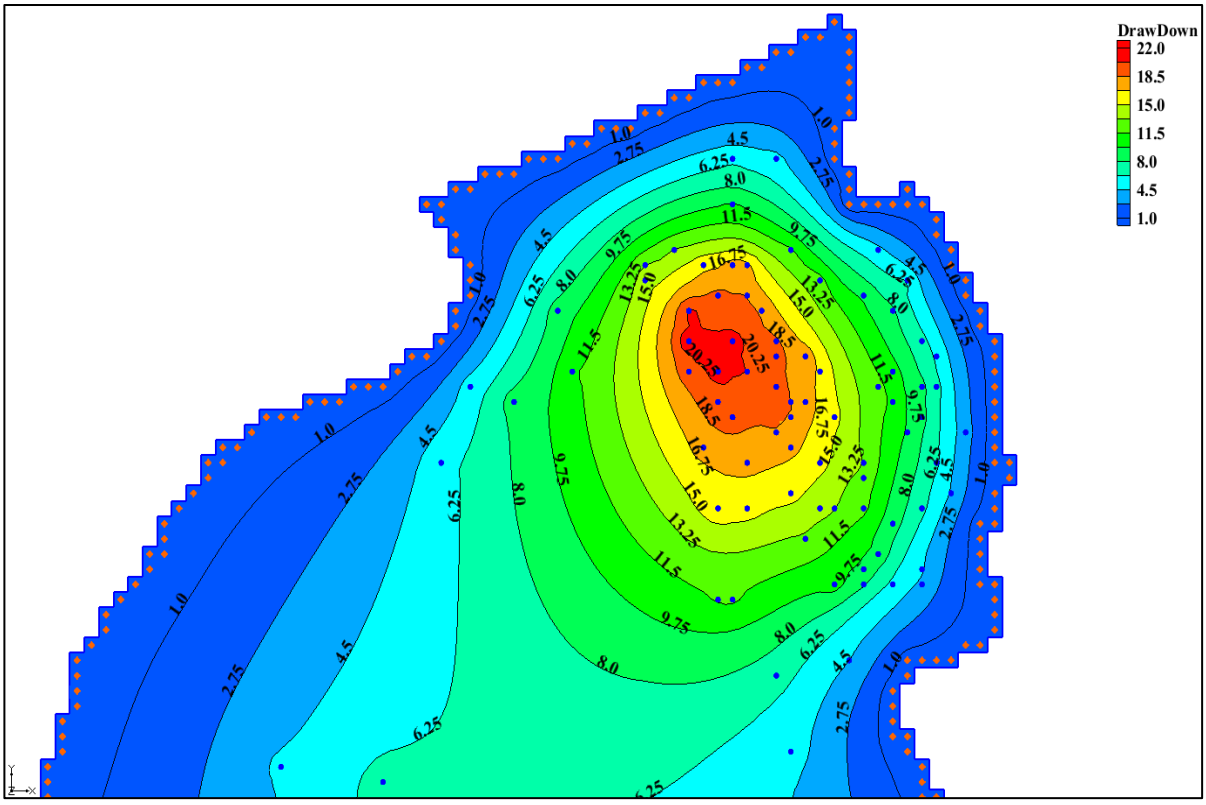
Figure (3.25): Groundwater head drawdowns in the first scenario.

3.13.2 Second Scenario

In the second operational scenario, an additional 139 pumping wells have been added to the existing wells to meet the water demand needs. The wells operated by their actual pumping rate of an average discharge rate equal to 6.73 ℓps. The discharging rates were found to be (1534.1 ℓps) from 228 pumping wells. Drawdown calculations indicate that the drawdowns range from 1m to 22 meters. The simulation results are shown in Figure (3.26: (a) and (b)).



(a) Drawdowns of groundwater piezometric head for Dammam aquifer.



(b) Drawdowns in the piezometric head of the Dammam aquifer northern part.

Figure (3.26 a and b): Groundwater head drawdowns in the second scenario.

The high drawdown values were recorded in populated areas as a direct result of the increased numbers and rate of discharge of water wells in such areas. The pumping rates increased from 582.3 ℓps to 1534.1 ℓps due to the increase in water demand, mainly attributed to the expansion in agricultural activities and the high population growth in the surrounding regions and extended to the study area. In contrast, low drawdown values are due to the distribution of the scattered wells. Consequently, many wells can be drilled in those areas of low drawdowns.

Chapter Four

Identification and Mapping of Groundwater Potential Zones

4.1 Introduction

Groundwater potential mapping (GPM) is a method for strategically developing and managing water resources that categorize different parts of a watershed or region according to the possibility of groundwater availability defined by Díaz-Alcaide and Martinez-Santos (2019) as a way to manage water resources.

Iraq experiences water scarcity, with a notable impact on its surface water sources. As a result, the use of groundwater has become the only feasible solution for meeting the water supply. Identifying appropriate sites and aquifer thickness through drilling tests and stratigraphic analysis is widely regarded as the most dependable and common method. However, this approach is also associated with high costs, time-consuming, and requires specialized labor. Additionally, random drilling without adequate planning often results in failure, and uncontrolled exploitation depletes groundwater potentiality (Prasad et al., 2008; Ratnakumari et al., 2012; Ahmed and Sajjad, 2018). As a result, it is essential to conduct an updated assessment of groundwater resources employing contemporary and effective methodologies to enhance management techniques and protect the sustainability of the aquifer system.

Geospatial and remote sensing methods are valuable tools for evaluating and conserving groundwater resources, identifying surface characteristics, and assessing regional groundwater (Jha et al., 2007; Rahman, 2008; Mondal et al., 2017). GIS-based techniques are used to map groundwater vulnerability and assess risk. Satellite and remote sensing data are combined to identify factors that

influence groundwater occurrence and flow, such as geology, soils, land use, and drainage. Recently, integrated RS, GIS, and multi-criteria decision-making (MCDM) approaches have been used to identify groundwater potential zones, particularly in arid regions.

This chapter aims to construct a groundwater potentiality map for the Dammam confined aquifer using integrated RS, GIS, and MCDM with an analytical hierarchy process (AHP) to facilitate groundwater management. A cost-effective MCDM approach is used to create a multiclass groundwater potential map by extracting relevant thematic layers based on factors that influence the groundwater potentiality, integrating all thematic by developing a normalized Pairwise comparison matrix based on groundwater potential zoning layers using the analytical hierarchy process (AHP) and validating the applied model using measured well yield to determine the accuracy of the results. Such a study will majorly affect the early planning of groundwater exploration since it offers preliminary and reliable information on the ideal zone before starting the drilling process, especially in arid locations with limited groundwater information to target appropriate drilling sites for groundwater wells.

4.2 Groundwater Potential Zones (GWPZ)

The term "groundwater potential" refers to the quantitative assessment of groundwater that can be safely extracted from a given area. The groundwater potentiality is a function of several hydrologic and hydrogeological factors. From a hydrogeological perspective, this term can be defined as the possibility of groundwater occurrence in an area. According to Jha et al. (2010), accurately identifying of groundwater potential zones can provide decision-makers with a useful framework to determine suitable groundwater strategies in a region where water resources are scarce. These zones are typically delineated using geospatial

and remote sensing methods, which provide information on several factors affecting groundwater occurrences and flow, such as geology, geomorphology, soils, land use, land cover, drainage, and lineaments. Proper delineation of these zones can help reduce the risk of over-extraction and depletion to ensure the sustainable use of groundwater.

Generally, identified three key groups of factors that influence groundwater accumulation: water sources, such as rainfall and distance from the channel network; possible infiltration pathways, such as geology, lineament density, distance from fractures, soil type, and land use; and opportunity, which refers to the amount of time available to infiltrate the water supply including elevation, slope steepness, and drainage density.

4.3 Role of Integration between Remote Sensing and Geographic Information Systems

RS data coupled with field observations will be adopted for resource inventory and assessment of the existing situation. Once such information is integrated into a GIS environment, it can be applied to quantify and analyze the decision determinants associated with the decision problem. The application of remote sensing data enables a comprehensive understanding of the functioning, patterns, and spatial changes of the earth system at the global, regional, and local levels. In addition, this data establishes a crucial link between the protection and administration of biodiversity on local, national, and international scales and the study of local water resources (Saha et al., 2010). The development of geoinformation technology, specifically RS and GIS, has significantly improved the process of groundwater mapping and exploration.

Methods for mapping groundwater potential zones are classified into two categories: knowledge-driven techniques, such as the Analytical Hierarchy

Process (AHP), which are based on expert experiences and thus influenced by expert knowledge and subjectivity, and data-driven techniques, which involve probabilistic, statistical, and data mining approaches (Rahmati et al., 2015, Yin et al., 2018). Therefore, the evaluation of the groundwater potential zones (GWPZ) and production of the final groundwater potential map (GPM) will be conducted during this study using both knowledge-driven approaches (AHP method) and data-driven techniques by applying machine learning classifiers.

4.4 Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) is a decision-making methodology developed by the Iraqi-American scientist Thomas Saaty in the 1970s that uses a hierarchical structure to break down complex problems into smaller and more manageable components. It is a widely used technique in various fields, including engineering and environmental management. In AHP, decision-makers evaluate the relative importance of criteria, alternatives, and sub-criteria in pairwise comparisons. The evaluations are then used to derive weights for each criterion or alternative, which are combined to obtain an overall score or ranking. AHP has been applied in many contexts, including land-use planning, environmental impact assessment, and resource allocation. It has been recognized as a valuable tool for supporting decision-making in complex and uncertain environments (Roy et al., 2020).

This study employs a multi-criteria approach that incorporates seven parameters that have been identified as highly correlated with groundwater accumulation through previous investigations, professional judgment, and local circumstances since there are no established criteria for selecting groundwater potential influencing factors. Several studies have been made to identify the GWPZ and suggest that the availability of groundwater is influenced by various

factors, as indicated by Roy et al. (2020), Ifediegwu (2022), Chatterjee and Dutta (2022), and Asgher et al. (2022). These factors include the lithological composition, drainage pattern, land use and cover, as well as climatic variables such as precipitation, temperature, and evapotranspiration. Also, geological features such as fractures and lineaments control groundwater occurrence and availability.

Consequently, the study area was analyzed using field data, satellite data, and existing maps to prepare seven thematic maps of the influencing parameters. These parameters include hydrogeological aquifer unit types, transmissivity, lineaments density, slope, soil type, land use and land cover, and drainage density. The preparation of these maps was carried out using ArcGIS 10.7.1 software. These parameters were assigned weights through the utilization of AHP. Within each parameter, various classes (sub-criteria) were ranked based on their relative significance for groundwater potentiality.

The implementation of geospatial technologies can efficiently and economically generate significant model data from various domains of geoscience, as evidenced by studies conducted by Adiat et al. (2012), Russo et al. (2015), and Hussien et al. (2017). Groundwater storage may be inferred indirectly through the use of high and moderate geographical resolution satellite data. These satellites have the potential to provide a substantial quantity of unprocessed data associated with resource inventory and mapping. To facilitate effective decision-making, the data must be treated appropriately to extract relevant data regarding indirect variables (Madani and Niyazi, 2015; Hussien et al., 2020). In several areas of natural resource management, environmental impact analysis, and regional planning, the AHP technique for multicriteria decision-making (MCDM) has been successfully used by assessing the relative significance parameters, as evidenced by studies conducted by Chowdhury et al. (2010),

Kaliraj et al. (2014), Ouma and Tateishi (2014), Rahaman et al. (2015), Mallick et al. (2019), Zghibi et al. (2020), Roy et al. (2020), and Morgan et al. (2022). The AHP is the most commonly used GIS-based to facilitate decision-making by evaluating multiple factors, which provides a flexible, easy, and understandable method for analyzing complicated problems; this technique has gained attention as a useful approach for groundwater prediction due to its efficient and accurate nature in calculating the spatial associations between dependent and independent variables according to scores assigned for each parameter based on major and minor influencers affecting the GWPZ (Abrams et al., 2018; Meena et al., 2019, Milevski, 2019; Abijith et al., 2020; Arefin, 2020).

4.5 Methodology Used to Apply Analytic Hierarchy Process

By applying this methodology to groundwater exploration, it is possible to identify and rank potential zones for developing groundwater resources. These steps were accomplished using the GIS software as follows:

(1) Assigning rank values to each class based on Saaty's 1-9 scale. This is done in accordance with their respective contributions to groundwater occurrence when compared to other factors. A layer having a significant influence on groundwater potential is represented by a high-weight parameter, while a low-weight parameter represents a layer with little impact.

(2) Creating a pairwise comparison matrix (PCM) to provide the normalized weights to each class.

(3) Assigning weights to sub-classes by multiplying the normalized weight of each class with the rank of its respective sub-classes.

(4) Overlaying all the thematic maps using the overlay spatial analysis tool in ArcGIS 10.7.1.

(5) Validating the model using actual field data of wells to calculate the overall accuracy.

The methodology used to delineate suitable sites for groundwater availability is presented in Figure (4.1) to show the integrated approaches used to create the final map of groundwater potential zones during the current study.

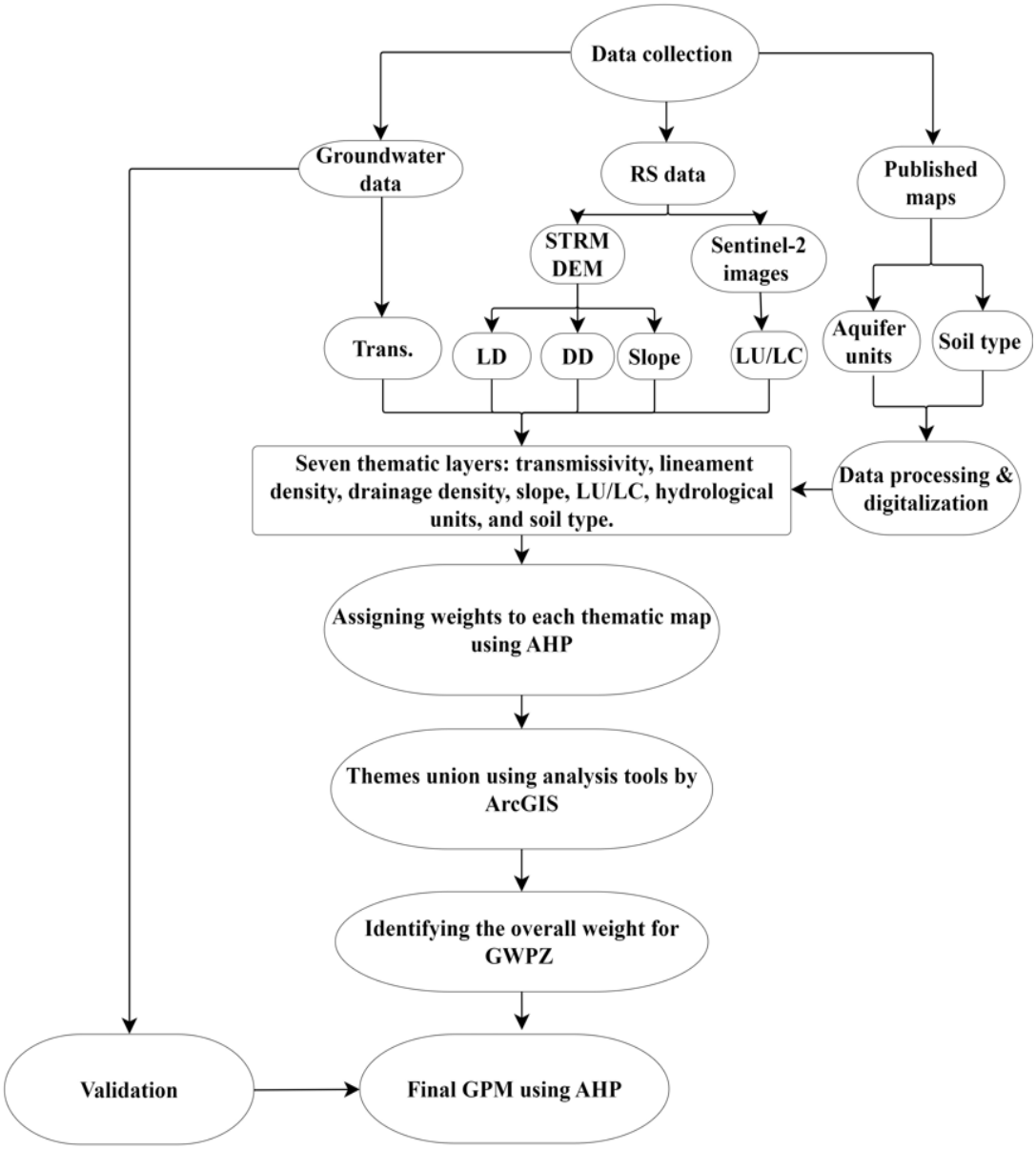


Figure (4.1): Framework to delineate the groundwater potential zones.

Remote sensing data, such as STRM DEM and sentinel-2 images, were prepared for digital image processing along with geological and soil maps in addition to the fieldwork data to create the characteristic groundwater layers needed to generate the potential groundwater map. The study area's DEM was obtained from the website (http://earthexplorer.usgs.gov) with 30m accuracy and SRTM type. In contrast, the sentinel-2 images were obtained from the website (https://www.arcgis.com/home/item.html?id=fc92d38533d440078f17678ebc20e8e2), which provides access to individual 10-meter resolution scenes for all land masses on the planet for each year from 2017-2021.

4.6 Creating an AHP Input Database

The present study has taken into account seven thematic layers to identify the groundwater potential zones: hydrogeological aquifer unit types, transmissivity (T), lineaments density (LD), slope (SLO), type of soils (SL), land use and land cover (LU/LC), and drainage density (DD). These factors significantly regulate groundwater availability, as reported in studies conducted by Mukherjee et al. (2012), Agarwal et al. (2013), and Roy et al. (2020).

4.6.1 Hydrogeological Aquifer Units

The study area exhibits four different aquifer units, which can be attributed to variances in geological formations. As illustrated in Figure (4.2), the hydrogeological units' thematic map was produced by applying existing data and georeferencing Iraq's primary groundwater aquifers map, created by Jassim and Goff (2006).

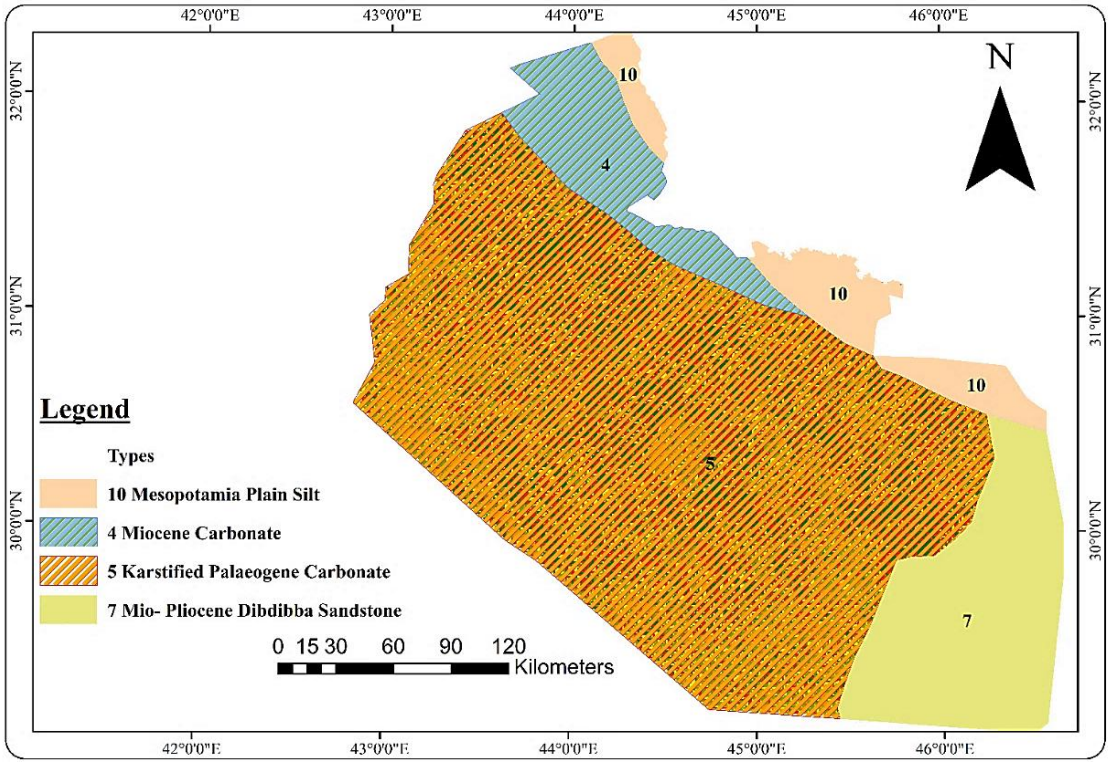


Figure (4.2): Thematic map for main hydrogeological units in the study area**.**

The ranking of hydrogeological units was found to be positively correlated with the static water level. Specifically, units with higher static water levels were assigned the highest rank, while aquifer units with lower static water levels were assigned the lowest rank. The Karstified Palaeogene aquifer comprises the Tayarat and Dammam aquifers and contains a significant amount of groundwater, indicating a high level of groundwater availability (Al-Jiburi, 2009). Furthermore, the Mio-Pliocene Sandstone, which is illustrated by the Dibdiba aquifer, exhibits a moderate level of groundwater availability. Conversely, the Miocene Carbonate, which is represented by the Euphrates and Zahra formations, displays a low level of groundwater availability. The Mesopotamia Plain Silt is distinguished by a notable scarcity of groundwater resources, resulting in very poor groundwater availability.

4.6.2 Transmissivity of Aquifer Layers

According to, the transmissivity of aquifer layers refers to the rate of flow across a unit-width cross-section of the fully saturated thickness of the water-bearing layer, under a unit hydraulic gradient. This parameter serves as an indicator of the aquifer's ability to transmit water through it. The Kriging interpolation technique was applied to produce a thematic map representing the spatial distribution of transmissivity in the study area, relying on the available data. The study region was classified into five groups, as shown in Figure (4.3).

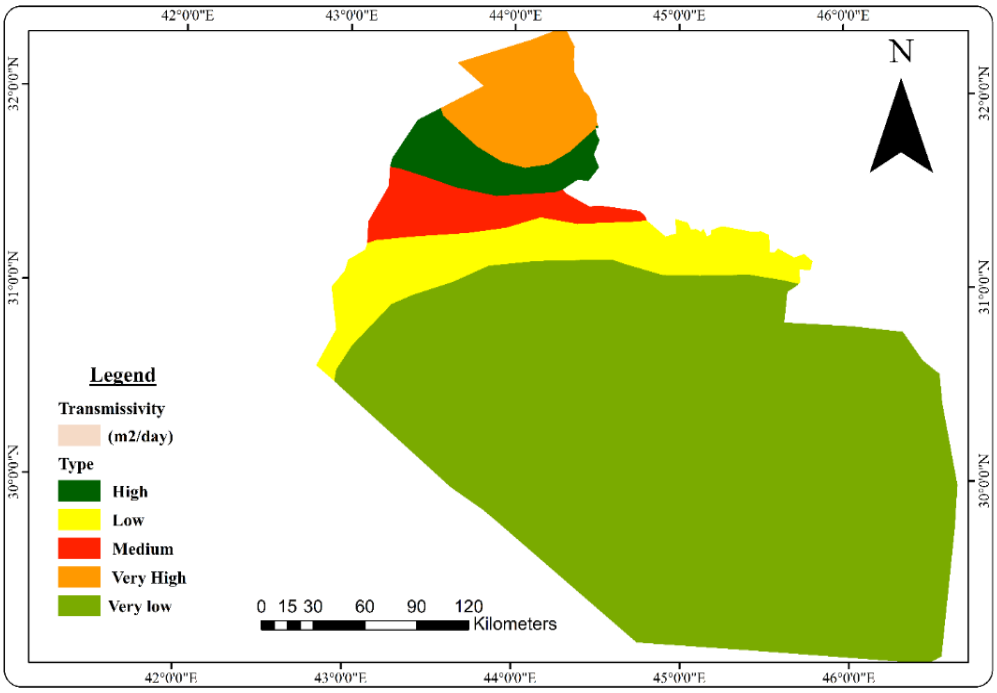


Figure (4.3): Thematic map of transmissivity.

According to Luijendijk (2022), a significant correlation exists between transmissivity, groundwater flow, and drainage density, in which high transmissivity results in low drainage density areas, while areas with low transmissivity have high drainage density. This is attributed to the fact that regions with high drainage density experience limited infiltration, leading to the formation of poor groundwater potential zones. As a result, the groundwater potentiality for

transmissivity range from 9.474 to 12.089 m^2/day was classified as "very poor," 12.089 to 17.081 m^2/day as "poor," 17.081 to 24.332 m^2/day as "moderate," 24.332 to 32.414 m^2/day as "good," and 32.414 to 39.784 m^2/day as "very good."

4.6.3 Lineament Density (LD)

According to Devi et al. (2001), it is expected that linear geological features will reflect a fractured zone characterized by favorable porosity and permeability. The topographic features of a region provide significant insights into the underlying fractures that can potentially impact the hydrological processes of groundwater flow and storage. According to Lemacha (2008), the presence of joints and fractures in a rock can increase its secondary porosity and permeability, thereby facilitating the availability and movement of groundwater. The density of lineaments density (LD) is considered a crucial factor in determining the availability and movement of groundwater in the study area, as evidenced by the tectonic map of Iraq. Limestone and dolomite beds are part of the carbonate aquifers found in the Dammam Formation. According to Al-Jibury and Al-Basrawi (2007), the presence of faults, fractures, and scattered karsts in these rocks allow regional groundwater flow into the southern desert area. In the context of groundwater potential analysis, it is important to consider the existence of lineaments. These linear features, representing secondary porosities, infiltrate surface runoff and recharge groundwater aquifers. A significant relationship exists between high lineament density and high groundwater potential (Al-Ruzouq et al., 2019). Figure (4.4) illustrates the spatial distribution of the structure lineaments within the examined region. Additionally, Figure (4.5) presents the classification of lineament density in the study area, which is separated into five distinct groups.

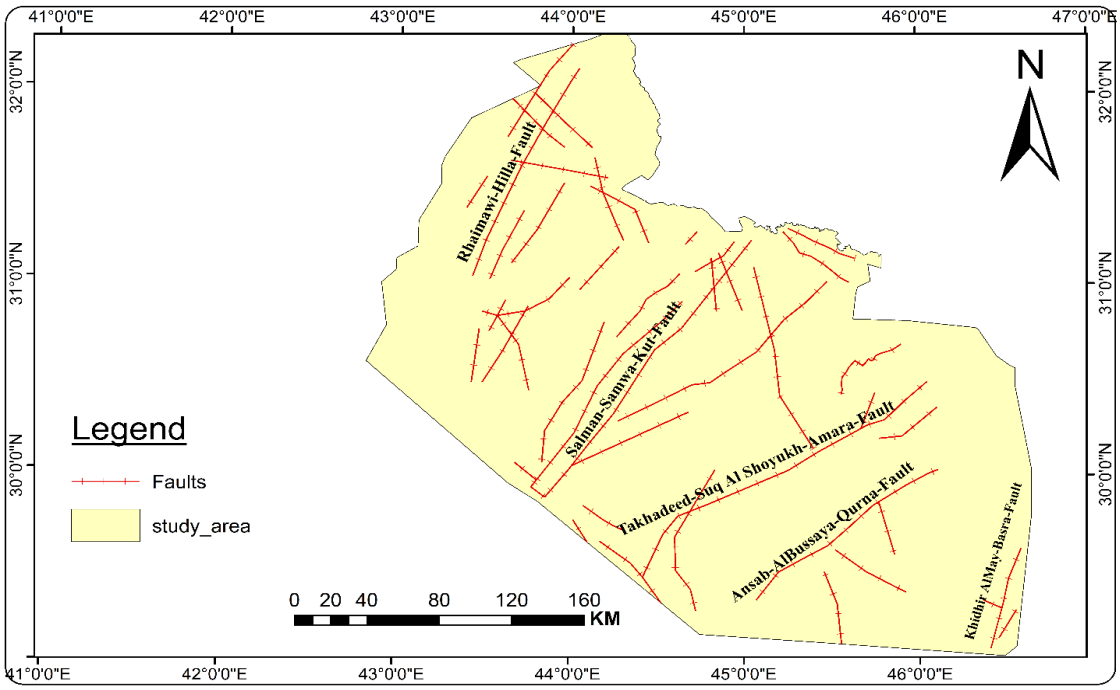


Figure (4.4): Faults map for the study area modified from the Iraq tectonic map (GEOSURV, 1994).

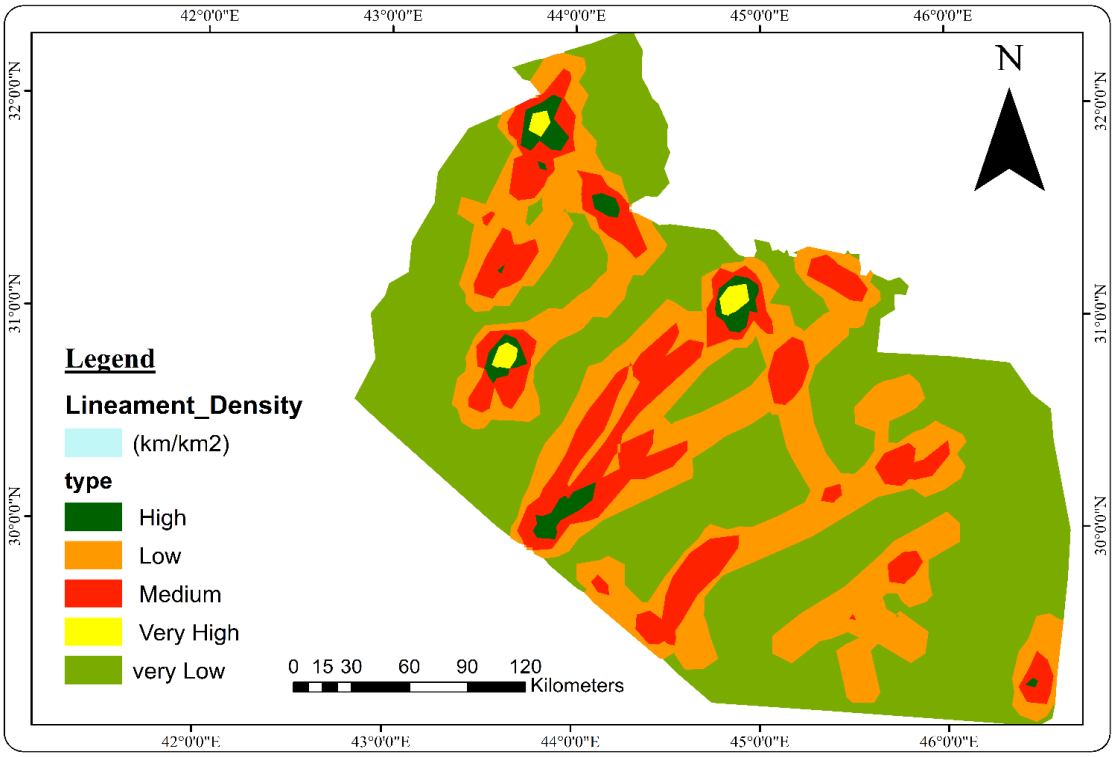


Figure (4.5): Thematic map of lineament density.

Based on the findings presented in Figure (4.5), it is evident that the regions exhibiting a high LD, ranging from 0.15 to 0.19 km/Km^2, are considered to possess significant potential for groundwater resources. This is attributed to their favorable porosity and infiltration characteristics, resulting in a higher ranking (very good) for such regions. The regions with a groundwater potentiality ranging from 0.11 to 0.15 km/Km^2 of LD were identified as good groundwater potential. On the other hand, regions revealing groundwater potentiality within the range of 0.07 to 0.11 km/Km^2 were classified as moderate. However, the areas of LD ranging from 0.04 to 0.07 km/Km^2 indicated poor groundwater potential zones. Regions characterized by a low LD ranging from 0 to 0.04 km/Km^2 are deemed to possess limited groundwater potential and thus assigned lower rankings.

4.6.4 Slope (SLO)

The slope has a considerable impact on the occurrence and movement of both surface water and groundwater. Studies have shown that the slope gradient influences both runoff and infiltration, with steeper slopes leading to increase runoff and decrease infiltration. Furthermore, the slope layer has been identified as a crucial variable in determining groundwater prospecting regions, as it can significantly impact the availability and distribution of groundwater resources. Hence, in the process of identifying groundwater-prospecting regions, the slope layer is considered an essential component (Ganapuram et al., 2009).

The present study involved the creation of a slope map, which was accomplished by employing a surface spatial analyst tool. Specifically, a slope tool was utilized to analyze the STRM DEM dataset with a resolution of 30 meters related to the year of 2018. The slope map generated for the study area indicated that the slope values varied between 0 and 2.9 degrees.

The process of reclassification was employed on the slope map in order to establish the weights and ranks for groundwater availability. This involved the division of the study area slope into five specific classes. These classes revealed a range of slope values, spanning from 0 to values exceeding 0.5°. The methodology employed in this study relied upon the classification of slopes as having low rank due to their limited ability to facilitate the infiltration of surface runoff, resulting in poor groundwater potential when their slopes exceed 0.5°. This classification was established by Roy et al. (2020). The resulting reclassified slope map is described in Figure (4.6).

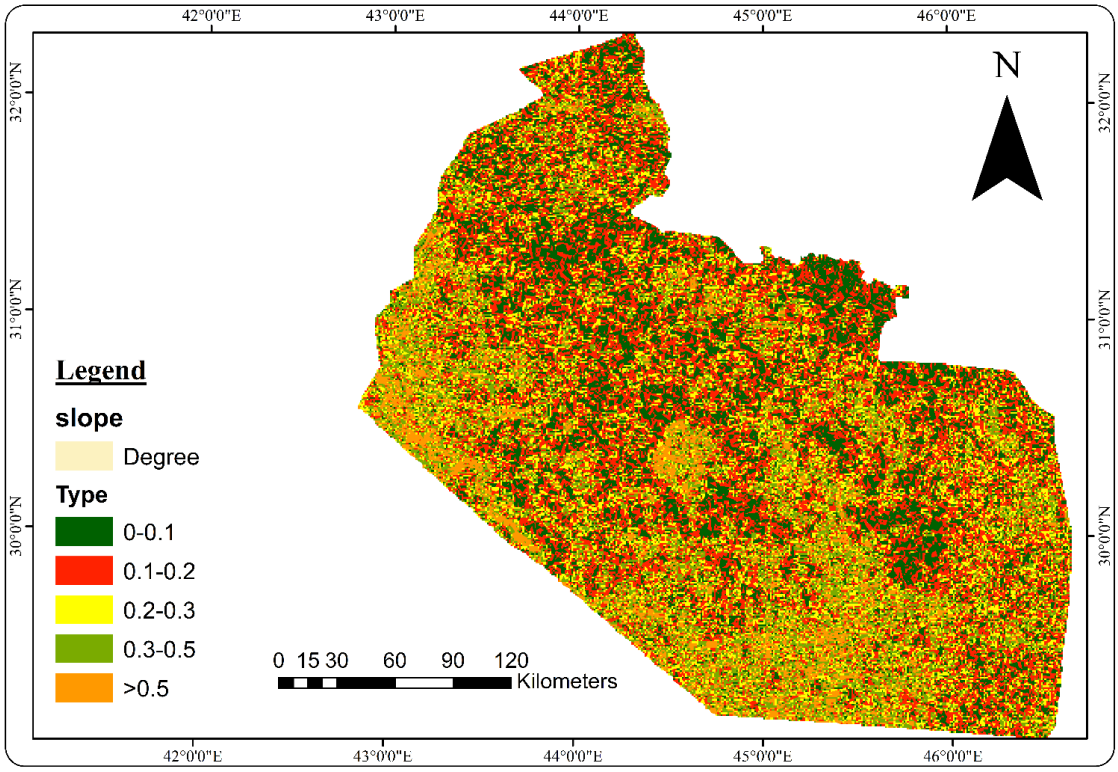


Figure (4.6): Thematic map of study area slope.

The slope of the study area was divided into five categories based on its variability. The areas with slope interval (0-0.1)° are considered (very good) due to flat topography and moderately decreased movement of runoff downstream, and increase the amount of infiltrated water, thus elevating the groundwater

potentiality. The slope of (0.1-0.3)^{o} classified as good to moderate, which is considered advantageous for groundwater recharge. Finally, the slope (0.3-0.5)^{o} (steep slope) and greater than 0.5^{o} (very steep slope) were regarded as (poor) to (very poor) groundwater potentiality, respectively, because the steep and very steep slopes were awarded low weights since they favour high amounts of runoff and do not give enough time for surface runoff to infiltrate. On the other hand, low-slope zones are suitable locations for water accumulation and infiltration.

4.6.5 Soil Type (SL)

Groundwater recharge is influenced by various factors related to soil properties, such as porosity, structure, adhesion, and consistency; these factors affect the ability of the soil to absorb and retain water, which in turn determines the rate at which groundwater is recharged (Lentswe and Molwalefhe, 2020).

Zghibi et al. (2020) stated that the gravel and sandy soil has a high degree of infiltration and, therefore, has a higher influence, while the clay and clay loam soil has the lowest degree of infiltration and, therefore, is given the lowest influence. The present study utilized soil texture data sourced from the digital soil map of the world, which was obtained in the ESRI shapefile format from the official website of the Food and Agriculture Organization (FAO) of the United Nations (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases). The digital global soil map was used to extract the soil map that corresponds to the study area. The map assigns a distinct numerical value to each soil type, which can be determined by referencing the user soil Excel file, as shown in Figure (4.7).

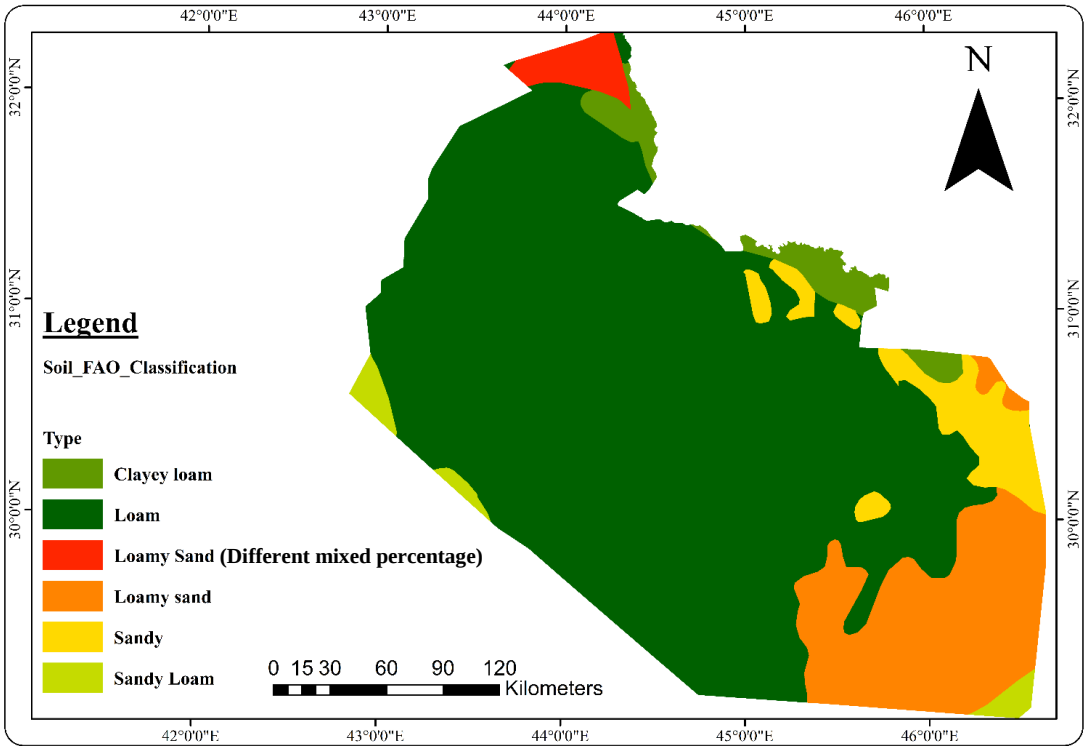


Figure (4.7): Thematic map for the soil classification in the study area.

The dominant soil textures in the study area are indicated by Figure (4.7), which shows that loam, loamy sand, sandy loam, sandy, and clayey loam were the prevailing attributes. Since the infiltration rate depends on soil grain size, coarse-textured soils (e.g., sand, loamy sand, and sandy loams) have a greater infiltration rate and, thus, a high groundwater potential. In contrast, medium-textured soils (loams and silt loams) have moderate to poor groundwater potentiality. On the other hand, fine-textured soils such as clay loams, silty–clay loams, and clays have the lowest infiltration rate and (very poor) groundwater potentiality.

4.6.6 Land Use/ Land Cover (LU/LC)

The groundwater resources in a region are heavily influenced by the area's land use and land cover (LU/LC), which also significantly impacts resource management. It impacts various hydrogeological water cycle phenomena, including evapotranspiration, infiltration, surface runoff, etc. (Kaliraj et al. 2014). LU/LC can significantly impact the potential zones of groundwater in a confined aquifer. These aquifers are recharged by water that seeps in from surrounding areas, typically at higher elevations. When land use changes occur in the recharge areas of a confined aquifer, the amount of water that enters the aquifer can be affected. The LU/LC map for the year 2021 was produced utilizing European Space Agency ESA Sentinel-2 imagery with a spatial resolution of 10 meters and a Universal Transverse Mercator (UTM) data projection. The map can be accessed from the following website:

(https://www.arcgis.com/home/item.html?id=fc92d38533d440078f17678ebc20e8e2). The relevant data covering the study area was obtained from a GeoTIFF file, which is a type of raster data. The data was processed using the raster proccing data management tools available in ArcMap 10.7.1 to clip the relevant raster data using the study area shape file. The conversion of the raster-generated map to a polygon is a crucial step that necessitates the utilization of conversion tools available in the arc toolbox to convert the generated map from raster to polygon format. This is due to the fact that the map assigns a unique numerical value to each LULC class, which can be identified by referring to the class definitions, as illustrated in Figure (4.8).

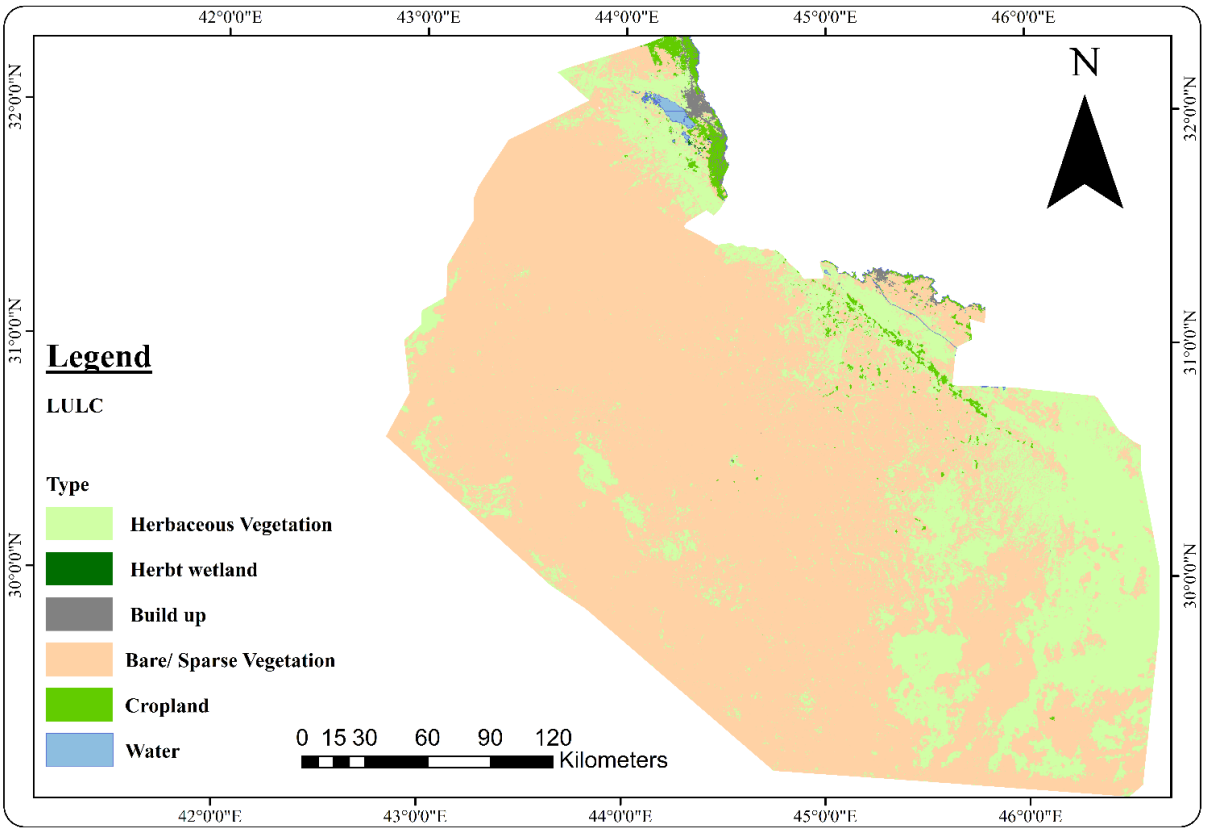


Figure (4.8): Thematic map of the study area Land Use and Land Cover.

Based on the thematic map of LU/LC for the study area shown in Figure (4.8). The study area has been divided into six classes: water, herbs wetland, herbaceous vegetation, cropland, and bare areas or sparse vegetation; each type has a weighted value representing the priority effect on groundwater accumulation. Runoff is often lower, and infiltration is higher in agricultural and forested regions compared to settlement and bare areas, where the infiltration rate typically decreases. As a result, settlement and bare areas are given a low rank, whereas agricultural land and the areas covered by vegetation are given a high rank (Roy et al., 2020).

4.6.7 Drainage Density (DD)

Drainage density can be defined as the ratio of the total length of streams within a given area to the corresponding drainage basin area. The unit of measurement is expressed in km/km^2, as stated by Harini et al. (2018); the drainage density or stream frequency is subject to diverse factors that can affect it, including:

i. Surface nature: the permeability of surfaces has an impact on the infiltration of water, leading to a reduction in channel formation and, subsequently, lower drainage densities.

ii. Climate: regions characterized by significant precipitation levels typically exhibit elevated drainage densities. However, this claim is not broadly true. For example, arid regions are recognized for showing higher drainage densities despite experiencing limited precipitation. This is a result of steep terrains associated with those regions, which facilitates the occurrence of surface water flow and the consequent development of rivers and channel systems. Additionally, the presence of hard impermeable surfaces is known to decrease infiltration. Conversely, regions with rainy climates, such as tropical areas, may have low drainage densities in cases where the terrain is relatively flat and there is a high rate of infiltration. Moreover, the existence of a greater amount of vegetation in the area accelerates the process of infiltration.

iii. Vegetation cover: the presence of vegetation cover may restrict the flow of runoff, as it serves as a barrier that intercepts and retains water, thereby reducing the probability of channel formation and subsequent contribution to streams. Drainage basins characterized by a low vegetation cover exhibit comparatively elevated drainage densities.

Accordingly, the drainage density indirectly influences groundwater availability due to its connection with the infiltration capacity and the permeability; higher drainage density reduces the infiltration, thus increasing the runoff, and higher runoff indicates less groundwater potentiality (Lentswe and Molwalefhe, 2020). Furthermore, different geologic and climate zone observations show that low drainage density is more likely to happen in a flat area with highly permeable subsoil and dense vegetation. High drainage density develops in areas with mountainous terrain, few plants, and impermeable subsurface (Waikar and Nilawar, 2014). As a result, Poor groundwater potential zones (GWPZs) result from limited infiltration in a high drainage density regions. Hence, a high rank is given the low drainage density area. The thematic map of drainage density for the study area was produced using DEM of the SRTM type, with a resolution of 30 meters for the year 2018, utilizing hydrology tools in spatial analysis tools included in ArcGIS 10.7.1 software. The drainage density map was generated after processing some steps, including:

(1) Fill in a surface raster using a fill tool to eliminate minor irregularities (small imperfections) in the dataset.
(2) Identify the flow direction that can be determined by utilizing the flow direction tool within the spatial analysis framework. The process generates a raster that delineates the flow direction from a given cell towards its steepest downslope neighbor or neighbors.
(3) Calculate the accumulated flow using a tool denoted as flow accumulation to compute the accumulated flow, which is defined as the sum of weights of all the cells that flow into each downslope cell present in the output raster. In the absence of a weight raster, a uniform weight of 1 is assigned to every cell. The resulting output raster contains the total number of cells that contribute to the flow of each cell.

(4) Assign the stream order; the process involves assigning a numerical sequence to the individual segments of a raster that correspond to the various branches of a linear network. The process of assigning a numerical order to links within a stream network is known as stream ordering. The present classification system serves as a means of stream identification and categorization, utilizing the number of tributaries as a primary determinant.

(5) Convert a raster representing a linear network to features representing the same linear network, which can be achieved by utilizing the stream to feature spatial analyst tool to allow the use of layer properties and change the symbology of the generated map.

(6) Calculates the density of linear features. The spatial analyst line density tool is utilized to compute the density of linear features by determining the density of linear features in the surrounding area of every output raster cell. A circle is drawn around each raster cell center and the length of each line that falls within the circle is divided by the circle's area. The density measurement is determined by the ratio of the unit length per unit of area, as shown in Figure (4.9).

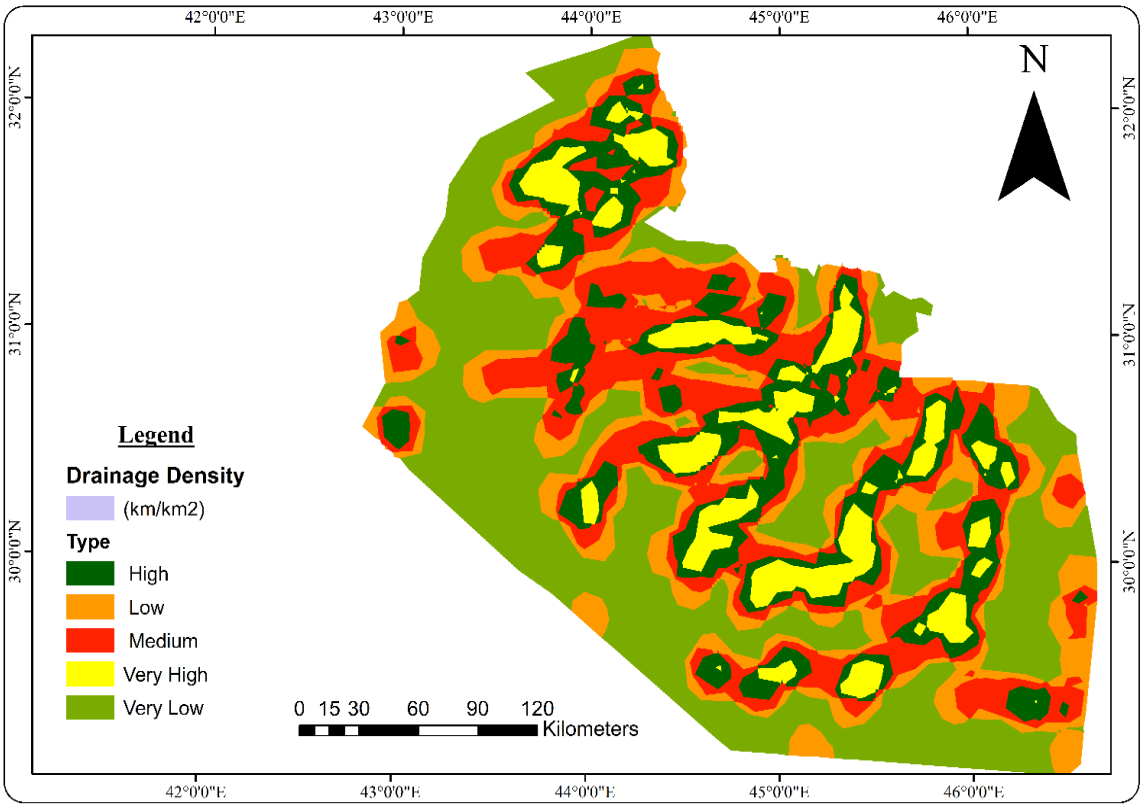


Figure (4.9): Thematic map of drainage density.

As shown in Figure (4.9), the area under study was categorized into five classes, each assigned a specific weight and assessed for its potential to yield groundwater. The assessment of groundwater potential for each class was conducted by considering its respective range of values.The findings indicate that higher drainage density reduces infiltration, thereby increasing runoff. Consequently, higher runoff values indicate a lower potential for groundwater availability. In light of this, the classes with higher drainage density values were assigned the lowest rank, while those with lower values were given a higher rank for groundwater availability, which were as follows: very poor (10.12 - 17.21) km/Km^2, poor (7.35 - 10.12) km/Km^2, moderate (4.66 -7.35) km/Km^2, good (1.75 - 4.66) km/Km^2, and very good (0 - 1.75) km/Km^2.

4.7 Assignment and Normalization of Weights for Thematic Layers

The creation of GWPZs was accomplished through the utilization of the overlay technique within the ArcGIS software, as described by Waikar and Nilawar (2014) and Chatterjee et al. (2020). This approach incorporated seven thematic layers and their corresponding percentage impact on the availability of groundwater. The overlay toolset comprises a set of tools that facilitate the superimposition of multiple feature classes for the purpose of merging (union), erasing, modifying, or updating spatial features, thereby generating a novel feature class. Spatial relationships between input features can be identified by overlaying one set of features with another, creating new information. The present study employed the union geoprocessing tools of the overlay toolset to perform a geometric union of the input features. The resulting feature class will contain all features and their corresponding attributes. To apply the union option, it is necessary for the feature layers to exhibit polygonal geometry.

Prior to the overlay process, a score ranging from 1 to 9 was assigned to each factor that affected groundwater potentiality based on its relative importance when compared to other factors through pairwise comparisons where a score of 1 indicates the equal influence of parameters and a score of 9 indicates the extreme importance of a parameter on groundwater recharge compared to the other factors, as directed by Saaty (1980) and presented in Table (4.1).

Table (4.1): Relative class rate scale for Analytic Hierarchy Process (AHP) as cited by (Saaty, 1980).

Importance	**Equal**		**Weak**		**Moderate**		**Moderate Plus**		**Strong**		**Strong Plus**		**Very Strong**		**Very, Very Strong**		**Extreme**	
Scale	1		2		3		4		5		6		7		8		9	
	1/9	1/8	1/7	1/6	1/5	1/4	1/3	1/2	1	2	3	4	5	6	7	8	9	> 9
Less Significant ←——————									**——————→ More Significant**									

Therefore, each thematic layer was transformed into a single rank ranging from 1 to 5, which corresponds to very poor, poor, moderate, good, and very good, respectively, in terms of their potential for groundwater storage.

Subsequently, a PCM is created, as described by Roy et al. (2020), Abrams et al. (2018), and Lentswe and Molwalefhe (2020), utilizing the Saaty scores obtained in the prior stage, as shown in equation (4.1):

$$A=\begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ x_{n1} & x_{n2} & x_{nn} \end{bmatrix} \quad (4.1)$$

Where A is a pairwise comparison matrix in which each element x_{nn} represents the relative importance of one parameter compared to another. The PCM matrix column was calculated in descending order of the influenced parameters and the normalized weights provided as presented in Table (4.2).

Table (4.2): Normalized pairwise comparison matrix (seven layers) developed for AHP-based groundwater potential zoning for Dammam aquifer in the study area.

Influence Factors	Aquifer-Units	T	LD	SLO	SL	LU/LC	DD	Normalized Weight
Hydrogeological Aquifer Units	7	6	5	4	3	2	1	0.38567
Transmissivity (T) m^2/day	7/2	6/2	5/2	4/2	3/2	2/2	1/2	0.19284
Lineament Density (LD) km/Km^2	7/3	6/3	5/3	4/3	3/3	2/3	1/3	0.12856
Slope (SLO) (Degree)	7/4	6/4	5/4	4/4	3/4	2/4	1/4	0.09642
Soil (SL)	7/5	6/5	5/5	4/5	3/5	2/5	1/5	0.07713
Land use Land cover (LU/LC)	7/6	6/6	5/6	4/6	3/6	2/6	1/6	0.06428
Drainage Density (DD)	7/7	6/7	5/7	4/7	3/7	2/7	1/7	0.05510

Good groundwater potentiality themes have the highest weights, while poor themes have the lowest. Following the weight assignment, ranks were assigned to sub-variables (subclasses) to determine the overall percentage weight. This was achieved by multiplying the normalized weight of each parameter with the rank of each sub-class. The resulting values were used to generate the groundwater potential zone map, which is given in Table (4.3).

Table **(4.3)**: Weights assigned for groundwater control parameters of Dammam aquifer within the study area.

Classes	Normalize weight %	Rank	Potential index	Overall weight
Hydrogeological Aquifer Units				
Karstified Palaeogene	38.567	4	good	154.268
Mio-Pliocene Sandstone		3	moderate	115.702
Miocene Carbonate		2	poor	77.135
Mesopotamia Plain Silt		1	very poor	38.567
Transmissivity (T) (m^2/day)				
9.474-12.089	19.284	1	very poor	19.284
12.089-17.081		2	poor	38.567
17.081-24.332		3	moderate	57.851
24.332-32.414		4	good	77.135
32.414-39.784		5	very good	96.419
Lineament Density (LD) (km/Km^2)				
0-0.022	12.856	1	very poor	12.856
0.022-0.054		2	poor	25.712
0.054-0.079		3	moderate	38.567
0.079-0.114		4	good	51.423
0.114-0.185		5	very good	64.279
Slope (SLO) (Degree)				
0-0.1	9.642	5	very good	48.209
0.1-0.2		4	good	38.567
0.2-0.3		3	moderate	28.926
0.3-0.5		2	poor	19.284
>0.5		1	very poor	9.642
Soil (SL) Classification				
Loam	7.713	2	poor	15.427
Loamy Sand		3	moderate	23.139
Sandy Loam		4	good	30.852
Sandy		5	very good	38.567
Clayey loam		1	very poor	7.713

Classes	Normalize weight %	Rank	Potential index	Overall weight
Land use Land cover (LU/LC)				
Water	6.428	5	very good	32.140
Herbst wetland		5	very good	32.140
Herbaceous Vegetation		4	good	25.712
Cropland		3	moderate	19.284
Bare/ Sparse Vegetation		2	poor	12.856
Build up		1	Very poor	6.428
Drainage Density (DD) (km/Km^2)				
0-1.754	5.510	5	very good	27.548
1.754-4.656		4	good	22.039
4.656-7.354		3	moderate	16.529
7.354-10.121		2	poor	11.019
10.121-17.205		1	very poor	5.510

Five groups were established from the final groundwater potential zone map: very good, good, moderate, poor, and very poor. A GIS-based process of analyzing and visualizing groundwater potential using multiple thematic layers and assigning weights to each layer based on their relative importance is performed as follows:

(1) Use the percentage effect that represents the normalized weight to determine the overall weight of each sub-class of the seven thematic layers with regard to their impact on groundwater availability.

(2) Add a new column to the attribute table of each of the seven layers shape files to assign overall weights to each sub-class.

(3) Use ArcGIS's spatial analysis tool to select the overlay option.

(4) Apply the union option by inserting the seven layers as input to create a new shape file with all seven weights in the attribute table.

(5) Calculate the sum of all seven weights to produce the final GPM in a rank range (1-5).

(6) The rank range indicates the groundwater potential zones, with 5 being very high potential and 1 being very poor potential.

4.8 Generation of Groundwater Potential Map

The study area was classified into the five zones of groundwater availability by using the results of the integrated Geographical Information Systems (GIS) and the Analytical Hierarchy Process (AHP) techniques. The spatial distribution of GWPZ classes and the percentages of areal distribution are shown in Figures (4.10) and (4.11), respectively.

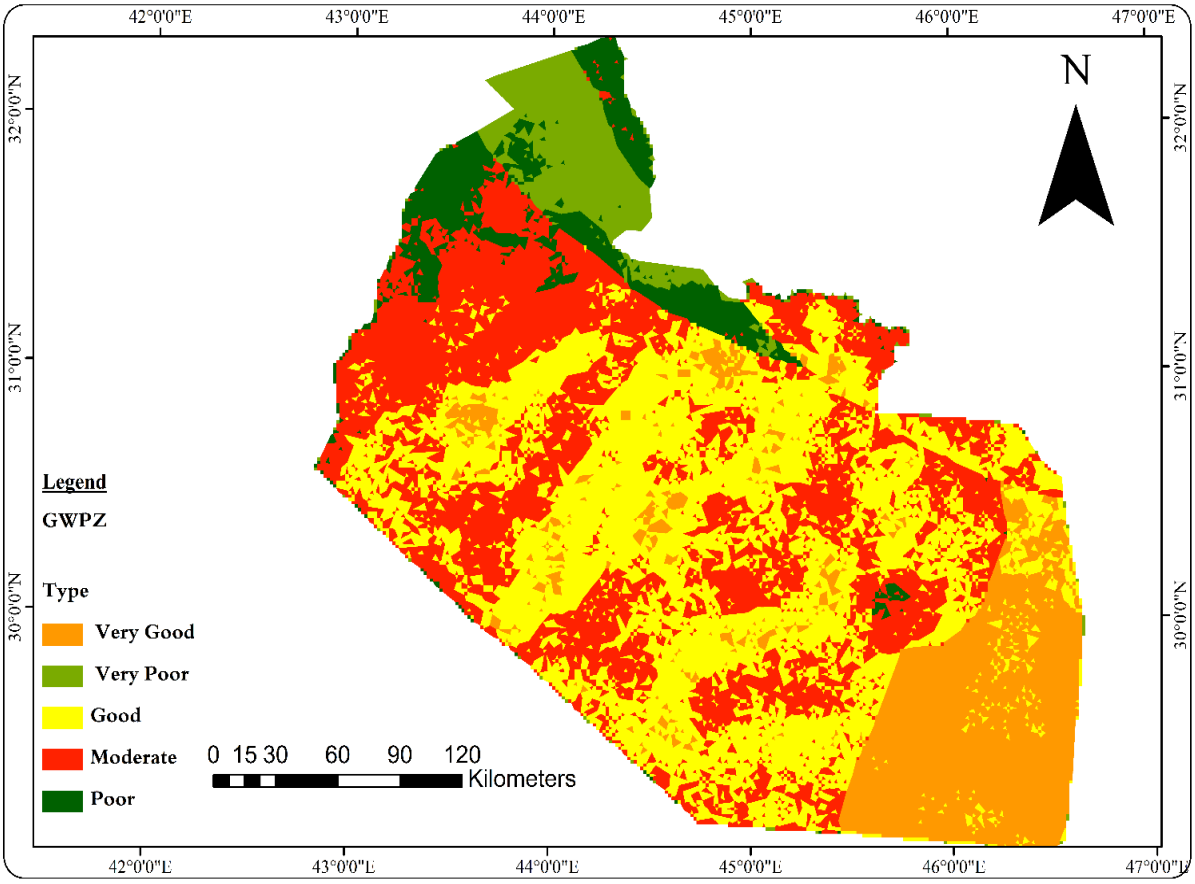


Figure (4.10): Groundwater potential zones (GWPs) of the study area.

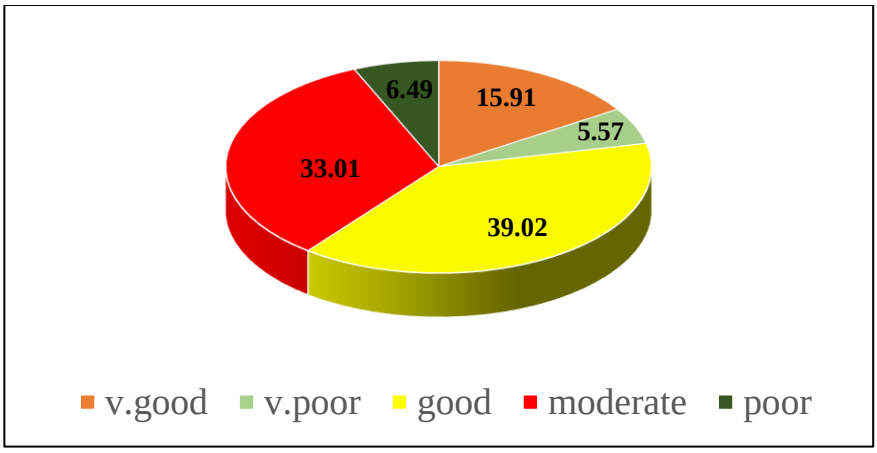


Figure (4.11): Percentage of GWPZs areal distribution.

Based on the findings illustrated in Figures (4.10) and (4.11), the classes of groundwater potential zone (GWPZ) map for the study area is given in Table (4.4).

Table (4.4): Groundwater potential zones of the study area.

Rank	Class attribution	Percentage Area	Area (km^2)
5	v.good	15.91	12208.51
4	Good	39.02	29942.26
3	Moderate	33.01	25326.31
2	Poor	6.49	4982.02
1	v.poor	5.57	4273.61

4.9 Verification of Groundwater Potential Zones

To validate the potentiality of groundwater, a total of 349 wells were used as an indicator of groundwater availability. The verification process involved analyzing the wells' yield data within the study area. The depths of the wells ranged from 75 to 346 meters, while their yield ranged from 2 to 20 ℓps. The well yield data was used to confirm the accuracy of the groundwater potential zones.

The data of the wells in the study area were categorized into five categories (very poor) for pumping rates equal to or below 3 ℓps, (poor) for pumping rates between 4 and 6 ℓps, (moderate) for pumping rates between 7 and 9 ℓps, (good) for pumping rate between 10 and 14 ℓps and (very good) for pumping rates above 14 ℓps, as proposed by Al-Abadi and Al-Shamma'a (2014) and Karim and Al-Manmi (2019). The number of wells with very poor, poor, moderate, good, and very good yields are (73, 32, 78, 138, and 28), with the proportion to the total wells (349) being (20.9, 9.2, 22.3, 39.5, and 8.0), respectively. A comparison was performed between the groundwater yields for each well and the corresponding groundwater potential zones to calculate the overall accuracy expressed as follows (Roy et al., 2020):

$$\text{Overall accuracy} = \frac{\text{No. of correct observation well location}}{\sum No. \text{ of observation well location}} \qquad (4.2)$$

Thus, the total number of correctly observed wells' locations is 288. Therefore, the overall accuracy is calculated to be (82.5 %).

Chapter Five

Theoretical Foundations and Practical Applications of Machine Learning

5.1 Introduction

Machine learning techniques are increasingly important in the management of groundwater resources due to their ability to analyze large amounts of complex data and provide insights that can inform decision-making. By applying machine learning algorithms to data from sensors, satellite imagery, and other sources, water resource investigators can more accurately monitor groundwater heads, predict changes in aquifer recharge and discharge, and identify potential areas of contamination or overuse (Liu et al., 2021). The implementation of this approach can lead to more effective and efficient management of groundwater resources, helping to ensure their long-term sustainability and availability for human needs. Additionally, machine-learning techniques can help to identify patterns and trends that may not be immediately apparent to human analysts, providing new opportunities for innovative and data-driven approaches to groundwater management (Ghobadi and Kang, 2023).

Usually, a variety of intricately interconnected complicated variables have an impact on hydrological processes. These processes are also featured by uncertainties in variable estimation. Hydrological systems' physical and numerical modeling demands comprehensive, accurate, and in-depth data on all affected aspects, which is not always attainable due to technological and economical constraints (Gao et al., 2020). Therefore, the potential outcome of this circumstance is the evidence of performance issues in the model, which could

subsequently exert an adverse influence on the formulation of hydrological plans and policies.

Data-driven models like artificial neural networks (ANNs) have shown the capability of input-output simulation and prediction due to their ability to learn the nonlinear relationship between input variables and outputs without explicitly referencing the underlying physical process (Li et al., 2020). Artificial intelligence is a computer programs designed to think and act like humans, and it can rationalize and take actions likely to achieve a specific goal with precise predictions, even with limited data (Sarker, 2022).

5.2 Techniques of Machine Learning

After an extensive review of relevant research on the utilization of artificial intelligence techniques for the analysis and management of groundwater resources, several studies have been identified as significant sources of information, including (Ripley, 2007; Witten and Frank, 2002; Kotsiantis, 2007; Kanevski et al., 2009; Marsland, 2015; Chipman et al., 2010; Ye, 2015; Da Silva et al., 2017; Mahesh, 2020). A concentrated definition can be deduced as Machine Learning (ML), a subset of Artificial Intelligence, which uses algorithms that can learn patterns and relationships from data to predict outcomes by connecting input data (variables) to the outputs. The type of algorithm chosen depends on the problem, the number of variables, and the best model that fits the data. ML has emerged as a tool for predictive analysis, using automated adaptive techniques to recognize patterns and make predictions and then apply the learned correlations to additional comparable data for classification and regression problems.

The goal is to understand the complicated relationships between natural occurrences to develop well-informed conclusions that aid human comprehension.

Supervised and Unsupervised Machine Learning are the primary approaches to ML, as shown in Figure (5.1).

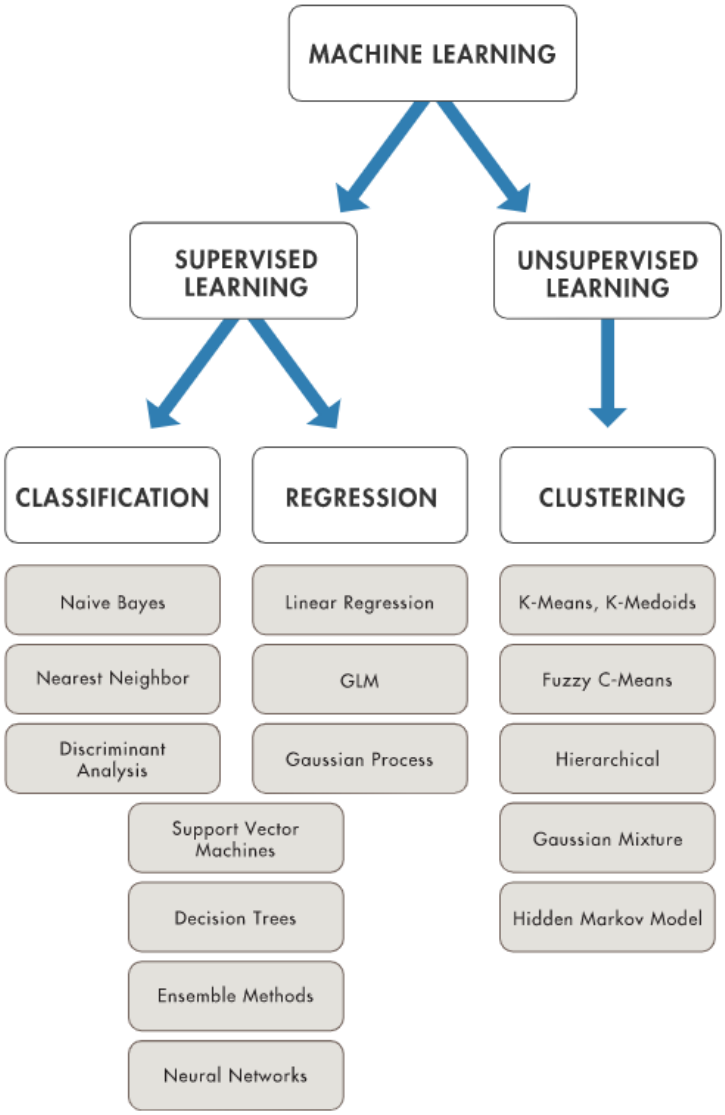


Figure (5.1): A diagram representing machine learning techniques (MathWorks user's guide: deep learning toolbox, 2020).

Unsupervised learning is a machine learning technique that uses unlabeled data to infer patterns and structures. Unlike supervised learning, there is no target output to guide the model training, which makes it impossible to evaluate the results quantitatively due to the lack of prior knowledge of expected outcomes.

Instead, the model explores the data's structure to extract feasible information and discover natural and coherent groupings within data to gain insight into how the data is organized (Da Silva et al., 2017; Hastie et al., 2009). Unsupervised learning technique called clustering divides the unlabeled dataset into a number of subgroups or clusters to identify and classify similarities between data points by minimizing the sum of the distances between the data points and their corresponding groups (Feng et al., 2019). In contrast, optimization is the method of iteratively training the model to provide an assessment of the maximum and minimum function, and it may be used in both supervised and unsupervised machine learning. To get the best results for machine learning models, the results are compared at each stage by adjusting the selected hyperparameters, such as the architecture of ML models, the number of epochs, the number of branches in a decision tree, etc. Until an accurate model with a low error rate was attained (Pereira et al., 2021).

Supervised learning involves using historical data to learn patterns and relationships between input and output variables. The data is split into training data, which contains known labels or observations, and testing data, which is used to evaluate the model's predictive performance (Agatonovic and Beresford, 2000). There are two types of supervised learning models: regression and classification models. Regression models are a type of supervised machine learning method that can be utilized for the purpose of predicting numerical outputs. This is achieved by examining the correlation between the input features, which are also known as independent variables, and the outcome, which is referred to as the dependent variable. The process of machine learning regression typically involves determining the optimal line of best fit for a given set of data points. The line of best fit is generated through the process of minimizing the distance between each data point and the line. The type of regression used in this case is simple

linear regression (Trunfio et al., 2021). Multiple linear and logistic regression are the most involved machine learning regression models; in case of more than one independent variable is employed, multiple linear regression is performed. While if the outputs of a regression model must be one of two values, the regression is known as logistic regression.

On the other hand, classification models adopted in this study are applied to predict categorical or class labels by classifying data into distinct groups and predicting the classification of new data points through the utilization of a supervised machine learning algorithm (Agatonovic and Beresford, 2000). The classification model is trained using the training data to connect input variables to their corresponding classes while minimizing the error criteria. Once trained, the model can predict the class labels for previously unseen data (Hastie et al., 2009). According to Kumari and Srivastava (2017), the classifier is the classification algorithm that recognizes the dataset classes. A multiclass classifier is used when there are more than two possible outcomes, while a binary classifier is used when there are only two possible outcomes for the classification task.

The present study involves the generation of a groundwater potential map through the utilization of multiclass and binary classifiers. This is accomplished by employing five distinct machine-learning approaches within the MATLAB R2020b software environment due to its capacity to manage big datasets, offer machine-learning applications, and extract features from signals and images through automated methods or generated codes. The main objective of this study is to conduct a comparative analysis of the outcomes produced by these approaches, with the aim of identifying the most accurate model. This model is subsequently utilized to delineate and map the groundwater potential zones within the specified study area. Using multiclass classifiers, the groundwater potential zone map was categorized into five classes, including very high, high, moderate,

low, and very low. While in binary classification, potential groundwater mapping was created as a two-classes system for identifying high-yield groundwater potential zones from those with low-yield potential based on a specified threshold value.

5.3 Approaches for Classification in Artificial Intelligence

The following is an overview of the foundations of five common supervised classification techniques that were used in the creation of hybrid machine learning models to produce the GPM of the study area:

5.3.1 Statistical Learning Algorithms

According to Guyon (2009), statistical learning algorithms are designed to construct a designated statistical model that can evaluate the probability of a given sample being assigned to a specific class. This approach encompasses various techniques, such as Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), which utilize density estimation approaches to identify optimal linear (or quadratic) combinations of variables that effectively partition groups into distinct categories. In contrast, Bayesian models such as Naive Bayes (NB) and Bayesian Networks (BN) offer potential alternatives for supervised categorization based on probability. In cases where there are high-dimensional input variables that exhibit nonlinear correlations among multiclass targets, the Bayesian models are more favorable compared to LDA and QDA, as stated by Kotsiantis (2007). The basic assumption of NB assumes that continuous variables correspond to a distribution that is normally distributed. In cases where this assumption is invalid, performance will be improved by using kernel density estimation methods that do not require a specific distribution. This method enables the user to conduct a more comprehensive analysis of the investigated probability distribution compared to a traditional histogram. In contrast to the histogram, the

kernel approach generates a smooth estimated curve for a specified dataset that incorporates all sample point positions (Węglarczyk, 2018).

The objective of kernel estimators is to estimate the probability density function of a given dataset. In the process of constructing a histogram, it is essential to take into account the width and endpoints of the bins. The width of the bins is determined by dividing the entire data interval into identical subintervals, while the endpoints of the bins represent the starting points of each bin. Histograms are known to have certain limitations, such as lack of smoothness, dependence on bin width and endpoints, and interpretational difficulties. However, these limitations can be overcome by utilizing kernel density estimators, as suggested by Xu et al. (2015). Kernel density estimators involve the utilization of two key parameters, namely the shape of kernel function (K) and bandwidth (h), to assign weights to the distances between observational data and a given point. The degree of required smoothness varies with different smoothing parameters, as shown in Figure (5.2) (Węglarczyk, 2018).

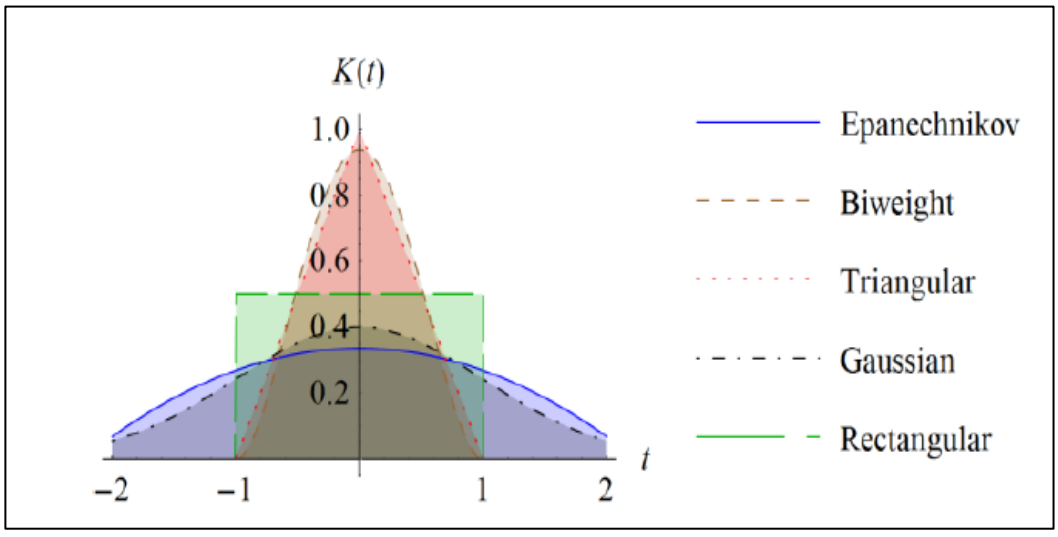


Figure (5.2): Shapes of symmetric kernels for different smoothing parameters (Węglarczyk, 2018).

5.3.2 Instance-Based Learners

In the field of supervised machine learning, there has been a recent development of several instance-based techniques. The concept of instance-based learning refers to a group of techniques employed for classification and regression tasks. These techniques generate a class label or prediction by assessing the distance or similarity metrics computations between the tested and the training samples. According to Alfeilat et al. (2019), the classification of a given data point is based on the majority class of its neighboring training data.

The k-nearest neighbor (kNN) classifier is an instance-based method that is commonly utilized due to its simplicity and effectiveness, showing performance that competes with the more intricate classifiers documented in the literature. The KNN classifier is used to generate categorizations or predictions regarding the grouping of an individual data point. This is based on the assumption that data points with similar attributes are likely to be located in close proximity to each other. In the context of classification tasks, the process of selecting a class label is based on the majority vote approach, in which the label assigned to a given data point corresponds to the one most frequently observed around that specific point. The process of classifying a new data point is shown in Figure (5.3).

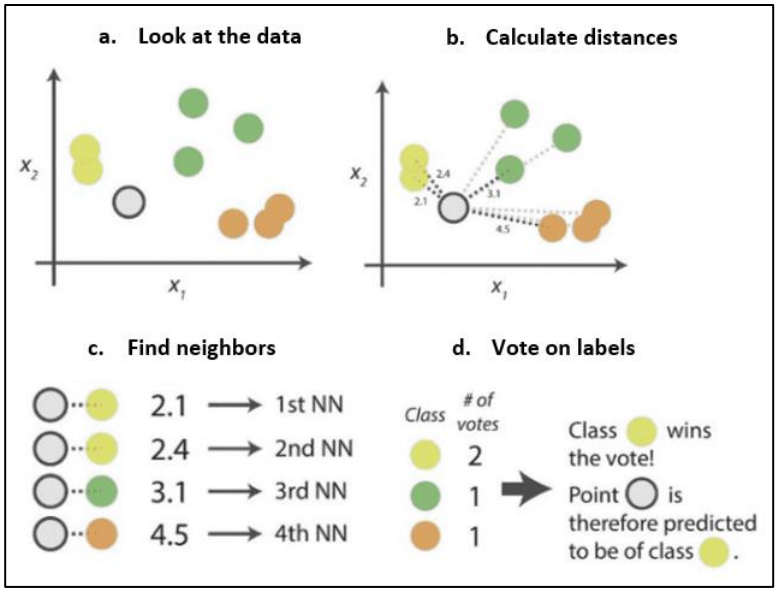


Figure (5.3): kNN classification (Alfeilat et al., 2019).

In Figure (5.3), the aim is to classify the grey data point, wherein three distinct categories are available, namely green, orange, and lime green. The kNN algorithm begins by computing the distance between the grey data point, which signifies the new data point to be categorized, and all other data points present in the figure. Subsequently, the identification of the nearest neighbors can be achieved through the process of arranging the locations in descending order based on their respective distances. The grey point's nearest neighbors (NNs) are defined as the data points that exhibit the closest proximity within the dataset. Ultimately, a decision will be made through a voting process regarding the predicted classification labels, which are determined by the classes of the k nearest neighboring data points. The labels were predicted utilizing the k-nearest neighbor algorithm with a value of k=3. The parameter 'k' in k-Nearest Neighbors (kNN) algorithm denotes the number of nearest neighbors to be considered for the purpose of voting, which ultimately determines the majority class. The kNN algorithm employs a similarity measure, commonly referred to as a distance function, to ascertain the distance between two given data points. The Euclidean

distance function is commonly utilized in addition to the Manhattan and Minkowski distances, as documented by Kotsiantis (2007) and Singh et al. (2013) and outlined in (Table 5.1).

Table (5.1): Distance metrics for multidimensional variables.

Euclidian	$D(x,y) = \sqrt{\sum_{i=1}^{n} \overline{\lvert x_i - y_i \rvert^2}}$
Manhattan	$D(x,y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$
Minkowski	$D(x,y) = \left(\sum_{i=1}^{n} \lvert x_i - y_i \rvert^p \right)^{1/p}$

Where x_i: stands for the new data point to be classified; y_i: stands for any data point within the training dataset; n stands for the feature (variables) size; and p: parameter when adjusting allows creating new distance measurements.

The Minkowski distance is a generalized metric measured in a normalized vector space. For a different value of p, Minkowski distance has different particular meanings. As p=1, Minkowski distance is specific to Manhattan distance, while p = 2, Minkowski distance is specific to Euclidean distance.

5.3.3 Logic-Based Learners

Classification algorithms have the ability to simultaneously process a significant amount of information in the field of data mining. The process of categorizing information based on training sets and class labels, frequently used for grouping purposes, allows for the establishment of assumptions regarding categorical class names and the classification of newly obtained data. Learners that adopt a logic-based approach will make predictions based on a predetermined

set of criteria or rules. In the domain of machine learning, the most basic form of logic-based learners consists of a sequence of Boolean decisions, such as "and," "or," and "if else," which establish classification rules and are generated automatically (Nikam, 2015; Gavankar and Sawarkar, 2017).

Decision Trees (DTs) serve as the fundamental basis for advanced algorithms that employ this particular learning methodology. Logic-based learners are represented through these algorithms, which are utilized to characterize, categorize, and generalize data. Decision Trees are a widely employed methodology that has proven to be effective in diverse domains, such as machine learning, image analysis, and pattern identification. The classification model takes the form of a tree structure, wherein internal nodes signify the attributes of a given dataset, branches represent decision rules, and each leaf node denotes the classification outcome. The model operates by following a path from the root to a leaf node, which is characterized by a sequence of data separation steps until a Boolean outcome is reached (Yang, 2019; Liang et al., 2019; Suresh et al., 2020; Jiao et al., 2020 and). The structure of DT is illustrated in Figure (5.4).

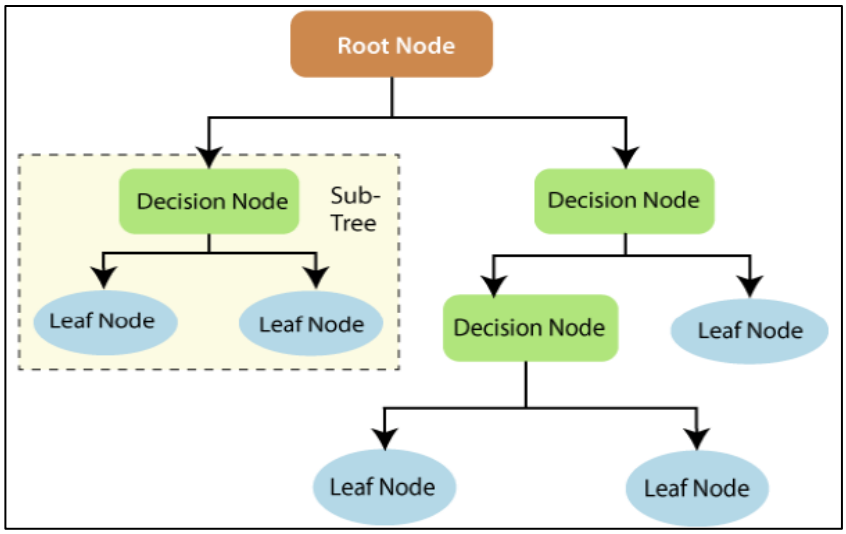


Figure (5.4): Structure of Decision Tree classifier (Suresh et al., 2020).

5.3.4 Support Vector Machines

Support vector machine (SVM) is a model for supervised machine learning that uses classification techniques. It has two key advantages: faster processing and better performance with data. Maximizing the margin between two classes can establish decision boundaries between classes in a high-dimensional variable space. Margin refers to the distance between the hyperplane and the data points closest to the hyperplane, known as the support vector, according to the work of Wang et al. (2017). A data point on either side of the hyperplane may be assigned to a different category, as shown in Figure (5.5), which illustrates the SVM main terms.

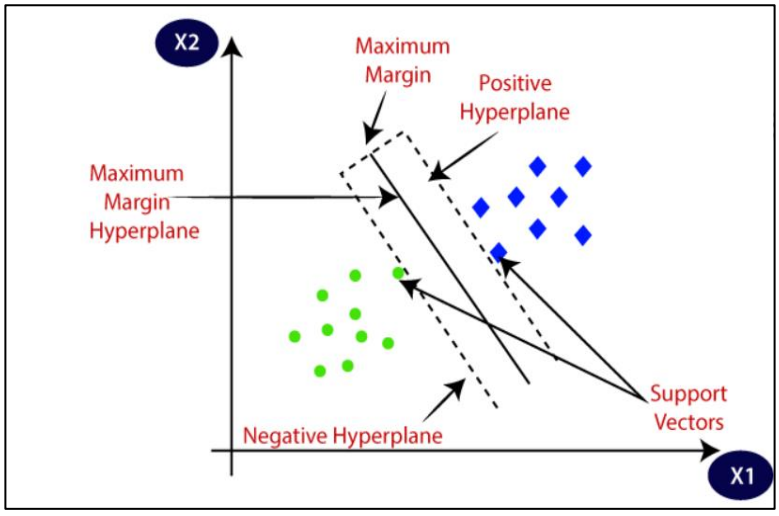


Figure (5.5): SVM main terms (Wang et al., 2017).

Nonlinear Support Vector Machines are employed in cases where the data cannot be separated into two distinct classes through a linear boundary in a two-dimensional feature space. The utilization of advanced techniques, such as the Kernel function, is necessary to classify data points. These techniques involve the use of specific quadratic functions that facilitate the conversion of a space with lower dimensions to a space with higher dimensions. This conversion enables the identification of a decision boundary that effectively separates the data points. In

SVM, there is no need to perform the transformation process manually or involve complicated calculations. SVM handles this process itself by choosing the appropriate Kernel type to locate a decision surface that correctly separates the data points, as presented in Figure (5.6) (Achirul et al., 2018).

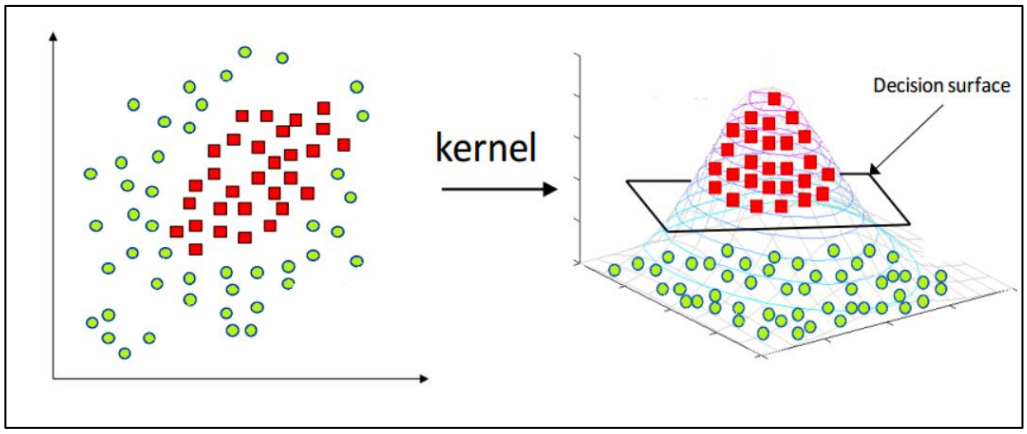


Figure (5.6): Kernel transformation to Gaussian distribution (Achirul,etal. 2018).

The kernel function is a crucial component in the development of a classification model using Support Vector Machines (SVM), as it facilitates the mapping of the dataset to a higher dimensional space, thereby enhancing the interpretability of the resulting classification model. Several kinds of Kernel functions can be used, such as linear, radial basis function, sigmoid, Gaussian, and polynomial.

In the current study, medium Gaussian SVM was applied for training data; the kernel function used for data transforming into the higher dimension to generate the Gaussian surface makes the data easily separable. Mahesh (2020) has stated that this function serves to establish boundaries between different categories. The margins are constructed in a way that maximizes the distance between the margin and the classes, ultimately leading to a reduction in classification error.

5.3.5 Perceptrons

The Perceptron algorithm is a supervised learning technique that comprises four primary components: input values, weights and biases, net sum, and an activation function. The principal components illustrated in Figure (5.7) and can be generalized, as stated by Dai et al. (2022):

$$y_i = f_k(\sum_i W_i X_i) \quad (5.1)$$

where W_{ji} represents the weight that may be adjusted for the *i*th instance and X_i represents one of the variables that are supplied. The activation function, denoted by f_k, has the flexibility to adopt various nonlinear forms.

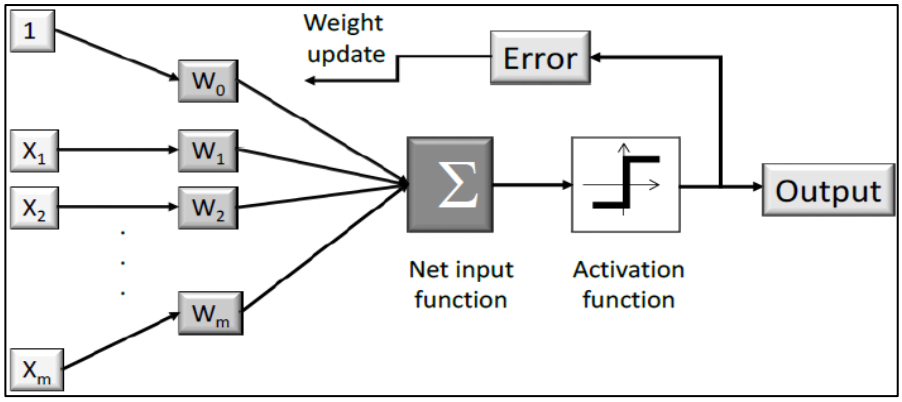


Figure (5.7): The Perceptron workflow (Dai et al., 2022).

Figure (5.7) represents how the perceptron gets the inputs X_1, X_2..., and X_m. The box in the top left with the symbol "1" symbolizes a bias in the received data. The inputs are combined with the weights W_i to calculate the net input function as $(\sum_i W_i X_i)$. The step-linear activation function receives the net input function. During the training process, the outputs are used to determine the error associated with the prediction. This information is then back-propagated to update the weights to minimize the misfit between the actual and predicted outputs (Dai et al., 2022). While in the case of nonlinear problems, the activation functions that are frequently employed include Uni-polar sigmoid, Bi-polar sigmoid, hyperbolic

tangent function (Tanh), softmax, Conic Section, and Radial Bases Function (RBF). These activation functions hold significant usefulness in training neural networks through the implementation of backpropagation techniques. This method employs the error between the computed output and the desired output to adjust the weight of the network, as noted by Raschka and Mirjalili (2017).

Artificial Neural Network Perceptron can be classified into two different types, namely the single-layer perceptron (SLP) and the Multilayer Perceptron MLP. The Single-Layer Perceptron (SLP) is a feed-forward network, that operates based on a threshold transfer function. This model does not include a hidden layer, unlike the Multilayer Perceptron. The computation of the single-layer Perceptron involves the summation of the input vector multiplied by its corresponding weight vector. The output layer will be determined by the results of the activation function, as stated by Wibowo and Wihayati (2021). A single-layer perceptron structure is presented in Figure (5.8).

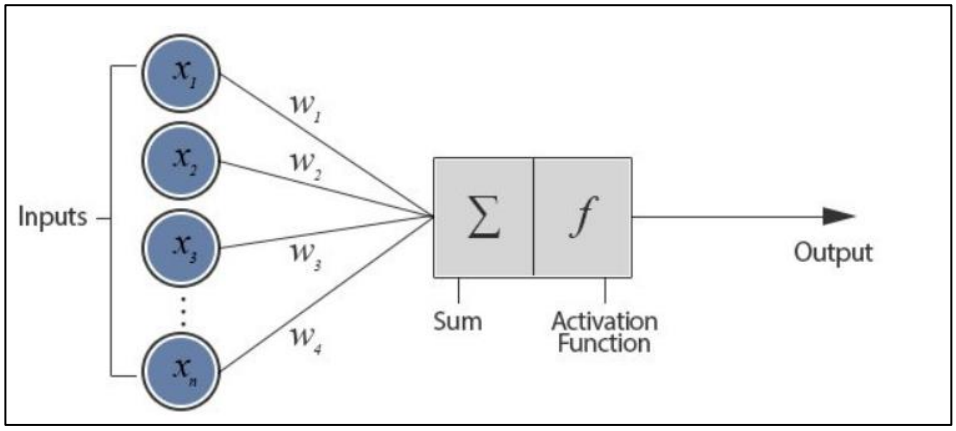


Figure (5.8): Single-layer perceptron (Wibowo and Wihayati, 2021).

The multilayer perceptron (MLP) is a type of neural network that employs the backpropagation algorithm. It shares a similar architecture with the single-layer perceptron, but is distinguished by the presence of one or more hidden layers. The backpropagation algorithm is composed of two distinct phases. The first

phase is known as the forward phase, during which the activations are transmitted data from the input layer to the output layer. The second phase is referred to as the backward phase, in which the difference between the predicted values and the actual values in the output layer is transmitted backward to update the weights and bias values. Subsequently, the outputs are calculated using a transformation function (Chen and Wang, 2020). The structure of the MLP is shown in Figure (5.9), which was generated using the draw.io software.

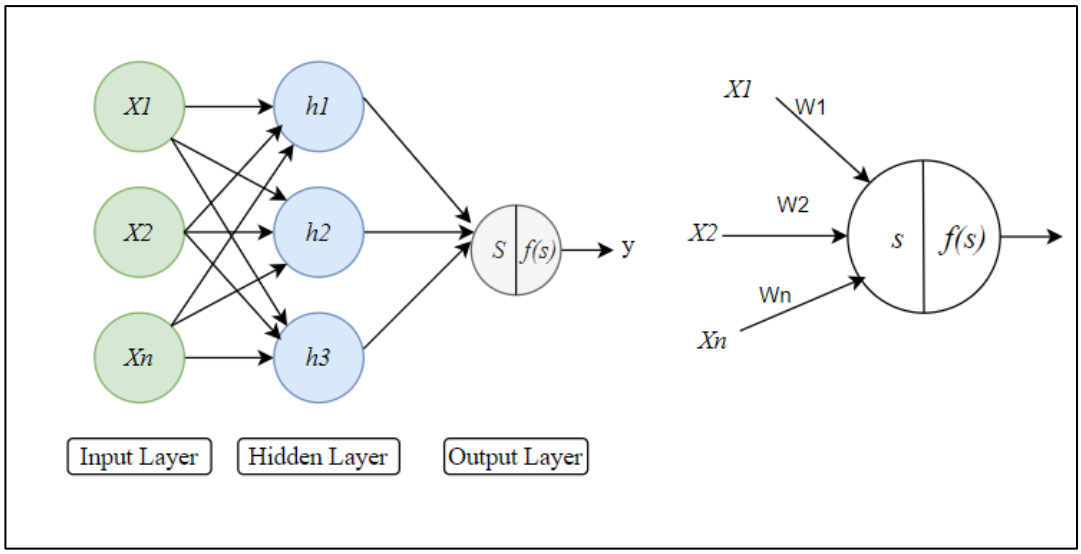


Figure (5.9): Architecture of the MLP network.

5.4 Applying Machine Learning: Train, Test, and Evaluate Models

The present study aims to create hybrid models for GPM by integrating supervised ML techniques and knowledge-driven GIS-based technologies to evaluate GWPZ of the Dammam confined aquifer whithin the study area. In most machine learning projects, the original dataset is split into training and testing data to build and evaluate the model's accuracy on new or unknown datasets. The training dataset is a subset of the original data used to train the model, while the testing dataset is used to check the model's accuracy. Typically, the training

dataset is larger than the testing dataset, and ratios of 80:20, 70:30, or 90:10 are commonly used to split the data (Dao et al., 2020; Pham et al., 2020; Moayedi et al., 2020). The accuracy of the model and its predictive ability are directly impacted by the data utilized for model training. Several researchers have suggested that the performance of a model can be enhanced by adjusting the ratio of training to testing data during the development and assessment of machine learning models. This study employed a data splitting ratio of 70:30, as recommended by Quang et al. (2021), due to the observed development in the predictive ability of machine learning models when this particular ratio is employed.

In this study, the supervised classification task was performed using five MLAs that represent different learning strategies. These strategies include Ensemble Boosted Trees (EBT)-logic-based learners; Naive Bayes (NB)-statistical learning algorithms; Support Vector Machines (SVM); Multi-Layer Perceptron (MLP)- Artificial Neural Networks (ANN); and *k*-Nearest Neighbours (kNN)- instance based learners. Despite their differences, all supervised MLAs essentially depend on the relevance of information included within input variables and previous knowledge of the problem, which is used to train and assess classifiers. The basic principle of supervised classification necessitates the availability of training data which includes labeled samples that provide information about the known characteristics of the target being inferred. The training of classification models is guided by the MLA structure and the observed data distributions, which are commonly optimized by minimizing the error function. The trained classification models are then used to predict classes based on similar input data used during training.

5.5 Supervised Classifier Implementation

The process of supervised classification using MLAs involves three fundamental stages subsequent to the acquisition of input data and target classes. These stages include data pre-processing, classifier training, and prediction evaluation, which must be properly planned and executed to produce effective classification models and accurately evaluate their predicted outcomes. Data pre-processing allows users to prepare given variables to include relevant data for the target purpose. Based on the imposed circumstances, one or more algorithm-specific model parameters must be selected during MLA training to optimize classifier performance. Methods such as cross-validation on a subset of labeled samples (training data) are often used in the parameter selection process. The capability of MLAs to categorize new samples is evaluated using testing data. Statistical measurements of classifier performance indicate how the future performance of classifications will be accurate. Figure (5.10) illustrates a generalized methodology, including the three main steps for supervised machine learning classification.

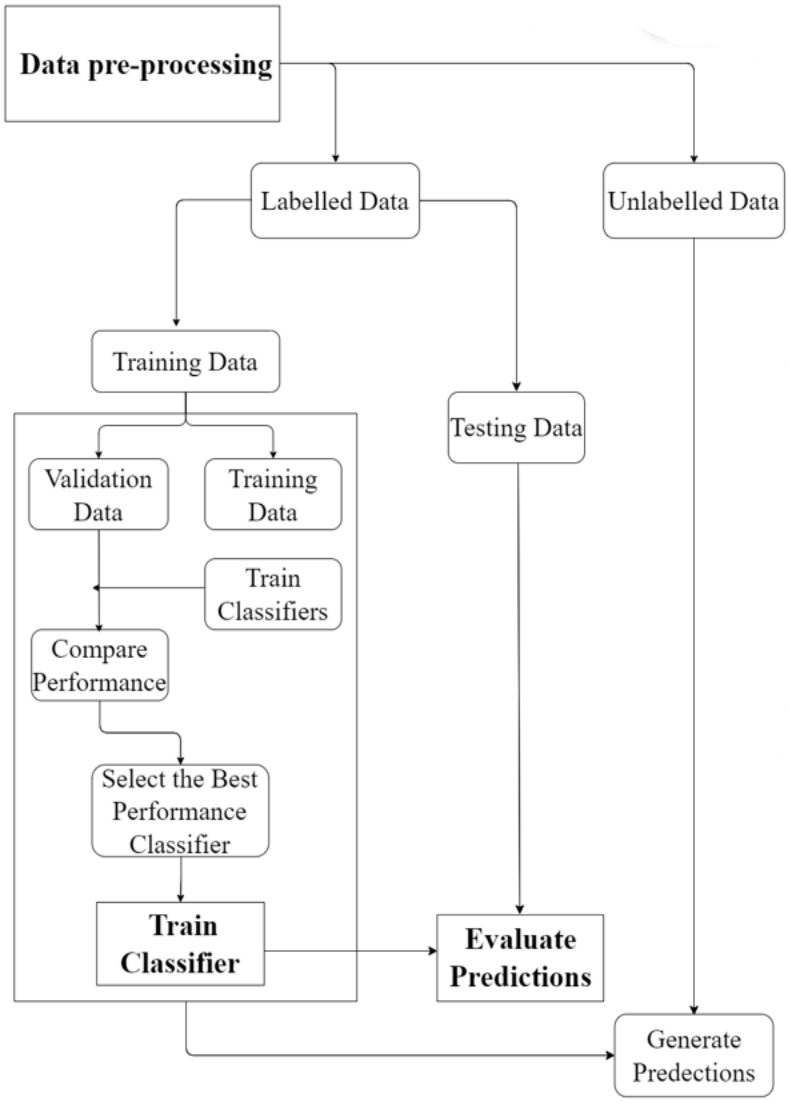


Figure (5.10): Generalized methodology for supervised ML classification

5.5.1 Data Pre-Processing

The first step in data pre-processing is preparing input variables and target classes for supervised classification in MLAs. Raw data usually contains inconsistencies that must be corrected before feeding the database to ML algorithms.The identified seven factors that influence groundwater potentialty, including hydrogeological aquifer unit types, transmissivity, lineament density, slope, soil type, land use and land cover, and drainage density required pre-

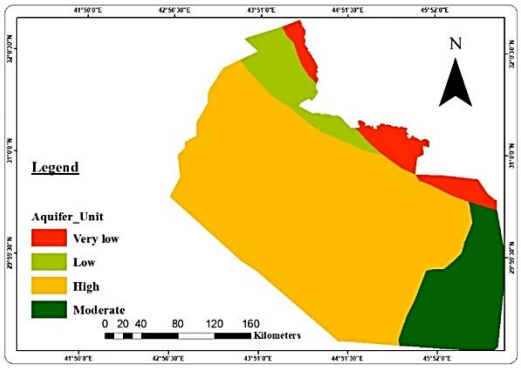
processing to be entered as input data to the MLAs. All maps were saved with BMP File (.bmp) with a dimension of (1009 pixels) for height and (683 pixels) for width and image resolution of (96 dpi) to ensure that the image data associated with the spatial coordinates of pixels of all seven thematic maps are accurately converted to a valued matrix. This conversion is based on the wells' position and their pumping rate. Matrices are generated to store data for variables on data layers, with each cell (pixel) containing a specific value. A pixel is the smallest unit that makes up a digital picture organized in a rectangular grid. The number of pixels in an image determines its resolution, which is often measured in (dpi).

The maps in this study employ a classification scale of 1, 2, 3, 4, and 5 correspond to the groundwater potentiality index of very low, low, moderate, high, and very high, respectively, as outlined by Kumar et al. (2014). Each map subclasses are assigned to a specific 24-bit true RGB color code. The RGB code defines a color by its constituent levels of Red, Green, and Blue, in which the three-part notation has any numerical value ranging from 0 to 255. The (R, G, B) code generates black when all three components are set to zero (0, 0, 0), and the brightest white is produced when all components are set to their maximum value of 255 in the form of (255, 255, 255). Table (5.2) displays the RGB code for each class of the seven thematic maps.

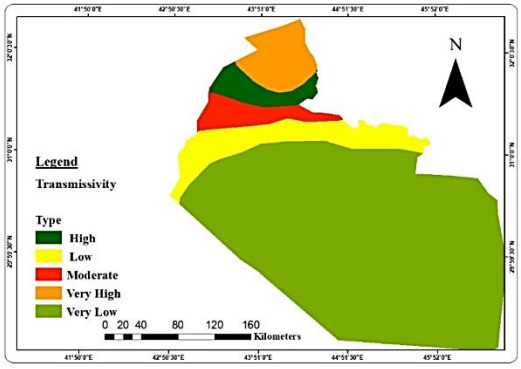
Table (5.2): RGB code for the influencing factors classes.

Factors	R	G	B	Legend	Color
Hydrological Aquifer Units					
Mesopotamia Plain Silt	255	38	0	very low	
Miocene Carbonate	164	196	0	low	
Karstified Palaeogene Carbonate	255	187	0	high	
Mio- Pliocene Sandstone	0	97	0	moderate	
Transmissivity (T) m^2/day					
24.332 - 32.414	0	97	0	high	
12.089 - 17.081	255	255	0	low	
17.081 - 24.332	255	34	0	moderate	
32.414 - 39.784	255	153	0	very high	
9.474 - 12.089	122	171	0	very low	
Lineament Density (LD) Km/Km^2					
0.079163905 - 0.114025074	0	97	0	high	
0.021788231 - 0.053744302	255	153	0	low	
0.053744302 - 0.079163905	255	34	0	moderate	
0.114025074 - 0.185199961	255	255	0	very high	
0 - 0.021788231	122	171	0	very low	
Slope (SLO) (Degree)					
0-0.1	0	97	0	very high	
0.1-0.2	255	34	0	high	
0.2-0.3	255	255	0	moderate	
0.3-0.5	122	171	0	low	
>0.5	255	153	0	very low	
Soil (SL) type-FAO Classification					
Clayey Loam	97	153	0	very low	
Loam	0	97	0	low	
Loamy sand (different mixed percentage)	255	38	0	moderate	
Sandy	255	217	0	very high	
Sandy loam	197	219	0	high	
Land use Land cover (LU/LC)					
Herbaceous Vegetation	233	255	190	high	
Herbs wetland	0	97	0	very high	
Bulid up	130	130	130	very low	
Bare/ Sparse Vegetation	255	235	190	low	
Cropland	76	230	0	moderate	
Water	0	77	168	very high	
Drainage Density (DD) Km/Km^2					
1.754258 - 4.65553	0	97	0	high	
7.354389 - 10.120718	255	153	0	low	
4.65553 - 7.354389	255	34	0	moderate	
0 - 1.754258	255	255	0	very high	
10.120718 - 17.205221	122	171	0	very low	

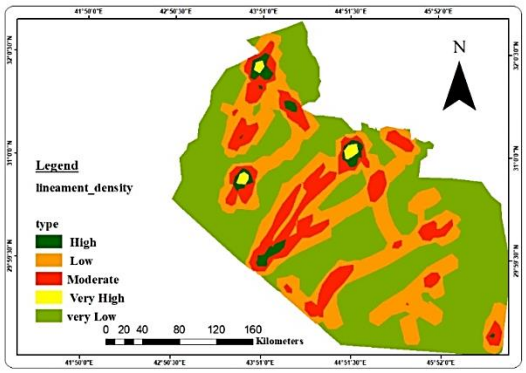
The conditioning factors of groundwater potentiality are shown in Figure (5.11) through data layers.



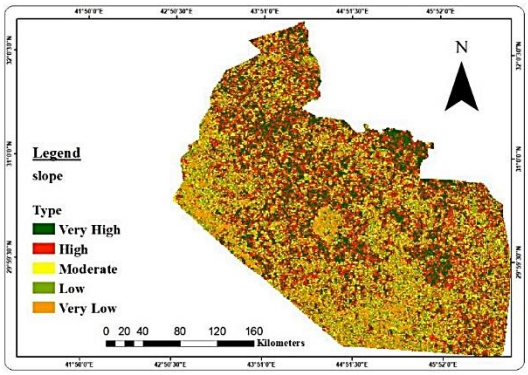
(a)



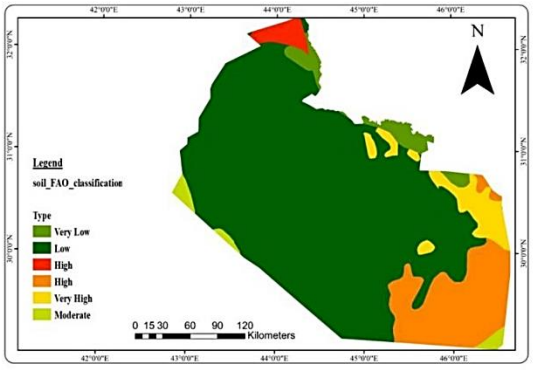
(b)



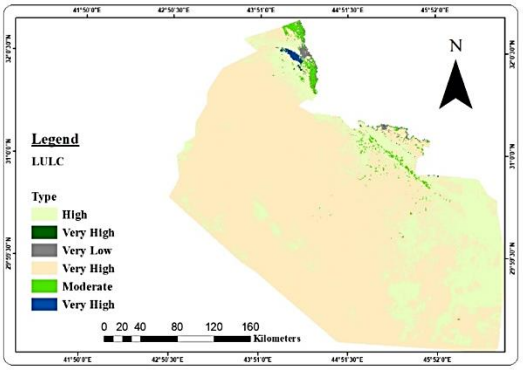
(c)



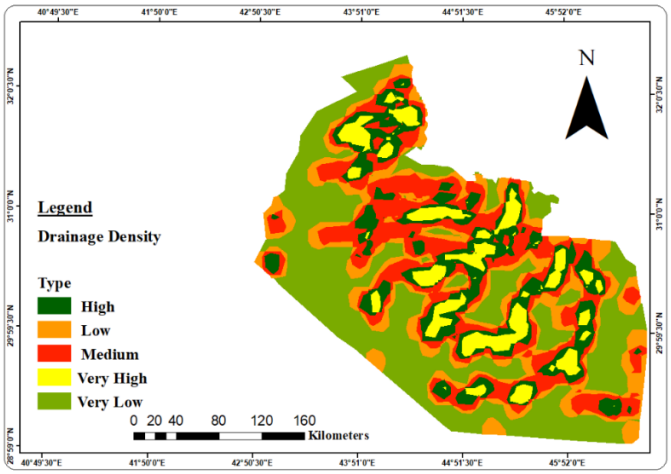
(d)



(e)



(f)



(g)

Figure (5.11): Data layers for groundwater potentiality conditioning factors: **(a)** Aquifer Units **(b)** Transmissivity **(c)** Lineament Density **(d)** Slope **(e)** Soil Classification **(f)** LULC **(g)** Drainage Density.

The process of generating features involves the utilization of the seven layers, which facilitate the conversion of image data into a valued matrix. This conversion is performed based on the position of 349 groundwater wells, and their respective pumping rates serve as targets for the supervised classification. Figure (5.12) illustrates the spatial distribution of these pumping wells.

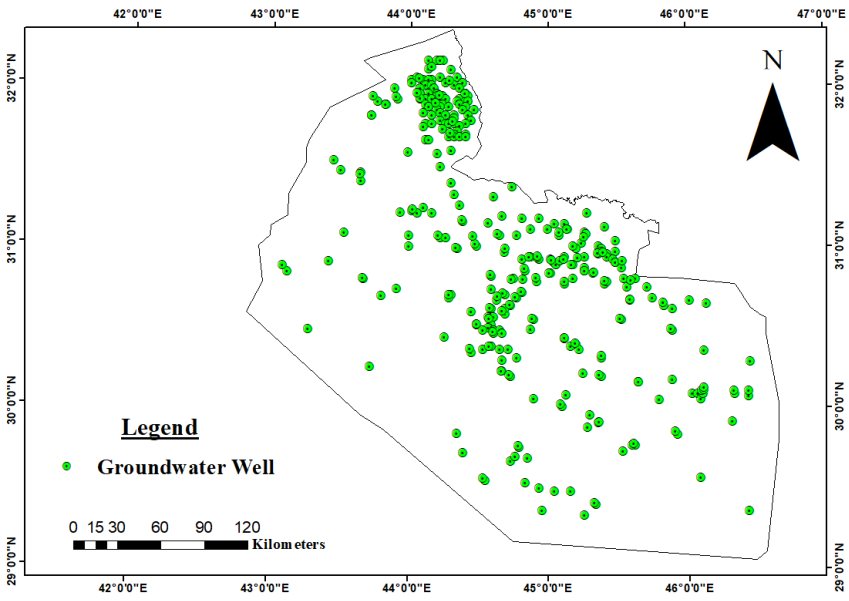


Figure (5.12): Wells locations for machine learning model.

The main difference between image and remote sensing data is that remote sensing (raster data) is linked to a geographical reference frame, a location on the earth's surface indicated by Latitude and Longitude. While the spatial coordinates of pixels indicate image data. Therefore, the subsequent phase of data pre-processing involves formulating an equation that can translate the latitude and longitude coordinates of a well into the corresponding pixel coordinates within an image, using polynomial curve fitting. This is achieved by utilizing the **polyfit** function in the MATLAB platform, which calculates the coefficients of a polynomial that can accurately fit a given dataset, expressed as follows:

$$x=polyfit(x_1, Y_1, n) \quad (5.2)$$

$$y=polyfit(y_1, X_1, n) \quad (5.3)$$

where: x_1 and y_1: are the longitude and latitude of the UTM coordinates of the wells position, respectively. X_1, Y_1: indicate the positions of the wells in the image's coordinate system, specifically denoting the row and column numbers of the image pixels, and n represent the degree of the polynomial to be fitted.

Then, the **polyval** function is employed to evaluate the polynomial at additional locations in (x), which is defined as:

$$Y_1\_fit=polyval(x, x_1) \quad (5.4)$$

$$X_1\_fit=polyval(y, y_1) \quad (5.5)$$

Accordingly, the developed functions are:

$$x= -0.0017 \times y_1 + 6.127e+03 \quad (5.2\text{-a})$$

$$y= 0.0017 \times x_1 - 3.025e+02 \quad (5.3\text{-a})$$

Where; x and y are the coordinates of image data representing the row and column of the image matrix.

Therefore, these equations are used to create the image pixel coordinates and then create the database of the collected wells that contain the features of thematic maps and the classes of the flow rate.

There were five classes of image targets identified based on the available data for wells' discharge rates: (very poor) for pumping rates equal to or below 3 ℓps, (poor) for pumping rates between 4 and 6 ℓps, (moderate) for pumping rates between 7 and 9 ℓps, (good) for pumping rates between 10 and 14 ℓps, and (very good) for pumping rates above 14 ℓps. The groundwater discharge classes were defined using an (if, else if, else) statement, a common MATLAB structure for conditional assignment. The generated database matrix has a dimension of (349×8) was employed to train hybrid models of classification learners.

The proposed hybrid model utilizes the concept of evaluating, analyzing, and combining multiple factors. This is achieved through the implementation of MLAs, which employ inductive analysis techniques to automatically identify patterns and relationships within the data. These algorithms are then applied to additional comparable data to provide predictions for data-driven classification problems. This approach enables the creation of a robust model that can effectively classify data based on its attributes and characteristics, thus enhancing the accuracy and reliability of data-driven decision-making processes. Furthermore, it can be easily updated with new information so that it may be used to examine additional areas and provide GPMs.

5.5.2 Training of Classifier

To effectively train and assess available data, it is important to divide the database into independent groups. Training data aim to identify patterns and correlations within the data and to optimize parameters and classification models, while testing data is used to objectively evaluate the predictive capabilities of

trained classifiers. The recommended splitting ratio for datasets is 70:30, where 70% of the database is used for training and 30% is used for testing to ensure the accuracy of the model.

The final goal is to train a classification model that fits the inputted training data and shows reliable results when applied to unfamiliar (new) data. An essential step in the MLA classifier training process is setting appropriate parameters for the task based on the available data, which is a significant step. A cross-validation method is often used to choose the appropriate parameters since this method is a simple and direct objective data-driven approach for selecting relevant algorithm parameters and evaluating the predictive accuracy of trained classification models (Kuhn and Johnson, 2013). In this study, the training data were subdivided into two subsets utilizing the cross-validation technique. One of these subsets is known as training data, which was used to modify the model parameters. The other subset, known as validation data, was used to measure the performance accuracy of trained classifiers.

5.5.3 Evaluation Criteria of Predictions

To ensure accurate prediction of the trained classifier, the testing data must be independent of the training data. This study uses several metrics, such as True Positive Rates (TPR), False Negative Rates (FNR), Positive Predictive Values (PPV), False Discovery Rates (FDR), the area under the receiver operating characteristics (ROC) curve, and classifier accuracy to assess the classifier's performance in the classification learner. The classifier's performance per true class was assessed using TPR and FNR. TPR represents the proportion of correctly classified observations per true class, while FNR represents the proportion of incorrectly classified observations per true, in which expressed by Hicks et al. (2022):

$$\text{True Positive Rates (TPR)} = \frac{TP}{TP+FN} \tag{5.6}$$

$$\text{False Negative Rates (FNR)} = \frac{FN}{TP + FN} \tag{5.7}$$

TP refers to true positives, which represents the number of positive samples that are accurately classified. FN, on the other hand, stands for false negatives and indicates the number of samples that are inaccurately classified as negative. FP denotes false positives and pertains to the number of samples that are inaccurately classified as positive. Lastly, TN stands for the observations that are negative and predected to be negative (accuratly classified as negative), as presented in Figure (5.13).

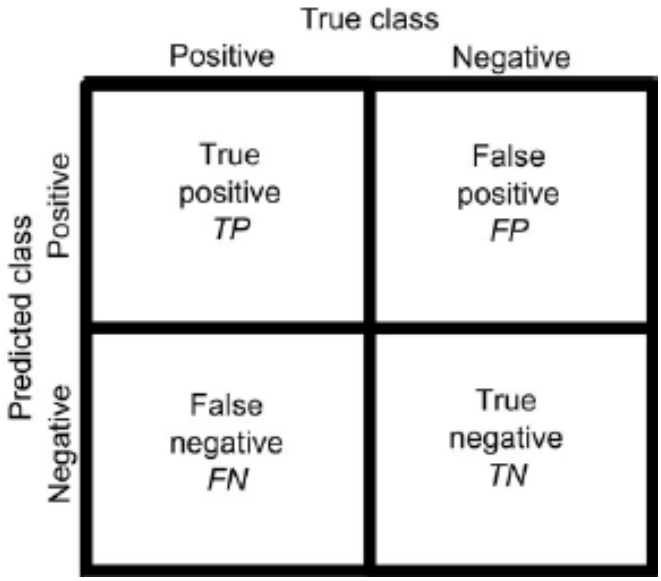


Figure (5.13): Confusion matrix (Vihinen, 2012).

Figure (5.14) presents the rates in which the performance of the classifier for each true class was evaluated using TPR and FNR for multiclass classification.

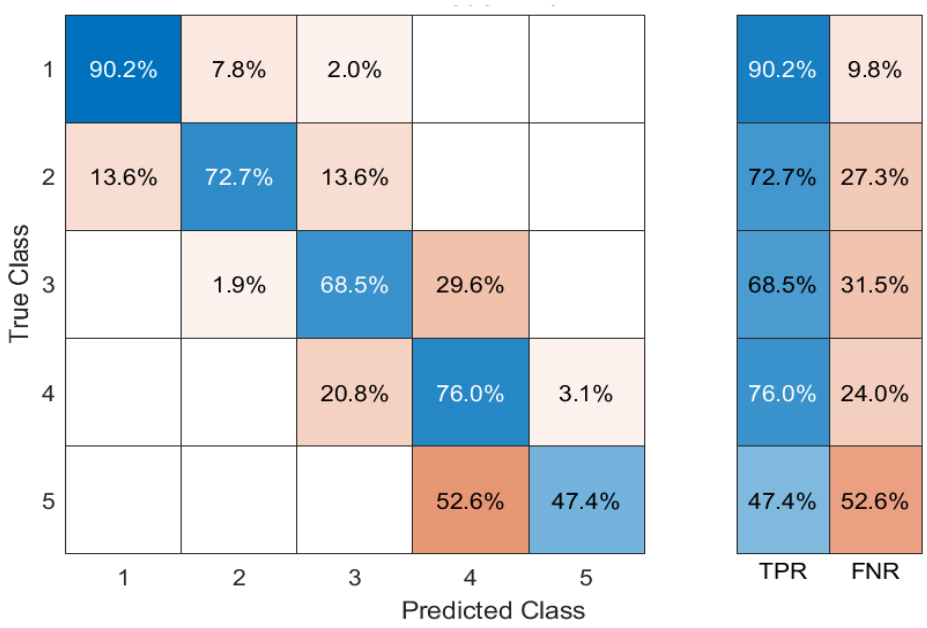


Figure (5.14): Classifier performance per true class using TPR and FNR.

A confusion matrix is used to evaluate categorical predictions generated by MLA-supervised classifiers. It presents a structured and organized representation of the classification model's performance. The dimensions of the confusion matrix are equal to the number of classes present in the given dataset. The numerical values present in every individual cell of the matrix denote the frequency of predicted classifications for samples that have been allocated to a specific class. This plot helps to identify areas where the classifier has performed poorly and can be viewed from the plots section of the classification learner tab after training the model. Figure (5.15) shows the confusion matrix of a five-class machine-learning model, indicating the number of correctly classified classes and the number of misclassified classes.

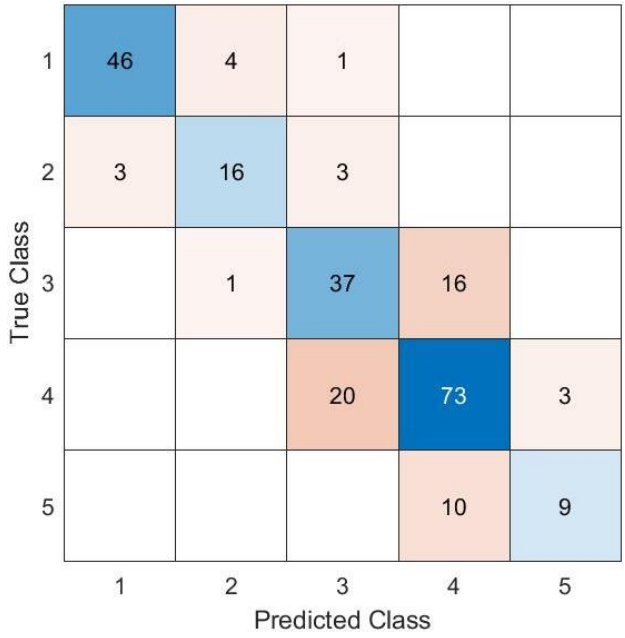


Figure (5.15): The confusion matrix of five-class ML model.

Examining the false discovery rates in the classification problem can be carried out by plotting results presented per predicted class if false positives are critical. Results per predicted class can be viewed by selecting PPV and FDR options. The positive predictive value, or PPV, indicates the percentage of observations properly classified as positive for each predicted class to total positively classified samples, including TP and FP. While the false discovery rate, or FDR, is the fraction of observations that were not properly categorized for each predicted class and all samples identified as positive. In which the PPV and FDR are generally expressed as follows (Hicks et al., 2022):

$$\text{Positive Predictive Values (PPV)} = \frac{TP}{TP+FP} \tag{5.8}$$

$$\text{False Discovery Rates (FDR)} = \frac{FP}{TP+FP} \tag{5.9}$$

Figure (5.16) displays classifier performance per predicted class using PPV and FDR options. The positive predictive values are shown in blue for the points

in each class that were properly predicted, and the false discovery rates are displayed in orange for each class that was not correctly predicted.

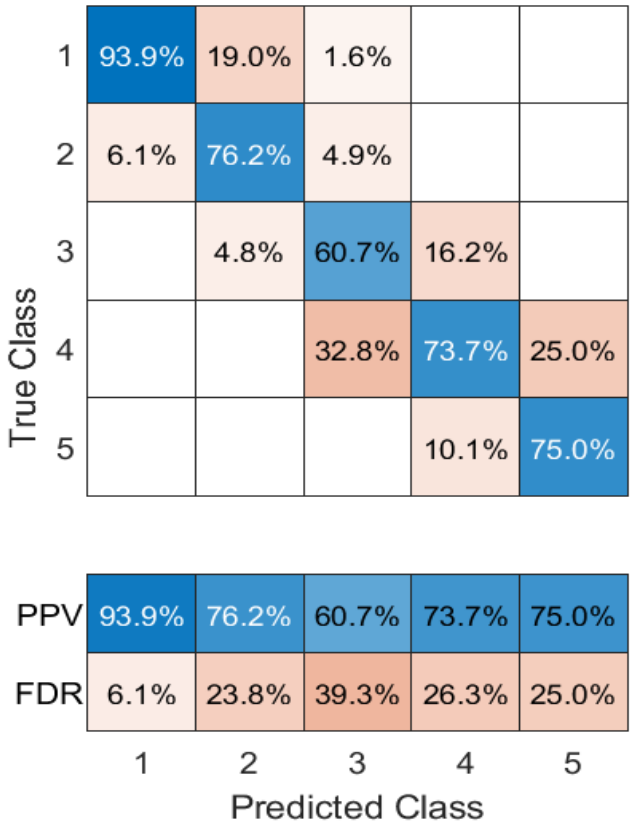


Figure (5.16): The classifier performance per predicted class using the Positive Predictive Values (PPV) and False Discovery Rates (FDR).

Another performance evaluation of models by using a receiver operating characteristic (ROC) curve can be performed. A standard statistical tool that is used to evaluate the performance of the models by plotting (TPR) and (FPR) on the y and x axes, respectively. This evaluation aimed to determine how well the models performed. The AUC is measured from (0, 0) to (1,1) and represents the entire two-dimensional area under the ROC curve, as shown in Figure (5.17).

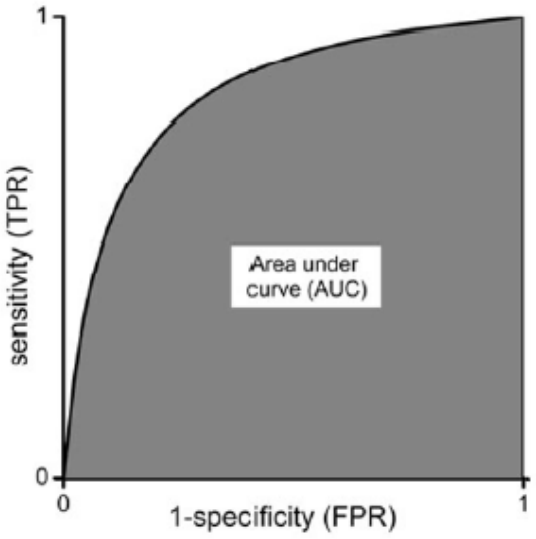


Figure (5.17): Area under the ROC curve.

The AUC values range from 0 to 1, with 0 indicating 100% incorrect predictions and 1 representing 100% accurate predictions. Figure (5.18) shows a plot using sensitivity (TPR) on the y-axis and 1-specificity (FPR) on the x-axis, illustrating some observations associated with the ROC curves. The sensitivity and 1-specificity, as described by Kuruvilla and Gunavathi (2014), can be calculated as follows:

$$\text{Sensitivity}=\frac{TP}{TP+FN} \tag{5.10}$$

$$\text{1-specificity}=\frac{FP}{FP+TN} \tag{5.11}$$

The value of FPR and TPR falls within the range [0, 1], and the green circle in the top left corner of Figure (5.18) indicates the best classification that a classifier can achieve. An (FPR, TPR) of (0.0, 1.0) indicates that the classifier can correctly classify all data without making false classifications. The area under the ROC curve (AUC) is under the red curve, as demonstrated by Amin and Yan (2011).

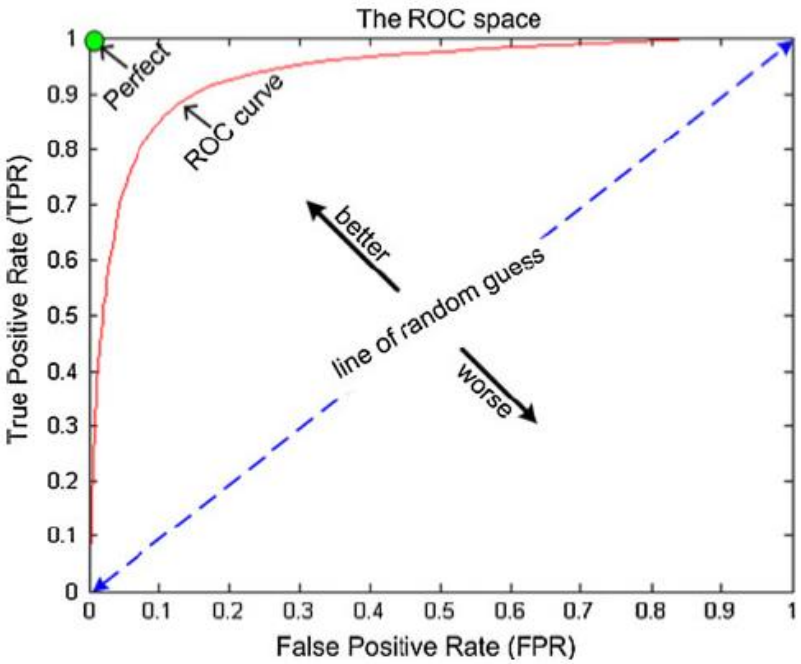


Figure (5.18): ROC curve (Amin and Yan, 2011).

In the ROC curve, a perfect classification is represented by a right angle in the upper left corner, wherein no points have been incorrectly classified. In contrast, a poor classifier is depicted by a line at a 45-degree angle, which is no better than random. The AUC metric serves as an indicator of the classifier's overall performance, as illustrated in Figure (5.19). A larger AUC value suggests a better performance of the classifier. By default, the ROC metrics in MATLAB utilize the FPR and the TPR to generate ROC curve and subsequently determine the AUC.

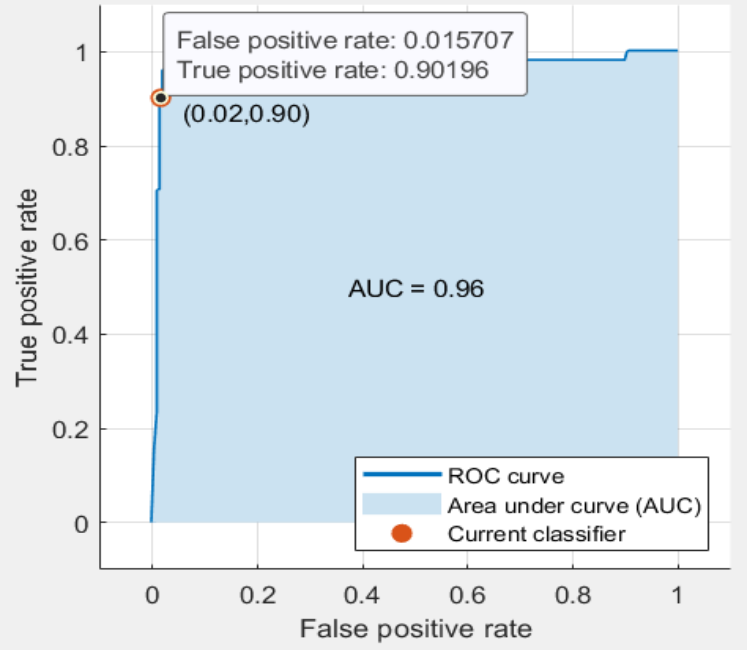


Figure (5.19): Receiver operating characteristic (ROC) curve and area under curve AUC for a classifier.

In the final step of model development, validation performance is calculated using model accuracy, with 30% of the dataset serving as testing data. The Accuracy (ACC) is a statistical metric that evaluates the ability of a classifier to correctly identify different classes and predict unlabeled data. Its calculation involves the use of equation (5.12), which determines the percentage of accurately classified samples in relation to the total number of samples in the evaluation dataset, as stated by Hicks et al. (2022).

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (5.12)$$

Chapter Six

Implementation of Machine Learning Algorithms for Groundwater Potential Zones Mapping

6.1 Introduction

The application of machine-learning algorithms for the purpose of mapping groundwater potential zones is a newly developed area of research that holds significant potential in the field of water resources management. Geological maps are frequently created using indirect observations to represent geological phenomena and features in a 2D spatial reference frame. They are crucial for many applications that rely on multivariate geospatial data, such as groundwater wells, remote sensing data, and published maps (Maiti and Tiwari, 2010). These data are utilized to generate thematic maps, produce a database, and train supervised machine-learning models

The utilization of MLAs has the capability to offer effective solutions for groundwater mapping and investigation in areas where acquiring field observations is difficult, time-consuming, and expensive. This study employed supervised machine learning classification, and the methodology for this approach is presented in Figure (6.1).

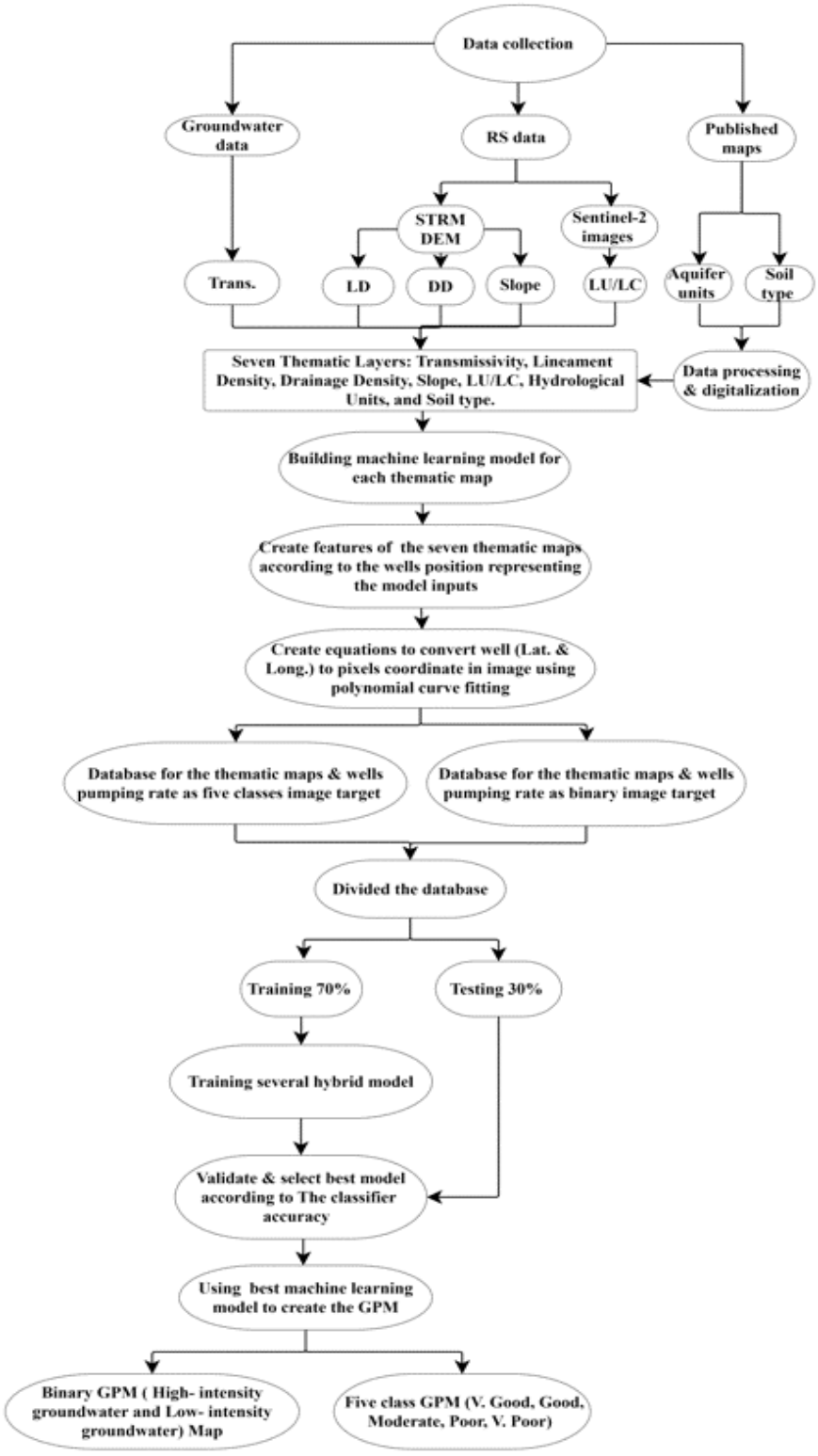


Figure (6.1): Methodology of the present work for classifying through supervised machine learning.

6.2 Splitting of Data

This study utilized supervised classification, which necessitates using labelled samples containing known information regarding the inference targets. Accordingly, the coordinates of 349 groundwater wells and their corresponding pumping rates were provided by the General Committee for Groundwater (GCGW)/ Ministry of Water Resources of Iraq. The study employed a splitting ratio of 70:30, wherein the locations of the pumping wells were randomly allocated into two groups.

The training data consisted of 70% of the groundwater wells with their coordinates transformed into pixel coordinates using polynomial curve fitting equations. The known pumping rates of the training data were used as the machine learning model target, and threshold values were applied to generate the categories.

The remaining 30% of collected wells were used as testing data to examine the model performance and to measure the capability of MLAs to categorize new samples that are unlabeled classes. MLAs are provided with the required information linking the location and pumping rate of the wells with the groundwater classes. Achieving an accurate machine-learning model can be improved by combining spatial values with a random distribution of samples across the study area during the phase of data splitting. Figure (6.2) illustrates the distribution of wells for the 70:30 splitting ratio used for training and testing data.

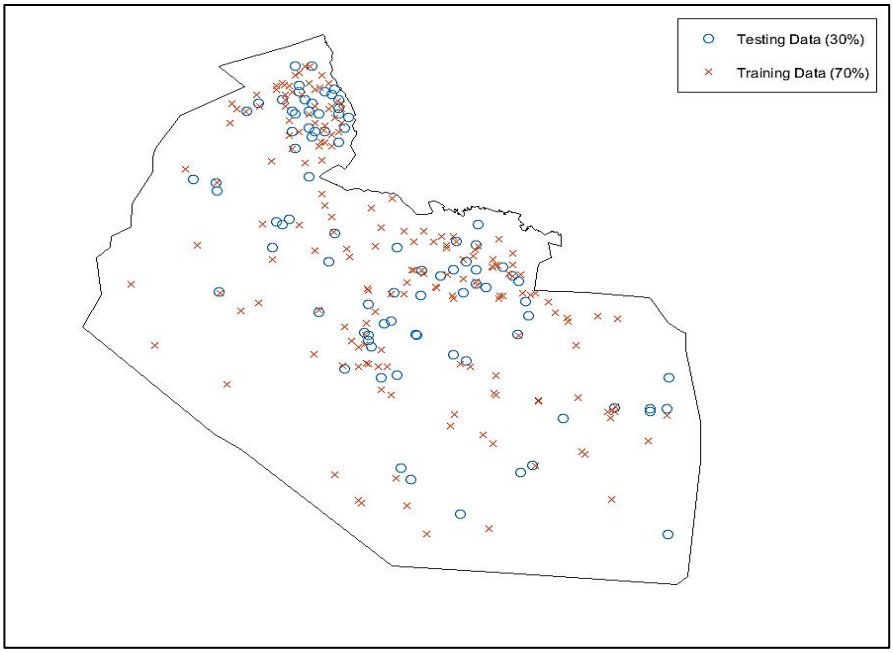


Figure (6.2): Training/testing data split for Five-classes mapping.

To ensure that the dataset was divided adequately, the process of splitting was carried out based on the target of the ML model, which was the pumping rate of wells; consequently, five matrices were created (A1-A5) containing the data features of the seven layers for groundwater potentiality conditioning factors, in addition to the target column. The 70:30 ratio was applied to the generated matrices, in which a total of (242) wells were chosen randomly for the training process, while (107) wells were selected for validation, as shown in Table (6.1).

Table (6.1): Dividing of Data

Generated matrix based on the target class	**No. of wells in matrix based on the target class**	**No. of training wells (70%)**	**No. of testing wells (30%)**
A1	73	51	22
A2	32	22	10
A3	78	54	24
A4	138	96	42
A5	28	19	9
Total	349	242	107

6.3 Application of Classification Learners

To create and train the supervised machine-learning model, a training matrix of (242×8) was applied. The ML and deep learning section in MATLAB R2020b software was employed to perform a supervised classification learner on the data. This was carried out by choosing the dataset and specifying a validation scheme. In Cross-Validation**,** the number of folds (or divisions) is selected to partition the data set using the slider control. The default option is 5-fold cross-validation, which protects against bias. When using the *k* folds cross-validation in which *k* refers to the number of groups the dataset is to be split into, the program will randomly partition the data into the selected folds for each unique class. The first fold is treated as a validation set known as in-fold data, while the remaining groups are treated as a training dataset (out-of-fold observations). The app will fit the model on the training set and evaluate it on the testing set; finally, the average error is calculated. Figure (6.3) illustrates the Cross-Validation process.

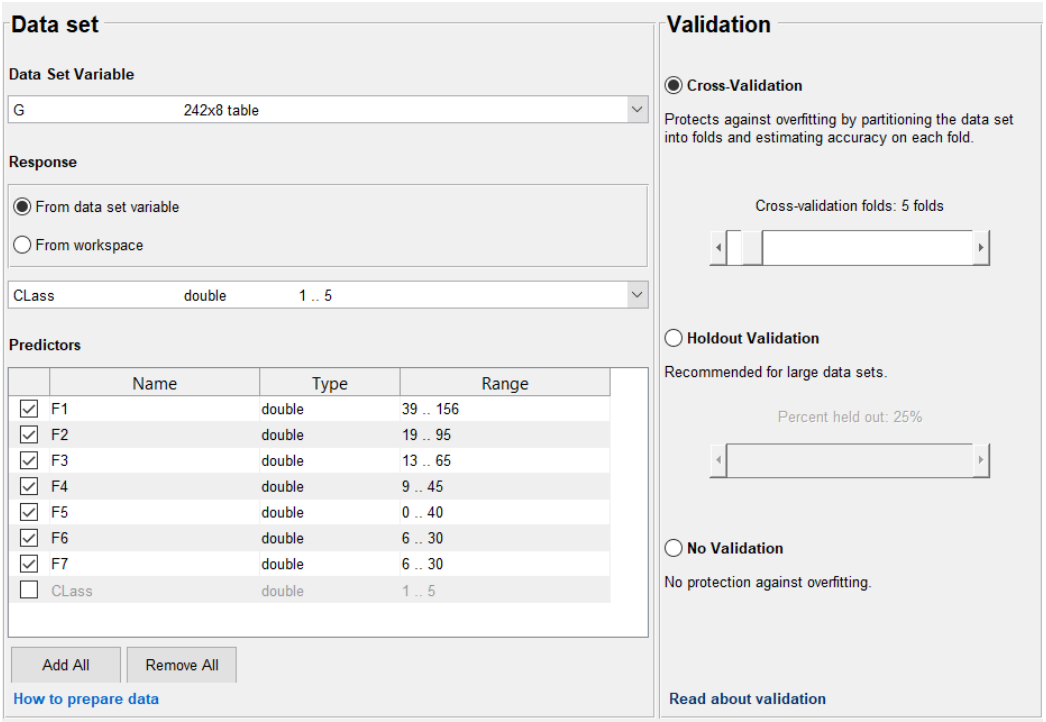


Figure (6.3): Cross-Validation process.

In applied ML, cross-validation is commonly employed for estimating the predictive performance of a model on data not utilized during the model's training. It is a popular technique as it offers a more practical evaluation and reduces bias in the model's performance, as explained by Kuhn and Johnson (2013).

6.4 Evaluating the Prediction of the Models

A compression analysis is carried out for the predictive ability of the developed five supervised MLAs. The performance of predictive models is evaluated using the TPR, FNR, PPV, FDR, the area AUC of the receiver operating characteristics (ROC), and accuracy of the classifier to to identify the most accurate model that could be employed to create the final GPM.

6.4.1 Evaluating the Logic-Based Learners-Ensemble Boosted Trees

The methodology involves the generation of several datasets from the primary dataset, followed by the application of multiple classifier learners to create several hypotheses and combine them for better predictive performance compared to a single model. This study used 30 decision tree learners. To evaluate the performance of the selected classifier in each class and identify areas where it performed poorly, a confusion matrix plot is shown in Figure (6.4).

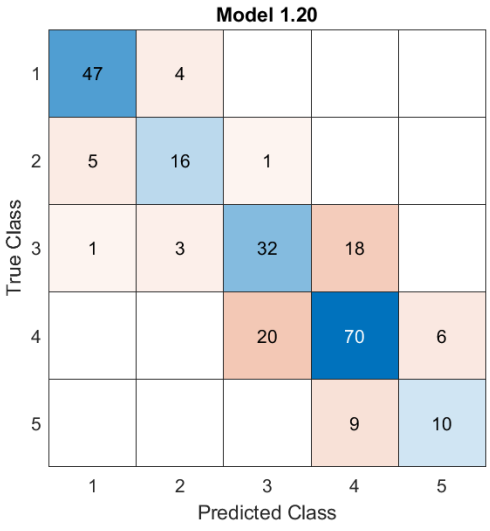


Figure (6.4): Confusion matrix of Ensemble Boosted Trees model.

The columns represent the predicted class, while the rows indicate the true (actual) class. In the confusion matrix figure, the cells along the diagonal indicate where the true class and predicted class are matched. When the diagonal cells are blue, it indicates that the classifier correctly identified the data in this true class. On the other hand, classes with orange color are those that were misclassified by the classifier. The confusion matrix in Figure (6.4) indicates that for wells classified as class-1 (i.e., very poor) with pumping rates below 3 ℓps, there were 47 correct identifications and 4 misclassifications. The True Positive Rates (TPR) and False Negative Rates (FNR) depicted in Figure (6.5) were computed using the correct and misclassified cells.

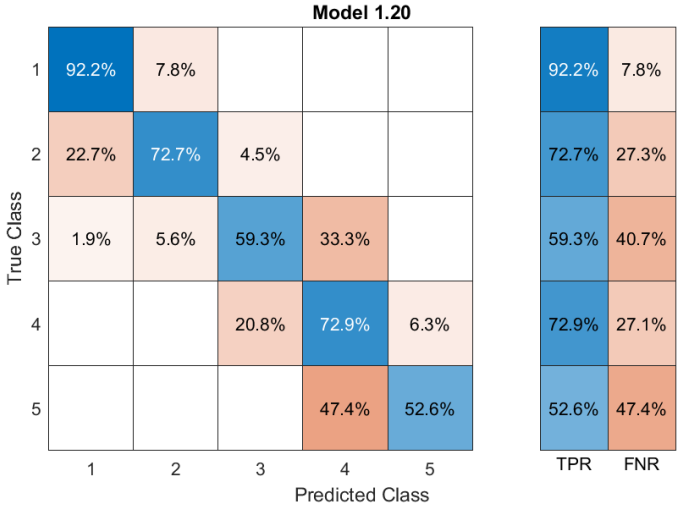


Figure (6.5): Performance of EBT per true class using TPR and FNR.

Figure (6.5) presents the TPR and FNR for each true class by a percentage summary provided in the last two columns on the right. The performance of the Ensemble Boosted Trees (EBT) model was evaluated in terms of its true positive rate (TPR) and false negative rate (FNR) for wells classified as class-1. The TPR was found to be 92.2%, indicating that 47 out of 51 observations were correctly classified. On the other hand, the FNR was 7.8%, indicating that 4 out of 51 observations were incorrectly classified. Likewise, the same methodology was

employed to compute the TPR and FNR for the remaining categories. Figure (6.6) displays the outcome per predicted class instead of true class with the aim of examining the false discovery rates (FDR) that refers to the proportion of observations that were inaccurately classified for each predicted class.

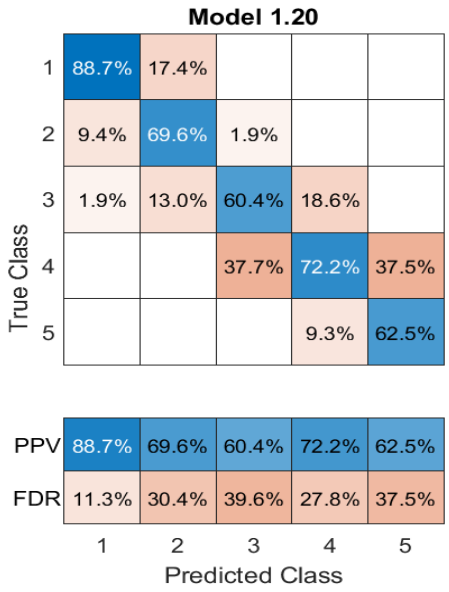


Figure (6.6): Performance of EBT per predicted class using PPV and FDR.

The number of true positive (TP) observations is 47, while the number of false positive (FP) observations is 6 wells per predicted class, resulting in a total of 53 observations. The calculations report the PPV and FDR for wells categorized as class-1, achieving values of 88.7% and 11.3%, respectively. The training data of the EBT model exhibited an overall accuracy of 72.3%. This value was obtained by computing the percentage of accurately classified classes in the diagonal cells of the confusion matrix, divided by the total number of observations (242) in the training dataset. Once a model has been trained, the receiver operating characteristic (ROC) curve can be plotted. The ROC curve shows the true positive rate versus the false positive rate for the currently selected trained classifier. The number that represents the area under the curve is a measurement of the overall quality of the classifier. The larger area under curve values indicates better

classifier performance. The EBT model's AUC value is 0.98 with (FPR,TPR) of (0.03,0.92), as illustrated in Figure (6.7), which displays the ROC curve and AUC for the current classifier.

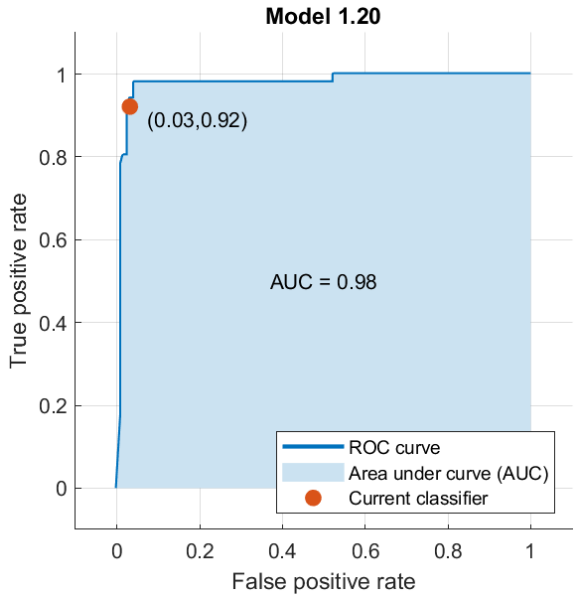


Figure (6.7): ROC curve for the Ensemble Boosted Trees model.

Ultimately, the model is assessed using 30% of the remaining data that was not utilized in model construction. The evaluation of the EBT model's accuracy is derived from the confusion matrix of the testing dataset, yielding a value of 89.7%, As illustrated in the confusion matrix presented in Figure (6.8). The calculation involves determining the proportion of accurate outcomes in the dataset's diagonal cells relative to the overall quantity of testing wells, which is 107.

22	0	0	0	0
2	8	0	0	0
0	2	20	2	0
0	0	1	41	0
0	0	0	4	5

True Class

Predicted Class

Figure (6.8): Confusion matrix of testing data for EBT classifier.

6.4.2 Evaluating the Support Vector Machine (SVM) Algorithms

The Support Vector Machine (SVM) is capable of identifying nonlinear decision boundaries through the utilization of kernel functions. There exist two fundamental stages. The initial stage of the process involves the transformation of the original input data into a higher dimensional space utilizing a kernel function such as the polynomial, Gaussian, radial basis function, exponential radial function, multi-layer perceptron, etc. The subsequent stage involves the search for a hyperplane that can effectively achieve linear separation of data in the transformed space, as illustrated in Figure (6.9).

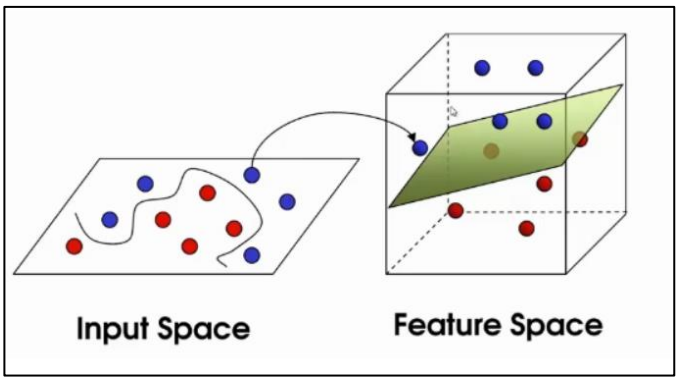


Figure (6.9): Nonlinear SVM (Melgani and Bruzzone, 2004).

This study used a medium Gaussian Support Vector Machine (SVM) to train the data. The kernel function employed for transforming the data into a higher dimension and generating the Gaussian surface facilitated linear separability of the data. The assessment of the SVM classification learning model was conducted through the application of the confusion matrix and statistical metrics, including True Positive Rates (TPR), False Negative Rates (FNR), Positive Predictive Values (PPV), False Discovery Rates (FDR), area under ROC curve (AUC) and the accuracy (ACC) of model for training and testing data. The results of this evaluation are displayed in Figure (6.10). To achieve optimal performance, it is desirable for both true positive rate (TPR) and positive

predictive value (PPV) to be high, while the false negative rate (FNR) and false discovery rate (FDR) should be low.

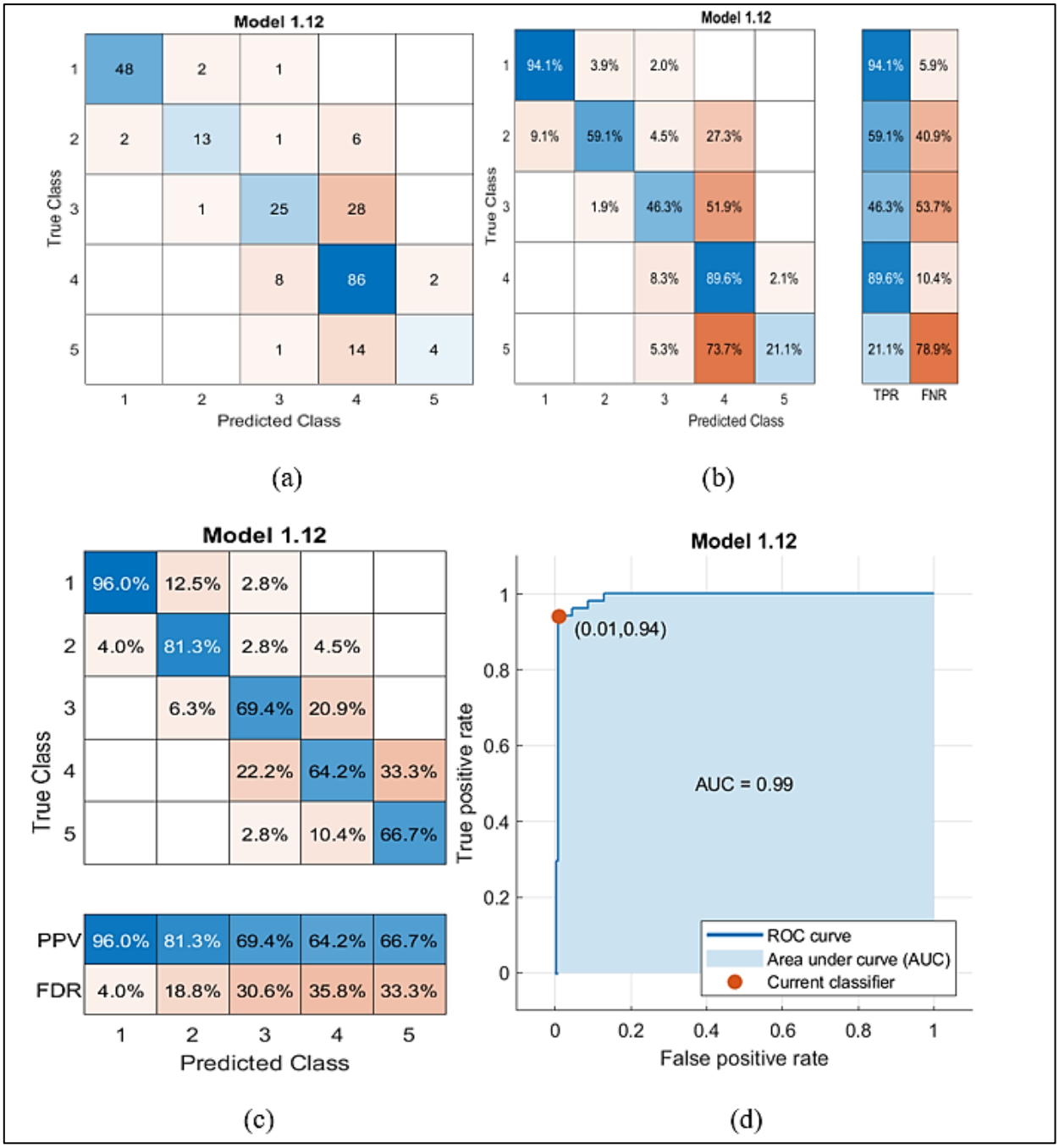


Figure (6.10): Performance of the SVM classification model:

(a) Confusion matrix in the form of observations number.

(b) TPR and FNR **(c)** PPV FDR **(d)** Area under ROC curve.

The results of the SVM classifier, as depicted in Figure (6.10), demonstrate that the True Positive Rate (TPR) for all identified classes is higher, with the exception of class-3 and class-5, which exhibit False Negative Rates (FNR) of 53.7% and 78.9%, respectively. These values are greater than the TPR of the same classes, indicating that the classifier's accuracy for data classification is unsatisfactory. This inadequacy can be attributed to the fact that the accuracy of the model's training data stands at 72.7%. Conversely, PPV exhibited greater values compared to FDR, signifying satisfactory classification for each predicted class with the area under the ROC curve of 0.99.

On the other hand, the accuracy of the SVM classifier when applied to testing data is 77.6%, as illustrated in Figure (6.11).

True Class

22	0	0	0	0
1	9	0	0	0
0	1	10	13	0
0	0	0	42	0
0	0	0	9	0

Predicted Class

Figure (6.11): Confusion matrix of testing data for SVM classifier.

6.4.3 Evaluating the Naive Bayes (NB) -Statistical Learning Algorithms

The performance of the Naive Bayes (NB) classifier was evaluated under the assumption that the dataset follows the normal distribution, allowing probabilities to be calculated based on this distribution., as presented in Figure (6.12).

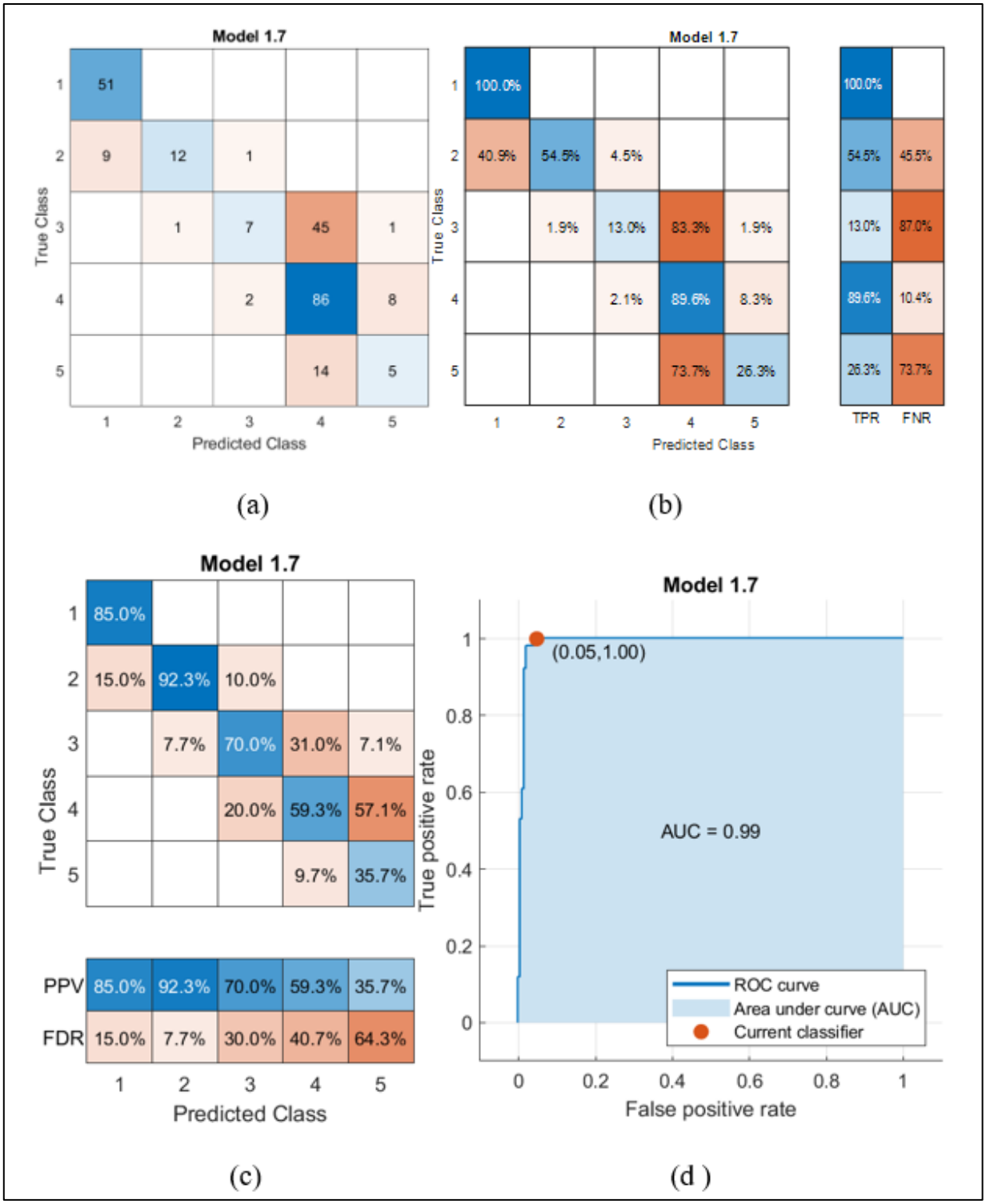


Figure (6.12): Performance of the Naive Bayes classification model

(a) Confusion matrix in the form of observations number.

(b) TPR and FNR **(c)** PPV FDR **(d)** Area under ROC curve.

In Figure 6.12 (b), the TPR for class-1 is 100%, signifying an ideal outcome with no reports of misclassified points. However, the TPR for classes 3 and 5 is comparatively lower than the FNR. Furthermore, it can be observed that the false discovery rate (FDR) associated with class-5 exceeds the positive predictive value (PPV), as illustrated in Figure 6.12 (c). Consequently, employing an alternative classifier that yields better performance is imperative, given that the overall accuracy of the Naive Bayes classifier for the training data stands at 66.5%. The AUC is 0.99, and the accuracy of the current classifier for testing data is 75.7%, as presented in Figure (6.13)

True Class					
	22	0	0	0	0
	2	8	0	0	0
	0	1	7	16	0
	0	0	0	42	0
	0	0	0	7	2

Predicted Class

Figure (6.13): Confusion matrix of testing data for Naive Bayes classifier.

6.4.4 Evaluating the of Multi-Layer Perceptron (MLP) - ANN

The multi-layer perceptron, denoted as MLP, is an improvement to the feed-forward neural network. As illustrated in Figure (6.14), the structure consists of three layers, namely the input layer, hidden layer, and output layer.

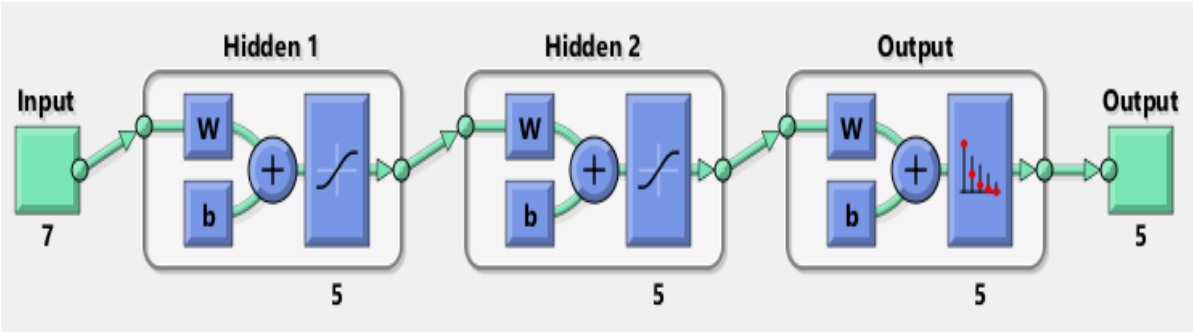


Figure (6.14): Scheme of the present study five-classes MLP neural network.

A multilayer perceptron (MLP) neural network was utilized to classify five classes. The network consisted of an input layer with seven features, representing the affecting factors of the groundwater potential map. Two hidden layers, each with five neurons, were also included in the network, along with an output layer of five neurons. The MLP neural network was trained for this purpose. The output layer is responsible for executing the necessary function, such as making predictions and conducting classifications. The output neurons were utilized to indicate the aimed target of the network. In MLP, the data is propagated in the forward direction from the input to the output layer, similar to a feed-forward neural network. The backpropagation learning algorithm was utilized to train the neurons of the output layer in the MLP. Equation (6.1) defines the hyperbolic tangent function as the activation function employed in the hidden layers. The function generates outputs within the range of [-1, +1]. Furthermore, the function generates an output for each value of (x).

$$Tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{6.1}$$

The output layer employed the softmax activation function. The transfer function is often used in the output layer of classified artificial neural networks (ANN) due to its ability to transform input data into probability. The Softmax function is employed to transform an input value into a set of values that conform to a probability distribution, where the sum of all values equals 1. It is preferred in the multiclass classification of the neural network model because the output values are in the range [0,1], unlike the binary classification, which can only accept the value 0 or 1.

The softmax function can transform small or negative inputs into a small probability, while large inputs are transformed into a large probability. However, it is important to note that the resulting probability will always be

confined within the range of 0 and 1. The mathematical representation of the softmax transfer function is included in equation (6.2).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{6.2}$$

Where $\sigma(z)_i$: is the softmax transfer function, z_i: the input elements to the softmax function, and they can take any real value, positive, zero, or negative, e^{z_i}: the standard exponential is applied to each input element that provides a positive number over 0, which is little if the input is negative and high if positive. But the resultant values are still not in the range (0, 1), and probability must be provided, $\sum_{j=1}^{k} e^{z_j}$: the normalization term. It ensures that all the function output values will sum to 1 and each be in the range (0, 1), thus forming a valid probability distribution. The present study reports a classification accuracy of 91.7% for the MLP-ANN model after conducting (100000) epoch/ iterations. The training confusion matrix of the MLP neural network is shown in Figure (6.15).

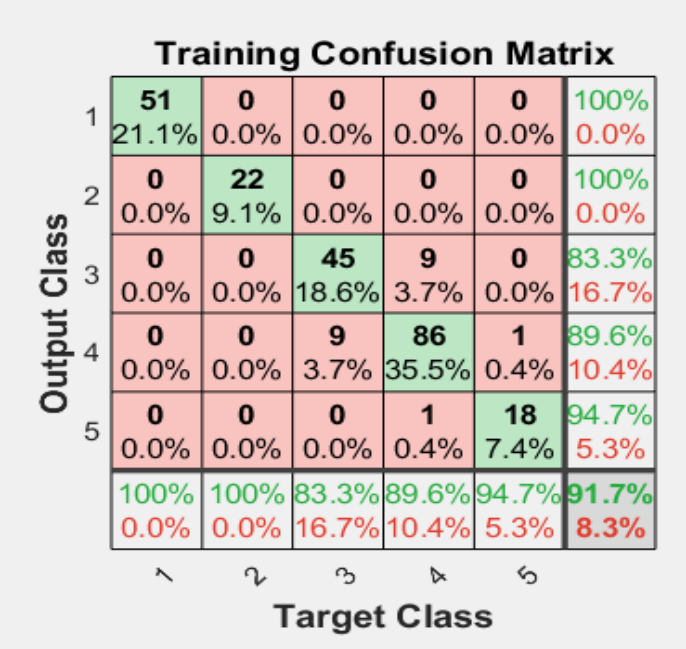


Figure (6.15): The training confusion matrix of the MLP neural network.

The rows represent the predicted neural network class (the output class), whereas the columns represent the actual class (target class). The observations that have been successfully categorized are shown by the diagonal cells. The observations that were not correctly categorized are shown by the off-diagonal cells. Each cell displays the overall number of observations and a percentage of the total observations. The percentage numbers shown in green reflect the proportion of correctly classified observations to the total observations. In contrast, the percentages displayed in red represent the proportion of data incorrectly classified.

The column on the far right of Figure (6.15) displays the percentages of all the predicted samples belonging to each class that was correctly and incorrectly classified. These metrics are often called the precision (or positive predictive value) for correctly classified classes and the false discovery rate for the incorrectly classified classes.

The row at the bottom of Figure (6.15) illustrates the percentages of all samples belonging to each target class that are correctly and incorrectly classified. These metrics are commonly referred to as recall also known as the true positive rate, and the false negative rate. The cell situated in the bottom right corner of the plot displays the overall accuracy. The optimal performance levels of a model are attained when the area under the ROC curve approaches 1 or when the ROC curve exhibits a straightened at the top of the graph, indicating 100% sensitivity (TPR) and 100% specificity (FPR). Accordingly, the current artificial neural network (ANN) classifiers show optimal performance for categories 1, 2, and 5. The ROC curve for classes 3 and 4 is almost straightened, leading to satisfactory outcomes, as shown in Figure (6.16)

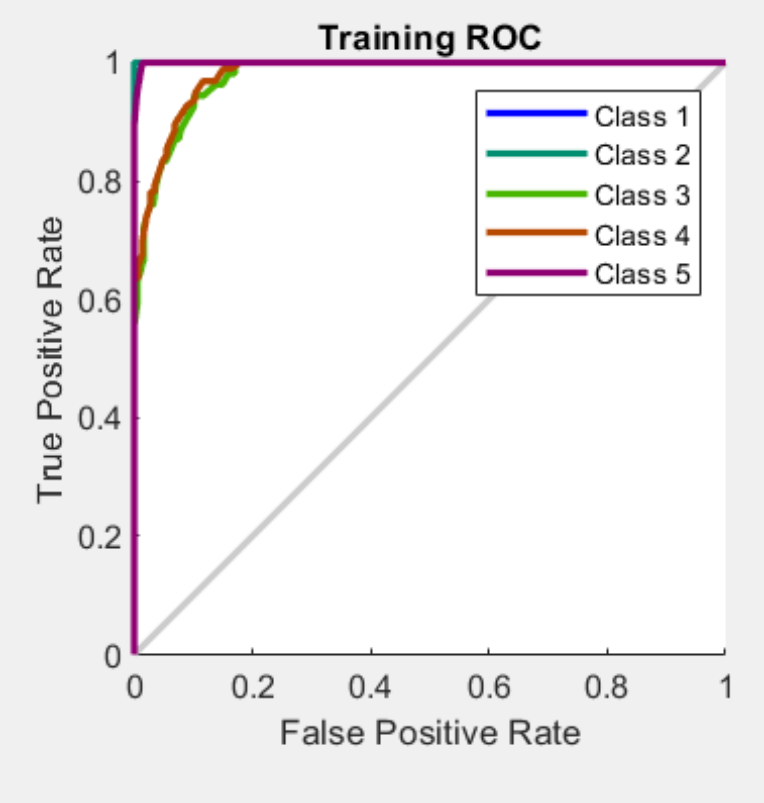


Figure (6.16): Training ROC curve for MLP neural network model.

Each predicted class probability receives a score or loss depending on how much it deviates from the actual value in the form of 0 or 1 in terms of cross-entropy loss. At various points in the learning process, model weights are updated using a loss function based on cross-entropy. The goal is to achieve a minimal loss, with a better model resulting from a smaller loss. The cross-entropy loss for a perfect model is 0. The best training performance for the current MLP model was achieved when the cross-entropy loss equals 0.0285 at epoch 100000, as presented in Figure (6.17).

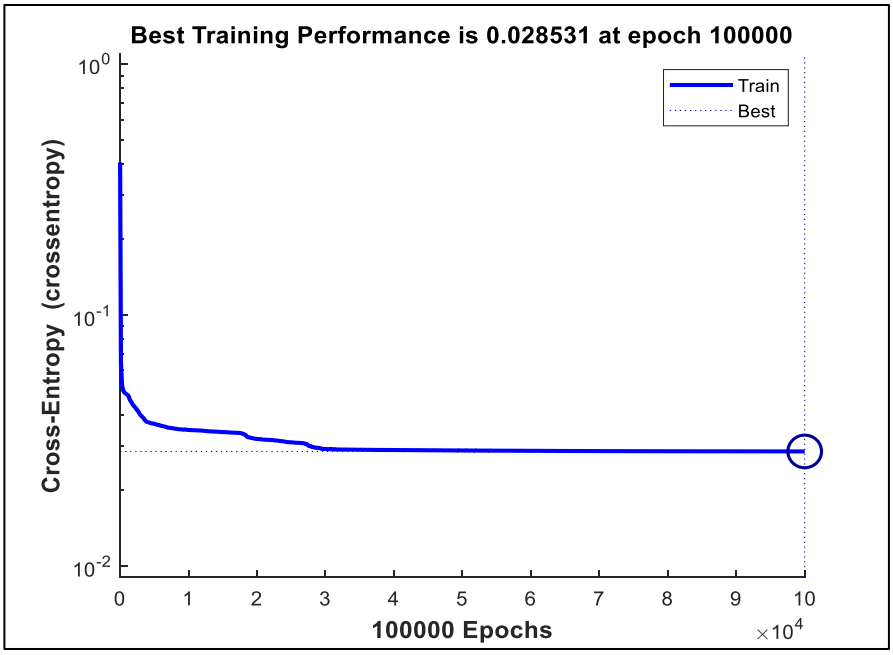


Figure (6.17): Cross-entropy error for MLP neural network model.

Finally, the model is tested using a testing dataset to calculate the model accuracy, which is 79.4%, as shown in the testing confusion matrix in Figure (6.18).

True Class					
	22	0	0	0	0
	3	7	0	0	0
	0	1	19	4	0
	0	0	5	35	2
	0	0	3	4	2

Predicted Class

Figure (6.18): Confusion matrix of testing data for MLP neural network.

Despite achieving a high accuracy of 91.7% for the training dataset, the ANN model cannot be deemed as the optimal classifier. This is due to the fact that a perfect MLP-ANN model should have a cross-entropy loss of 0, whereas the current classifier attained a cross-entropy loss of 0.0285 at epoch 100000. Furthermore, the accuracy for the testing dataset was only 79.4%, necessitating the training of an alternative classifier to yield improved classification outcomes.

6.4.5 Evaluating the k-Nearest Neighbours(kNN)-instance-based learners

The present investigation employed a weighted kNN classifier to train the dataset, in which the classification of a sample is determined by the majority vote of the nearest sample classes with a specified number of neighbors set at 10. The current model's performance is demonstrated in Figure (6.19).

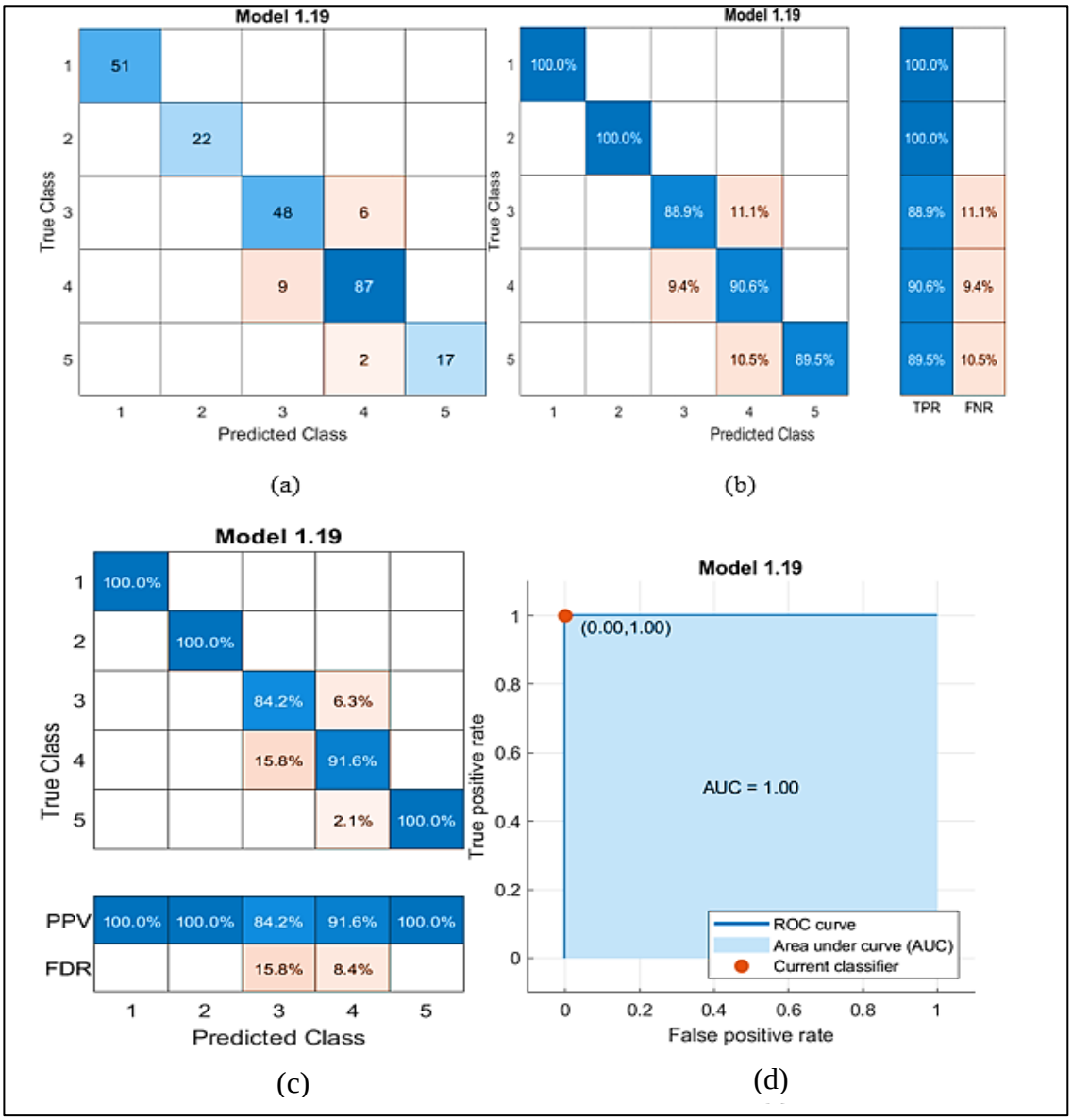


Figure (6.19): Performance of k-Nearest Neighbors (kNN) model.

(a) Confusion matrix in the form of observations number.

(b) TPR and FNR **(c)** PPV FDR **(d)** Area under ROC curve.

The total number of misclassified cells is (17), knowing the total number of the trained dataset is (242) through the confusion matrix presented in Figure 6.19 (a), which indicates an accuracy of 93.0% for the trained model. The wells of classes 1 and 2 are perfectly classified (100%) of TPR with (0%) of FNR, while for classes 3,4 and 5, the TPR is (88.9%, 90.6%, and 89.5%) and the FNR is (11.1%, 9.4%, 10.5%) respectively. The TPR is much higher than FNR, which indicates better performance for the kNN classifier.

From the ROC curve in Figure 6.19 (d), it can be noticed the area under curve AUC that measures the entire two-dimensional area underneath the ROC curve is 1 with (false positive rate, true positive rate) is (0,1), which indicates 100% accurate predictions. The testing confusion matrix in Figure (6.20) indicates that the weighted k-nearest neighbor (kNN) model achieved an overall accuracy of 90.7% for the testing dataset.

True Class					
	22	0	0	0	0
	0	10	0	0	0
	0	1	20	3	0
	0	0	5	37	0
	0	0	0	1	8

Predicted Class

Figure (6.20): Confusion matrix of testing data for the weighted k-nearest neighbor (kNN) model.

The assessment of the five supervised machine learning classifiers based on predetermined evaluation criteria reveals that the weighted k-nearest neighbor (kNN) model achieves better results, with a higher accuracy rate of 93.0% and 90.7% in the training and classification tasks, respectively. Hence, it is recommended to employ this model for the development of a groundwater potential map in the study area.

6.5 Machine learning -Based Mapping of Groundwater

In this study, the reliable classifier for accurate groundwater potential zone mapping was identified by comparing the predictive abilities of five machine-learning techniques based on the classifier's performance accuracy and evaluation criteria, such as True Positive Rates (TPR), False Negative Rates (FNR), Positive Predictive Values (PPV), False Discovery Rates (FDR), and the area under the curve of the receiver operating characteristics (ROC).

The weighted k-Nearest Neighbours (kNN) algorithm, which is an instance-based learning approach, has demonstrated higher accuracy in both the training and testing phases of machine learning. Specifically, it has achieved classification accuracies of 93.0% and 90.7%, respectively. Furthermore, the True Positive Rate (TPR) exhibits a significantly higher value in comparison to the False Negative Rate (FNR), in which a total of 37 samples (cells) out of 349 data points were misclassified during both the training and testing phases, thereby implying a more effective performance of the weighted k-Nearest Neighbors (kNN) classifier.

The study determined that the weighted-kNN model had an area under the receiver operating characteristic (ROC) curve (AUC) of 1, which denotes a perfect level of predictive accuracy. This was determined by measuring the entire two-dimensional area beneath the ROC curve, where the rate of false positives was set at 0, and the rate of true positives was set at 1. Consequently, this model was employed to produce the groundwater potential map shown in Figure (6.21).

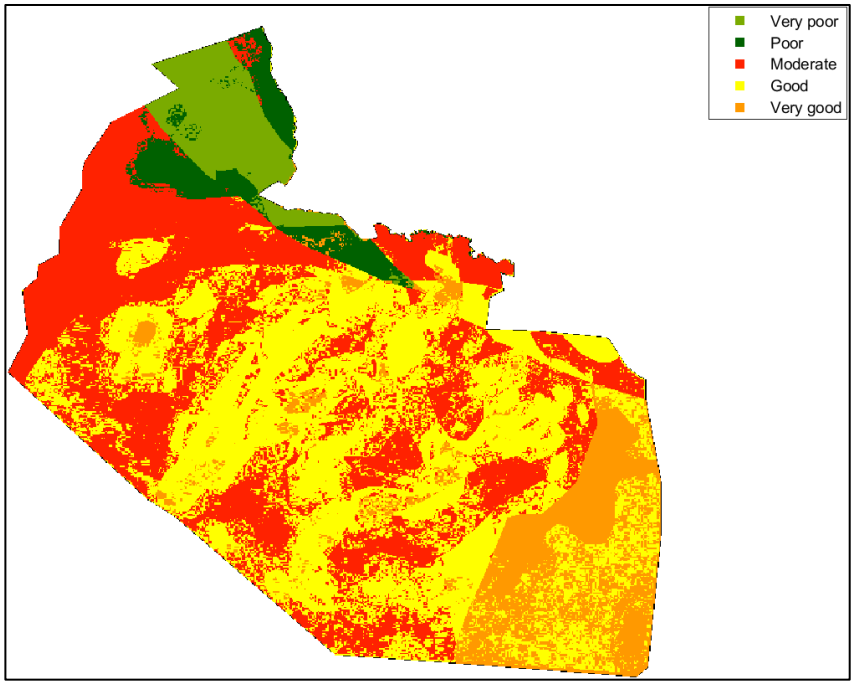


Figure (6.21): GWPZ map using the weighted-kNN classifier.

The findings of the weighted-kNN classifier-GWPZ map depicted in Figure (6.21) have led to the classification of the study area into five distinct groundwater potential zones. These zones have been identified as very good, good, moderate, poor, and very poor, and they respectively cover 12.37%, 42.12%, 35.69%, 4.24%, and 5.58% of the total study area. Table (6.2) presents the spatial distribution of the groundwater potential zone (GWPZ) map in the investigated region.

Table (6.2): Groundwater potential zones of the study area using the weighted-kNN classifier.

Rank	Class attribution	Percentage Area	Area (km^2)
5	v.good	12.37	9491.84
4	good	42.12	32319.82
3	moderate	35.69	27385.90
2	poor	4.24	3253.47
1	v.poor	5.58	4281.69

6.6 Binary Classification of Groundwater Potential Mapping

In this type of classification technique, groundwater potential mapping was formed as a two-classes system. The goal was to differentiate groundwater potential zones with high yield from those with low yield. A well yield of (9 ℓps) was used as a threshold value; wells with a pumping rate equal to or below (9 ℓps) were identified as low-yield groundwater (class-1), while any values greater than (9 ℓps) were identified as high-yield groundwater (class-2). The dataset was randomly divided into (70%) of the dataset (244 wells) for training and (30%) of the datasets (105 wells) for the testing/validation process. Figure (6.22) illustrates the data splitting.

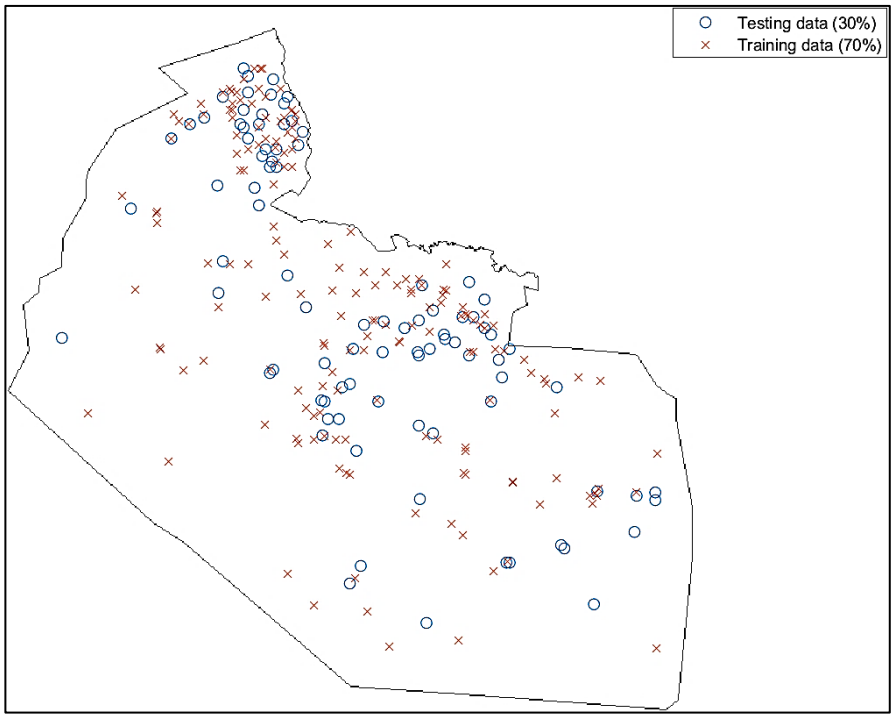


Figure (6.22): The binary training/testing data split.

The same five machine learning algorithms: Ensemble Boosted Trees (EBT); Naive Bayes (NB); Support Vector Machines (SVM); Multi-Layer Perceptron (MLP); and *k*-Nearest Neighbours (kNN), which are used to train and validate the five class models will be used in the development of the binary classification of

GWPZ mapping (two classes classification). In this case, a comparison of five machine-learning approaches will be assessed based on the classifier's performance using some metrics such as True Positive Rates (TPR), False Negative Rates (FNR), Positive Predictive Values (PPV), False Discovery Rates (FDR), the area under the curve of the receiver operating characteristics (ROC), and classifier accuracy to find out the best classifier performance for the accurate mapping of the groundwater potential zones. Figure (6.23) presents the performance of logic-based learners- Ensemble Boosted Trees.

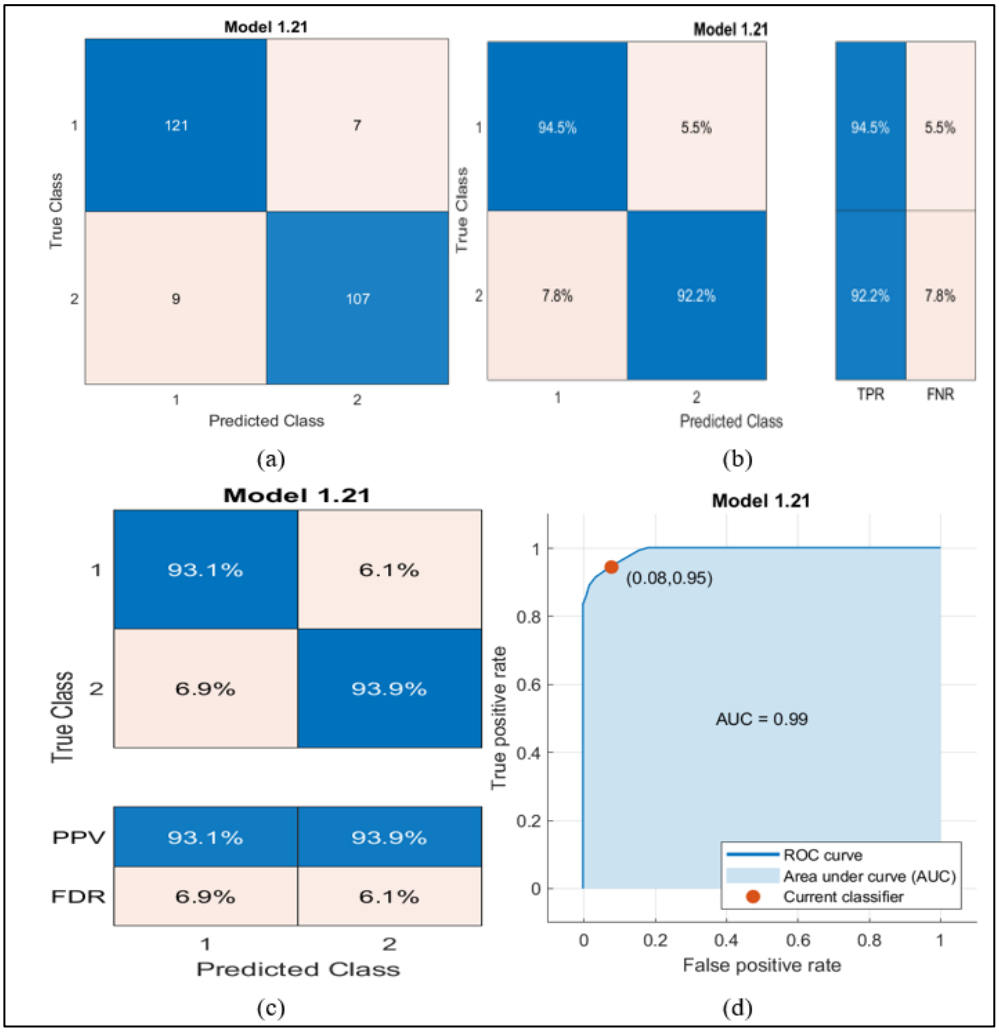


Figure (6.23): Performance of the EBT for binary classification.

(a) Confusion matrix in the form of observations number.

(b) TPR and FNR **(c)** PPV FDR **(d)** Area under ROC curve.

The performance of the Ensemble Boosted Trees (EBT) model was evaluated in terms of its true positive rate (TPR) and false negative rate (FNR) for wells classified as class-1. The TPR was found to be 94.5%, indicating that 121 out of 128 observations were correctly classified. On the other hand, the FNR was 5.5%, indicating that 7 out of 128 observations were incorrectly classified. Likewise, the same methodology was employed to compute TPR and FNR for class-2, which were found to be 92.2% and 7.8%, respectively. The calculations report the PPV and FDR for wells categorized as class-1, achieving values of 93.1% and 6.9%, respectively. The training data of the EBT model exhibited an overall accuracy of 93.4%. This value was obtained by computing the percentage of accurately classified classes in the diagonal cells of the confusion matrix, divided by the total number of observations (244) in the training dataset. The area under the ROC curve (AUC) for the current classifier is 0.99, as presented in Figure 6.23 (d).

The evaluation of the EBT model's accuracy is derived from the confusion matrix of the testing dataset, yielding a value of 93.3%, as illustrated in Figure (6.24). The calculation involves determining the proportion of accurate outcomes in the dataset's diagonal cells relative to the overall quantity of testing wells, which is 105.

True Class		
	51	4
	3	47

Predicted Class

Figure (6.24): Confusion matrix of testing data for the EBT model.

A comparative analysis was conducted on the performance of five machine learning classifiers utilized in this binary classification study to assess the effectiveness of predictive classifier capabilities. The findings that are based on

the six evaluation criteria are presented in Table (6.3), where the positive predictive values are shown in blue for the points in each class that were properly predicted, and the false discovery rates are displayed in orange for each class that was not correctly predicted.

Table (6.3): Training performances of the models.

Training Dataset	Classifier	No. of Observations			TPR	FNR	PPV	FDR	Accuracy	AUC
			1	2						
		1	Correctly Classified (TP)	Wrongly Classified (FP)						
		2	Wrongly Classified (FN)	Correctly Classified (TN)						
244	EBT	1	121	7	94.5%	5.5%	93.1%	6.9%	93.4%	0.99
		2	9	107	92.2%	7.8%	93.9%	6.1%		
244	SVM	1	108	20	84.4%	15.6%	95.6%	4.4%	89.8%	0.97
		2	5	11	95.7%	4.3%	84.7%	15.3		
244	NB	1	96	32	75.0%	25%	88.9%	11.1%	82.0%	0.92
		2	12	104	89.7%	10.3%	76.5%	23.5%		
244	MLP	1	120	5	96.0%	4.0%	93.8%	6.2%	94.7%	0.97
		2	8	111	93.3%	6.7%	95.7%	4.3%		
244	kNN	1	120	8	93.8%	6.2%	96.0%	4.0%	94.7%	0.99
		2	5	111	95.7%	4.3%	93.3%	6.7%		

Table (6.3) illustrates that k-Nearest Neighbors (kNN), which are instance-based learners, exhibit better outcomes in a training model. The classifier achieves a 94.7% accuracy rate. TPR and PPV for class-1 are 93.8% and 96.0%, while the FNR and FDR are 6.2% and 4.0%, respectively. The TPR and PPV for class-2 are 95.7% and 93.3%, while the FNR and FDR for class-2 are 4.3 and 6.7%, respectively. Consequently, the classifier's performance is better than that

of other trained classifiers, as demonstrated by its higher TPR and PPV values relative to FNR and FDR. Additionally, the AUC value of 0.99, which is nearly 1, indicates a high degree of predictive accuracy.

The present study employs the Multi-Layer Perceptron (MLP) structure of Artificial Neural Networks (ANN) for binary classification. The MLP structure comprises an input layer with seven neurons corresponding to the thematic maps' seven data features. Additionally, the structure includes two hidden layers, each consisting of five neurons, and an output layer with two neurons that execute the binary classification task. Figure (6.25) presents the MLP-ANN structure.

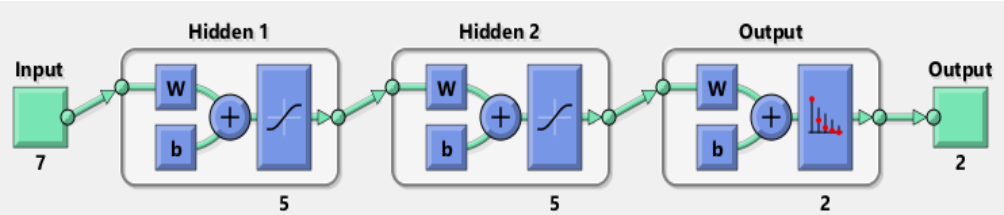


Figure (6.25): Structure of two classes MLP neural network.

The best training performance of the MLP neural network is when the cross-entropy loss is equal to 0.0485 at epoch 7481, as presented in Figure (6.26).

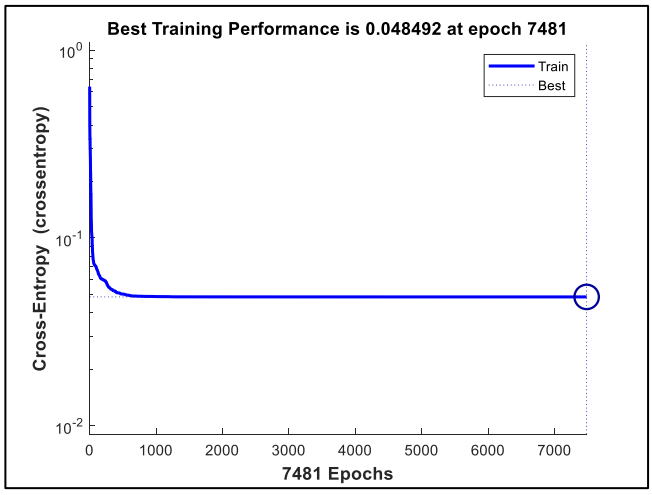


Figure (6.26): Cross-entropy error for binary MLP neural network model.

Every classifier is subject to the testing phase with a subset of 30% of the dataset, comprising 105 data points, for the purpose of model validation. Table

(6.4) presents the validation outcomes of the five training models, comprising the confusion matrix (count of observations) and the accuracy of model validation.

Table (6.4): Testing/Validation performances of the models.

Testing Dataset	Classifier	No. of Observations			Accuracy
			1	2	
		1	Correctly Classified (TP)	Wrongly Classified (FP)	
		2	Wrongly Classified (FN)	Correctly Classified (TN)	
105	EBT	1	49	6	89.5%
		2	5	45	
105	SVM	1	48	7	88.6%
		2	5	45	
105	NB	1	41	14	83.8%
		2	3	47	
105	MLP	1	51	4	90.5%
		2	6	44	
105	kNN	1	51	4	93.3%
		2	3	47	

According to Table (6.4), the k-Nearest Neighbours (kNN) algorithm is the optimal classifier for constructing the GPM due to the classifier's high level of accuracy during the testing phase, which was determined to be 93.3%. Additionally, out of a total of 105 cells, 98 were correctly classified, while only 7 were misclassified. Consequently, the kNN machine learning classifier employed to produce the groundwater potential map GPM, as shown in Figure (6.27). The

map will be displayed using binary classification, which aims to differentiate areas with low-yield groundwater potential from those with high-yield groundwater potential.

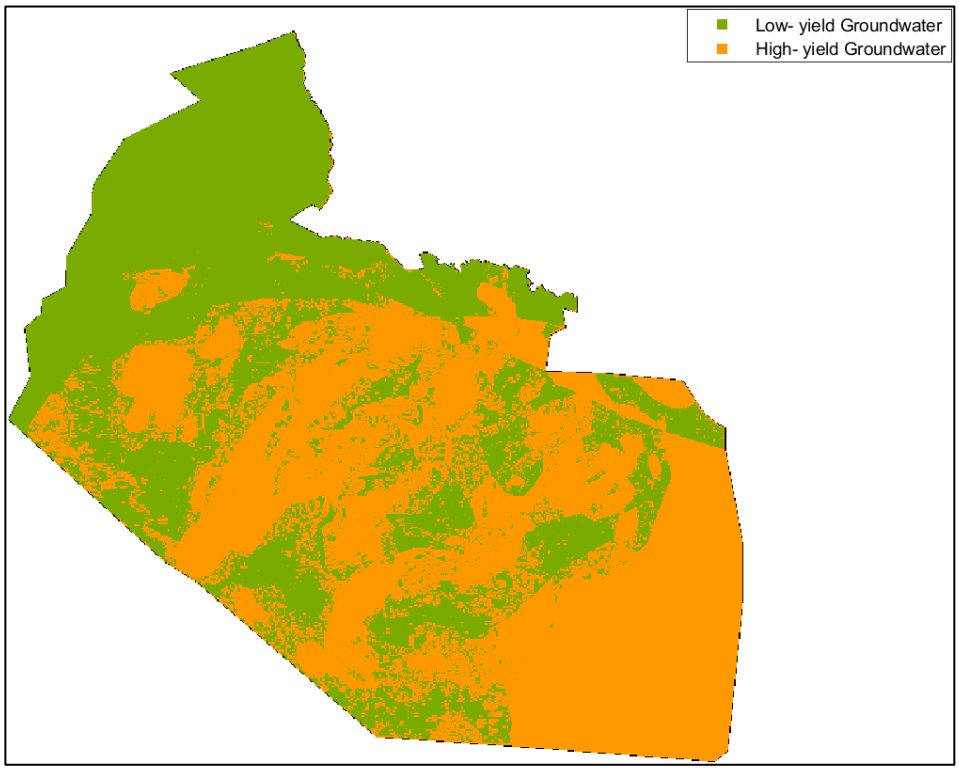


Figure (6.27): Binary GWPZ map using the kNN classifier.

The areal distribution of GWPZs for class-1, which represents low-yield groundwater zones, and class-2, which refers to high-yield groundwater zones, account for 45.51% and 54.49% of the total study area, respectively. The computed extent of the region characterized by groundwater of high- yield was 41809.27 Km^2, while the area with the groundwater of low- yield was determined to be 34923.43 Km^2.

6.7 Summary of the Key Findings and Discussion Points

The findings obtained from training and testing five different classifiers indicated that the weighted k-Nearest Neighbours (kNN) classifier, an instance-based learning approach, outperformed other classifiers in both multiclass and

binary supervised classification tasks. The model achieved classification accuracies in both the training and testing phases of 93.0% and 90.7% for multiclass classification, and 94.7% and 93.3% for binary classification, respectively. Furthermore, the True Positive Rate (TPR) showed a significantly higher value compared to the False Negative Rate (FNR), resulting in an area under the curve (AUC) of 1. According to Sarkar et al. (2022), an AUC of 1 with a corresponding (0,1) false positive rate and true positive rate signifies accurate model output. Given the satisfactory results in identifying areas with significant groundwater potential, the generated hybrid model can be effectively utilized to examine additional locations and provide Groundwater Potential Maps (GPMs) in a cost-effective manner without requiring extensive field investigations.

Data-driven machine learning models demonstrate superior predictive capabilities for assessing groundwater potential compared to knowledge-driven models like the Analytical Hierarchy Process (AHP) techniques. The AHP method yielded an accuracy rate of 82.5%, with 288 correctly identified well locations out of 349 wells. In the multiclass classification context, the ML weighted-kNN model accurately classified 322 observations, while the binary model accurately classified 329 out of 349 wells with validation accuracy of 90.7% and 93.3% for both multiclass and binary classification. Moreover, the supervised machine learning methodology is preferred due to its ability to learn patterns and correlations among variables based on training data and apply this knowledge to make predictions on unseen or unfamiliar data in the testing phase.

Conversely, the AHP method relies on expert experiences, which can be influenced by expert knowledge and subjectivity. The Analytic Hierarchy Process involves assessing the relative importance of criteria and sub-criteria through pairwise comparisons made by decision-makers. These assessments are then utilized to determine weights for each criterion, which are combined to calculate

an overall score or ranking. Consequently, machine learning is well suited for mapping complicated spatially distributed variables, such as the occurrence of groundwater, as reported by Gómez-Escalonilla et al. (2022).

6.8 Development of MATLAB Graphical User Interface Application

Graphical user interfaces (GUIs), commonly referred to as applications, offer users the ability to control software applications through a visual interface, thereby avoiding the necessity for users to acquire proficiency in a programming language or manually input commands to execute the application.

The App Designer tool in MATLAB enables the creation of advanced apps that utilize the drag-and-drop feature to arrange visual components for the purpose of designing a graphical user interface (GUI) and employing the integrated editor to program its behavior efficiently. The App Designer tool allows the incorporation of graphical user interface elements such as sliders, dropdown menus, text fields for editing, and buttons into the software application without the need for manual coding. Conversely, establishing a connection between the GUI elements and the hybrid models that were formulated using the five different classifiers is crucial for generating outputs of binary or multiclass classifications.

To develop a software application, the App Designer tool facilitates the integration of the two fundamental tasks involved in the process of building an application: creating a user interface and defining app behavior. During the initial stage of GUI design, the design canvas is used to arrange visual components through the drag-and-drop functionality, as shown in Figure (6.28).

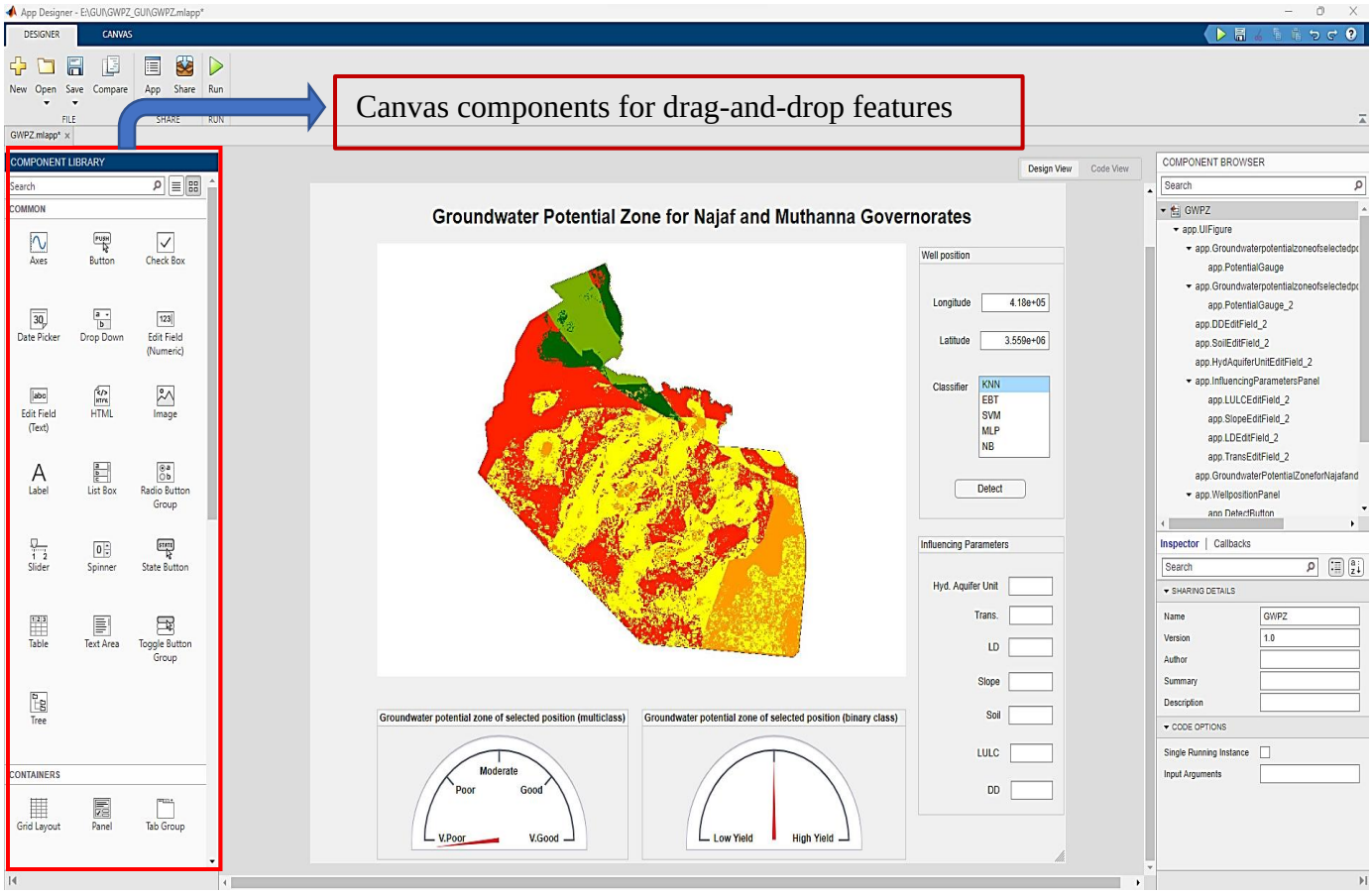


Figure (6.28): The interface of the App Designer

In the second stage of the process, it is recommended to use the integrated version of the MATLAB editor in order to specify the behavior of the application. The Code Analyzer feature in App Designer has the capability to automatically detect coding issues. During the process of writing code, it is possible to observe warning and error messages and make necessary modifications to the application based on these messages, as presented in Figure (6.29).

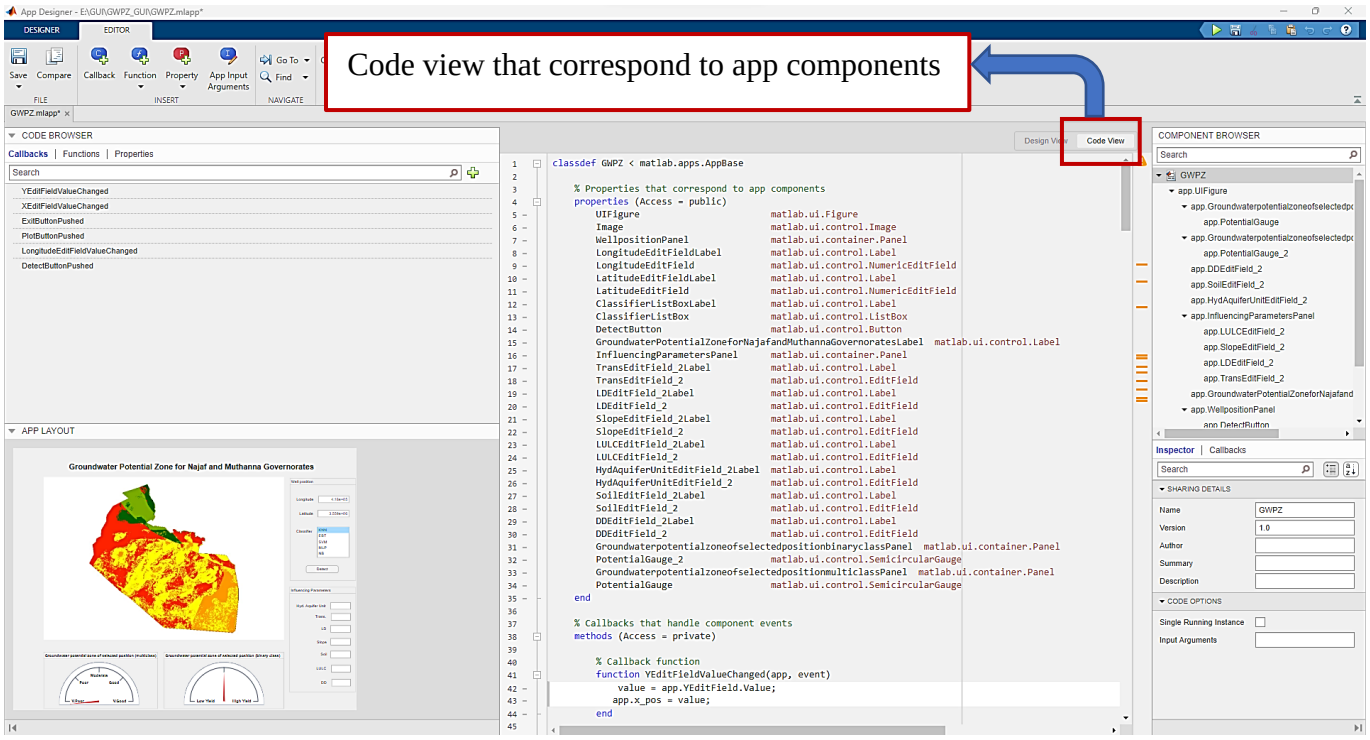


Figure (6.29): The Code View feature in App Designer.

The GUI application that was developed for the purpose of identifying potential groundwater zones within the study area has been organized into four different regions. The creation of each region was facilitated through the application of drag-and-drop functionalities, in which the necessary components were inserted and organized within a designated container, commonly referred to as a panel. The design of the groundwater potential application primarily involves the use of several key elements. These include a numeric edit field for inputting the longitude and latitude coordinates of the well to be classified, a state button to start the detection process for the corresponding groundwater potential zone, a list box for selecting the desired machine learning classifier, and a text-based edit field for presenting the influencing factors' features. The output region included the semicircle gauge from the instrumentation section to display the final classification of the inserted well location. This application allows the efficient management of groundwater resources and provides preliminary and reliable

information on the ideal wells sites before starting the drilling process, especially in areas of limited groundwater information to target appropriate drilling sites for groundwater wells. The four designated regions are shown in Figure (6.30).

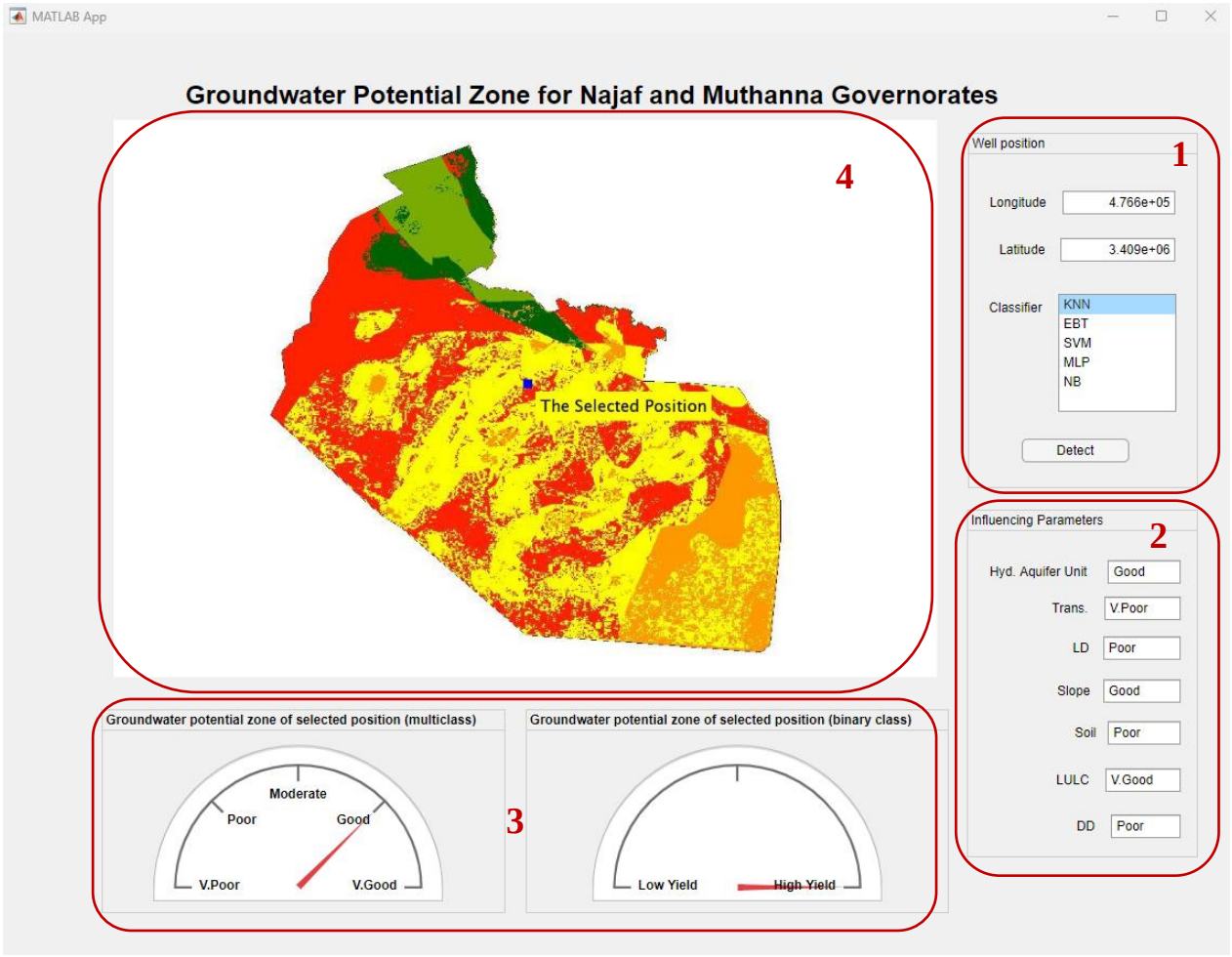


Figure (6.30): GWPZ application with a GUI in MATLAB.

(1) Region-1 has been designated for the purpose of inputting the geographical coordinates (longitude and latitude) of wells using the UTM Projection, for which classification is necessary. Additionally, this region includes the process of choosing the appropriate classifier and detection button.

(2) Region-2 has been created to analyze the features of the seven thematic maps that represent different influencing factors. These factors include hydrogeological aquifer unit types, transmissivity, lineaments density, slope,

soils, land use and land cover (LU/LC), and drainage density that significantly influence groundwater availability.

(3) Region-3 has been developed to display the results of a selected classifier that has been utilized for training and testing features extracted from seven thematic maps. The output is a multiclass and binary classification, which identifies the class to which the selected well belongs. In the case of multiclass classification, the selected well is categorized into one of five classes, namely V. poor, Poor, Moderate, Good, or V.good. On the other hand, in binary classification, the selected well is evaluated to determine whether it belongs to the Low-yield or High-yield category.

(4) Region-4 shows the study area map covering the Najaf and Muthanna governorates. Its purpose is to illustrate the predicted zone of the chosen well, whether it falls within or outside the study area.

The packaging of App Designer applications for installation into the MATLAB Apps tab is also a feasible option. MATLAB Compiler enables the creation of standalone desktop and web applications from apps, which can be shared with individuals who do not possess MATLAB proficiency; therefore, it is possible to release applications as standalone desktop or web applications. Within the App Designer interface, users can use the share function located within the Designer tab to select from a range of three available options: the MATLAB App, which generates an installation file for sharing with other MATLAB users; the Web App, which deploys web applications using the MATLAB compiler; and the Standalone Desktop App, which creates a standalone desktop application using the MATLAB compiler. This study aims to develop a standalone application that can be shared with individuals who do not possess proficiency in MATLAB using the Standalone Desktop App. The process of packaging apps in App Designer is presented in Figure (6.31).

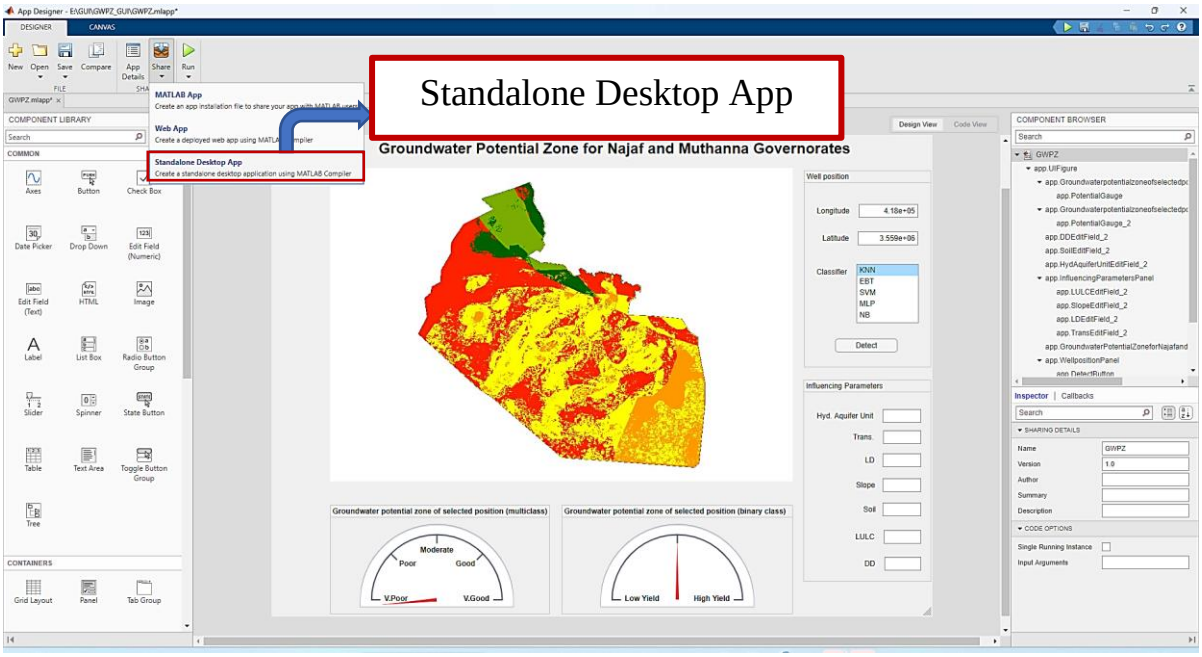


Figure (6.31): App designer packages.

Following selecting the standalone Desktop App option, a window containing library information will be displayed, enabling the user to input application details such as the developer's name, email address for contact purposes, company name, and a summary describing the application, as presented in Figure (6.32). This information is optional and displayed in the Windows add/remove programs control panel.

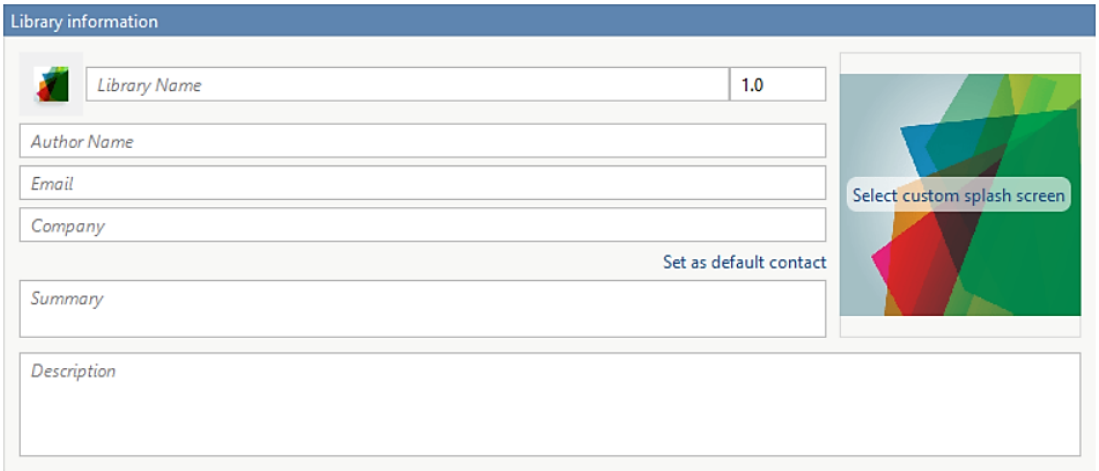


Figure (6.32): Application information.

Chapter Seven

Conclusions and Recommendations

The key results from the study provide insights into the research's importance. Therefore, a comprehensive summary of the research conclusions and practical recommendations based on the results were outlined in the following.

7.1 Conclusions

The significant conclusions of this research can be summarized as follows:

1. In the static circumstances simulation conducted using the MODFLOW package in GMS v10.4 software, it was observed that the groundwater within the Dammam confined aquifer in the study area exhibits a flow pattern originating from the southwest in the direction of Saudi Arabia and moving towards the northeast, close to the Euphrates River. The head contour map revealed that elevated heads are detected in the southwestern section of the study area at approximately 260 meters, representing a recharge zone. Conversely, lower heads are observed near the Euphrates River within the discharge zone, exhibiting a piezometric head of 10 meters. This observation is consistent with the groundwater flow map constructed using field data, indicating that the groundwater piezometric head varies between 208.423 and 257.65 meters in the southwestern regions, while ranging from 20.78 to 69.08 meters in the northeastern areas.

2. The sensitivity analysis of aquifer hydraulic conductivity and recharge rate under steady-state and specific storage during transient conditions indicates that the model is more sensitive to variations in hydraulic conductivity than recharge rate. The hydraulic conductivity of the Dammam confined aquifer in the study area varies from 1.47 to 20.0 m/day. The recharging rate was

determined to be 1.66×10^{-6} m/day, and the storativity values ranged between 5×10^{-5} and 6.5×10^{-4}.

3. The operation of 89 wells at their full operational capacity caused a decline in groundwater heads with a range of 0.4 to 5.8 meters. While operating 228 wells, the drawdown increased significantly from 1 to 22 meters. This increase can be attributed to the rise in overall pumping rates, which increased from 582.3 to 1534.1 ℓps. However, regions with minimal drawdowns continue to show promising prospects for the implementation of further drilling operations, as the pumping process has a moderate influence on these areas.
4. The groundwater availability in the Dammam confined aquifer within the study area was significantly influenced by hydrological aquifer unit, transmissivity, lineament density, slope, soil type, land use and land cover, and drainage density. The normalized weights for these factors were determined using a pairwise comparison matrix (PCM) and were found to be 38.567%, 19.284%, 12.856%, 9.642%, 7.713%, 6.428%, and 5.510%, respectively.
5. The developed groundwater potential map (GPM) classifies the Dammam confined aquifer of the study area into five zones, including very good, good, moderate, poor and very poor GWPZs. The distribution of these zones in terms of percentage to the total area of 76732.71 Km^2 is as follows:15.91%, 39.02%, 33.01%, 6.49%, and 5.57%, respectively. The AHP method accurately delineates a groundwater potential map with 82.5% accuracy. Specifically, out of 349 well locations, 288 were correctly identified using this method.
6. Among the five developed hybrid ML models (Ensemble Boosted Trees, Naive Bayes, Support Vector Machines, Multi-Layer Perceptron, and weighted k-Nearest Neighbors), the multiclass weighted-kNN model proves most effective, with an accuracy of 90.70% in identifying areas with significant groundwater potential, followed by other models with validation accuracies of

89.7%, 79.4%, 77.6%, and 75.7% respectively. Out of 349 observation wells, the kNN model accurately classified 322 wells. Class 1 and 2 wells achieved perfect ratings with 100% True Positive Rate (TPR) and 0% False Negative Rate (FNR), while classes 3-5 had TPRs of 89%, 90%, and 89.5%, respectively. Moreover, the area under curve achieves a value of 1 when the (FPR, TPR) pair equals (0,1), signifying accurate predictions. The areas of zones identified by the kNN hybrid model are approximately 12.37%, 42.12%, 35.69%, 4.24%, and 5.58% of the total area, respectively.

7. The binary classification of high-yield and low-yield groundwater potential zones was conducted using a threshold value of 9 ℓps. The kNN classifier showed the highest performance, achieving an accuracy of 93.3%, followed by MLP, EBT, SVM, and NB, with accuracies of 90.5 %, 89.5 %, 88.6 %, and 83.8%, respectively.
8. The high-yield groundwater area was estimated to be 41809.27 km^2, while the low-yield groundwater area covered 34923.43 km^2, representing 45.51% and 54.49% of the total area, respectively.
9. The developed standalone application using MATLAB's Graphical User Interface (GUI) has the ability to accurately identify the category of a selected well in both the multiclass and binary classification by only inputting its UTM coordinates (longitude and latitude). This application facilitates the effective management of groundwater resources. It offers preliminary and dependable data concerning the optimal location for the drilling process, particularly in regions with limited groundwater information. The developed hybrid models have the capability to detect locations with significant groundwater potential since they can be easily updated with new data, enabling their application in examining other areas.

7.2 Recommendations

Recommendations for further works and future studies based on the provided points of conclusion are:

1. Conducting further investigations to analyze the groundwater flow patterns and dynamics in the study area, considering temporal variations and seasonal changes. Exploring the impact of climate change and land-use changes on groundwater resources, incorporating these factors into the modeling and prediction processes.
2. Applying the developed models and techniques to other regions with similar hydrogeological characteristics to enhance the accuracy and efficiency of groundwater potential mapping and monitoring.
3. Investigating the potential interactions between surface water and groundwater systems, particularly in regions where groundwater discharge contributes to streamflow or affects environmental systems.

REFERENCES

Abijith, D., Saravanan, S., Singh, L., Jennifer, J.J., Saranya, T., and Parthasarathy, K.S.S., 2020. "GIS-based multi-criteria analysis for identification of potential groundwater recharge zones-a case study from Ponnaniyaru watershed, Tamil Nadu, India", HydroResearch, Vol. 3, pp.1-14. DOI: https://doi.org/10.1016/j.hydres.2020.02.002.

Abrams, W., Ghoneim, E., Shew, R., LaMaskin, T., Al-Bloushi, K., Hussein, S., AbuBakr, M., Al-Mulla, E., Al-Awar, M. and El-Baz, F., 2018. "Delineation of groundwater potential (GWP) in the northern United Arab Emirates and Oman using geospatial technologies in conjunction with Simple Additive Weight (SAW), Analytical Hierarchy Process (AHP), and Probabilistic Frequency Ratio (PFR) techniques", Journal of Arid Environments, vol. 157, pp.77-96. DOI: https://doi.org/10.1016/j.jaridenv.2018.05.005

Achirul, N., Kudang, B. S., Dodi, N, and Akhiruddin, M., 2018. "A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection," Information, vol. 9, no. 1, p. 5. DOI: https://doi.org/10.3390/info9010005

Adiat, K.A.N., Nawawi, M.N.M., and Abdullah, K., 2012. "Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool–a case of predicting potential zones of sustainable groundwater resources", Journal of Hydrology, vol. 440, pp.75-89. DOI: https://doi.org/10.1016/j.jhydrol.2012.03.028

Agarwal, E., Agarwal, R., Garg, R.D., and Garg, P.K., 2013. "Delineation of groundwater potential zone: An AHP/ANP approach. Journal of Earth System Science", vol. 122, pp.887-898. DOI: https://doi.org/10.1007/s12040-013-0309-8

Agatonovic-Kustrin, S., and Beresford, R., 2000. "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research", Journal of Pharmaceutical and Biomedical Analysis, vol. 22, no. 5, pp. 717-727. DOI: https://doi.org/10.1016/S0731-7085(99)00272-1.

Ahmed II, J.B., and Mansor, S., 2018. "Overview of the application of geospatial technology to groundwater potential mapping in Nigeria", Arabian Journal of Geosciences, vol. 11, p. 504. DOI: https://doi.org/10.1007/s12517-018-3852-4.

References

Ahmed, R., and Sajjad, H., 2018. "Analyzing factors of groundwater potential and its relation with population in the Lower Barpani Watershed", Assam, India. Natural Resources Research, vol. 27, no. 4, pp.503-515. DOI: https://doi.org/10.1007/s11053-017-9367-y

Al-Abadi, A. M. and Al-Shamma'a, M. A., 2014. "Groundwater Potential Mapping of the Major Aquifer in Northeastern Missan Governorate, South of Iraq by Using Analytical Hierarchy Process and GIS." Journal of Environment and Earth Science, vol. 4, pp.125-149.

Al-Areedhi, H.H., and Khayyun, T.S., 2019. "Modelling of Groundwater Flow for the Iraqi Aquifers", Doctoral dissertation, University of Technology.

Al-Basrawi, N.H., 1996. "Hydrogeology of Razzaza Lake Iraq's western desert", Doctoral dissertation, PhD Thesis. College of Science, University of Baghdad, Baghdad, Iraq.

Alfeilat, H.A., Hassanat, A., Lasassmeh, O., Tarawneh, A.S., Alhasanat, M.B., Salman, H.E., and Prasath, V.B., 2019. "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review", Big data, vol.7, no. 4, pp 221-248. DOI: http://doi.org/10.1089/big.2018.0175

Al-Fugara, A.K., Pourghasemi, H.R., Al-Shabeeb, A.R., Habib, M., Al-Adamat, R., Al-Amoush, H. and Collins, A.L., 2020. "A comparison of machine learning models for the mapping of groundwater spring potential", Environmental Earth Sciences, vol. 79, pp. 1-19. DOI: https://doi.org/10.1007/s12665-020-08944-1.

Al-Hassoun, S.A., and Mohammad, T.A., 2011. "Prediction of water table in an alluvial aquifer using Modflow", Pertanika Journal of Sciences and Technology, vol. 19, no. 1, pp. 45-55.

Ali, H. A., 2021. "Hydrogeology Study of the Upper Al-Dammam Aquifer-Southwest-Iraq", Ph.D. Thesis, University of Baghdad, Baghdad, Iraq.

Ali, S.S., and Hamamin, D.F., 2012. "Groundwater Vulnerability Map of Basara Basin, Sulaimani Governorate, Iraqi Kurdistan Region", Iraqi Journal of Science, vol. 53, no. 3, pp. 579-594.

Al-Jiburi, H.K., and Al-Basrawi, N.H., 2009. "Hydrogeology". Iraqi Bulletin of Geology and Mining, no. 2, pp. 77-91, DOI: https://doi.org/10.5194/esurf-10-1-2022.

References

Al-Jiburi, H.K., and Al-Basrawi, N.H., 2021, "Hydrogeological and Hydrochemical study of Najaf sheet NH- 38 - 2 scale 1: 250000", Library Report No. 2404.

Al-Kharabsheh, A., 2000. "Ground-water modelling and long-term management of the Azraq basin as an example of arid area conditions (Jordan)", Journal of Arid Environments, vol. 44, no. 2, pp. 143-153. Doi: https://doi.org/10.1006/jare.1999.0580.

Al-Mussawy, W.H., and Khalaf, R.M., 2013. "Optimum management models for groundwater use in Karbala desert area", PhD Thesis, Al-Mustansiriyah University, Baghdad, Iraq.

Al-Ruzouq, R., Shanableh, A., Yilmaz, A.G., Idris, A., Mukherjee, S., Khalil, M.A., and Gibril, M.B.A., 2019. "Dam site suitability mapping and analysis using an integrated GIS and machine learning approach", Water, vol. 11, no. 9, p.1880. DOI: https://doi.org/10.3390/w11091880.

Al-Siba′ai, M., 2005. "Modeling of groundwater movement (Euphrates lower basin)", Damascus University for Basic Sciences J, vol. 21, no. 2, pp.91-114.

Al-Taee T. M., and Al- Sadiq A. A., 2003. "A considerable model for determining groundwater movement trends and estimating quantities in the Sulayfani plain, north of the Mosul Dam Lake", Journal of Damascus University, vol. 20, no. 2, (in Arabic).

Amin, M.A., and Yan, H., 2011. "High speed detection of retinal blood vessels in fundus image using phase congruency", Soft Computing, vol. 15, pp. 1217-1230. DOI: https://doi.org/10.1007/s00500-010-0574-2

Anderson, M.P., Woessner, W.W., and Hunt, R.J., 1992. "Applied groundwater modeling: Simulation of flow and advective transport", Academic Press Inc., San Diego, CA. Journal of Hydrology, vol. 140, pp. 393-395.

Anderson, M.P., Woessner, W.W., and Hunt, R.J., 2015. "Applied groundwater modeling: simulation of flow and advective transport", Academic Press.

Aquaveo, L.L.C., 2017. "Mesh editing", SMS 12.2 tutorials.

Arefin, R., 2020. "Groundwater potential zone identification at Plio-Pleistocene elevated tract, Bangladesh: AHP-GIS and remote sensing approach", Groundwater for sustainable development, vol. 10, p. 100340. DOI: https://doi.org/10.1016/j.gsd.2020.100340

Asgher, M.S., Kumar, N., Kumari, M., Ahmad, M., Sharma, L., and Naikoo, M.W., 2022. "Groundwater potential mapping of Tawi River basin of Jammu District, India, using geospatial techniques", Environ. Monit. Assess, vol. 194, no. 240. DOI: https://doi.org/10.1007/s10661-022-09841-9

Carrard, N.; Foster, T.; Willetts, J., 2019. "Groundwater as a Source of Drinking Water in Southeast Asia and the Pacific: A Multi-Country Review of Current Reliance and Resource Concerns". Water , vol.11, no.8 p.1605. https://doi.org/10.3390/w11081605

Chatterjee, R.S., Pranjal, P., Jally, S., Kumar, B., Dadhwal, V.K., Srivastav, S.K., and Kumar, D., 2020. "Potential groundwater recharge in north-western India vs spaceborne GRACE gravity anomaly based monsoonal groundwater storage change for evaluation of groundwater potential and sustainability", Groundwater for sustainable development, vol. 10, p. 100307. DOI: https://doi.org/10.1016/j.gsd.2019.100307

Chatterjee, S. and Dutta, S., 2022 "Assessment of groundwater potential zone for sustainable water resource management in south-western part of Birbhum District, West Bengal", Applied Water Science, vol. 12, no. 40. DOI: https://doi.org/10.1007/s13201-021-01549-4

Chen, J-C, Wang, Y-M., 2020. "Comparing Activation Functions in Modeling Shoreline Variation Using Multilayer Perceptron Neural Network", Water, vol.12, no. 5, p. 281. DOI: https://doi.org/10.3390/w12051281

Chipman, H.A., George, E.I., and McCulloch, R.E., 2010. "BART: Bayesian additive regression trees", The Annals of Applied Statistics, vol. 4, no. 1, 266-298, DOI: https://doi.org/10.1214/09-AOAS285

Chowdhury, A., Jha, M.K., and Chowdary, V.M., 2010. Delineation of groundwater recharge zones and identification of artificial recharge sites in West Medinipur district, West Bengal, using RS, GIS and MCDM techniques. Environmental Earth Sciences, vol. 59, pp. 1209-1222. DOI: https://doi.org/10.1007/s12665-009-0110-9

CHRISTENSEN, S., and COOLEY, R.L., 2003. "Experience gained in testing a theory for modelling", In Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality: Proceedings of the ModelCARE 2002 Conference: Held in Prague, Czech Republic, 17-20 June, 2002 (No. 277, p. 22). International Assn of Hydrological Sciences.

Da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., and dos Reis Alves, S.F., 2017. "Artificial neural network architectures and training processes", Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-43162-8_2.

Darcy, H.,1856. "Les Fontaines Publiques de la ville de Dijon", Dalmont, Paris.

Dai, Ce, Jie Tang, Zhaoyang Li, Yucong Duan, Yunke Qu, Yao Yang, Hang Lyu, Dan Zhang, and Yongtai Wang. 2022. "Index System of Water Resources Development and Utilization Level Based on Water-Saving Society", Water, vol.14, no. 5, p. 802. DOI: https://doi.org/10.3390/w14050802

Dao, D. V., Adeli, H. Ly, H.-B. et al., 2020. "A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation," Sustainability, vol. 12, no. 3, p. 830.

Devi, S.P., Srinivasulu, S., and Raju, K.K., 2001. "Delineation of groundwater potential zones and electrical resistivity studies for groundwater exploration", Environmental Geology, vol. 40, no. 10, pp.1252-1264. DOI: https://doi.org/10.1007/s002540100304

Di Liu, Ashok K. Mishra, Zhongbo Yu, Haishen Lü, Yajie Li, 2021."Support vector machine and data assimilation framework for Groundwater Level Forecasting using GRACE satellite data", Journal of Hydrology. vol. 603, Part A, 126929, https://doi.org/10.1016/j.jhydrol.2021.126929.

Díaz-Alcaide, S., and Martínez-Santos, P., 2019. "Advances in groundwater potential mapping", Hydrogeology Journal, vol. 27, no. 7, pp.2307-2324. DOI: https://doi.org/10.1007/s10040-019-02001-3.

Doherty, J., 1998. "PEST; Model Independent Parameter Estimation, 1998; User's Manual; Watermark Computing, Australia", Unpublished M.S.c thesis, ITC, Enschede.

Doherty, J.E., and Hunt, R.J., 2010. "Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration", (Vol. 2010), Reston, VA, USA: US Department of the Interior, US Geological Survey. DOI: https://doi.org/10.3133/sir20105169

Doherty, J.E., Fienen, M.N., and Hunt, R.J., 2010. "Approaches to highly parameterized inversion: Pilot point theory, guidelines, and research

directions", vol. 2010, US Department of the Interior, US Geological Survey. URL: https://pubs.usgs.gov/sir/2010/5168/

Elbeih, S. F., 2015. "An overview of integrated remote sensing and GIS for groundwater mapping in Egypt", Ain Shams Engineering Journal, vol. 6, pp.1-15. DOI: https://doi.org/10.1016/j.asej.2014.08.008

Falah, F., and Zeinivand, H., 2019. "Gis-based groundwater potential mapping in khorramabad in lorestan, Iran, using frequency ratio (fr) and weights of evidence (woe) models", Water Resources, vol. 46, pp.679-692. DOI: https://doi.org/10.1134/S0097807819050051

Feng, Z.K., Niu, W.J., Zhang, R., Wang, S., and Cheng, C.T., 2019. "Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization", Journal of Hydrology, vol. 576, pp. 229-238. DOI: https://doi.org/10.1016/j.jhydrol.2019.06.045

Ganapuram, S., Kumar, G.V., Krishna, I.M., Kahya, E., and Demirel, M.C., 2009. "Mapping of groundwater potential zones in the Musi basin using remote sensing data and GIS", Advances in Engineering Software, vol. 40, no. 7, pp.506-518. DOI: https://doi.org/10.1016/j.advengsoft.2008.10.001

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., and Lin, Q., 2020. "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation", Journal of Hydrology, vol. 589, p. 125188. DOI: https://doi.org/10.1016/j.jhydrol.2020.125188

Gavankar, S.S., and Sawarkar, S.D., 2017. "Eager decision tree", 2nd International Conference for Convergence in Technology (I2CT), pp. 837-840. IEEE. DOI: https://doi.org/10.1109/I2CT.2017.8226246

GCGW, 2021. General Committee for Groundwater, Ministry of Water Resources, Iraq.

Ghasemian, D., 2016. "Groundwater management using remotely sensed data in high plains aquifer," Doctoral dissertation, University of Arizona, pp. 2643–2672.

Ghobadi, F., and Doosun, K. 2023. "Application of Machine Learning in Water Resources Management: A Systematic Literature Review" Water, vol. 15, no. 4, p. 620. DOI: https://doi.org/10.3390/w15040620

GMS User Manual, 2018. "v10.4, Groundwater Modeling System software tutorial".

Gómez-Escalonilla, V., Martínez-Santos, P., and Martín-Loeches, M., 2022. "Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions, Mali. Hydrology and Earth System Sciences", vol. 26, no. 2, pp. 221-243. DOI: https://doi.org/10.5194/hess-26-221-2022

Guyon, I., 2009. "A practical guide to model selection", Proc. Mach. Learn. Summer School Springer Text Stat, pp.1-37. URL: http://eprints.pascal-network.org/archive/00005768/01/guyonmlss.pdf

Guzman, S.M., Paz, J.O., Tagert, M.L.M., and Mercer, A., 2015. "Artificial Neural Networks and Support Vector Machines: Contrast Study for Groundwater Level Prediction", ASABE annual international meeting (p. 1). American Society of Agricultural and Biological Engineers. URL: https://elibrary.asabe.org/abstract.asp?aid=45795

Hadi, S. H., and Alwan, H. H., 2020. "The effect of Shatt Al Diwaniya on the groundwater levels of Shatt Al- Diwaniya", M.Sc. Thesis, University of Kerbala, Kerbala, Iraq.

Harini, P., Sahadevan, D.K., Das, I.C., Manikyamba, C., Durgaprasad, M., and Nandan, M.J., 2018. "Regional groundwater assessment of Krishna River basin using integrated GIS approach", Journal of the Indian Society of Remote Sensing, vol. 46, pp. 1365-1377. DOI: https://doi.org/10.1007/s12524-018-0780-4

Hasan, A. A. and Al-Taiee, T. M., 2006. "Simulation and prediction of groundwater paths and flow vectors at Mosul cit", Journal of Al- Rafidain Engineering, vol. 14, no.4, pp.73-81.

Hastie, T., Tibshirani, R., Friedman, J.H., and Friedman, J.H., 2009. "The elements of statistical learning: data mining, inference, and prediction", Vol. 2, pp. 1-758. New York: Springer. DOI: https://doi.org/10.1007/978-0-387-21606-5

Hicks, S.A., Strümke, I., Thambawita, V. et al., 2022. "On evaluation metrics for medical applications of artificial intelligence", Sci Rep 12, 5979. DOI: https://doi.org/10.1038/s41598-022-09954-

Hogeboom, R.H., van Oel, P.R., Krol, M.S., and Booij, M.J., 2015. "Modelling the influence of groundwater abstractions on the water level of Lake

Naivasha, Kenya under data-scarce conditions", Water Resources Management, vol. 29, pp. 4447-4463. DOI: https://doi.org/10.1007/s11269-015-1069-9.

Hoopes, J. A. and Harleman, D. R. F., 1967. "Waste water recharge and dispersion in porous media," ASCE Journal of the Hydraulics Division, vol. 93, no. 5, pp. 51-71.

Hussien, E.A., Thron, C., Ghaziasgar, M., Bagula, A. and Vaccari, M., 2020. "Groundwater prediction using machine-learning tools", Algorithms, vol.13, no.11, p.300. DOI : https://doi.org/10.3390/a13110300.

Hussien, H.M., Kehew, A.E., Aggour, T., Korany, E., Abotalib, A.Z., Hassanein, A., and Morsy, S., 2017. "An integrated approach for identification of potential aquifer zones in structurally controlled terrain: Wadi Qena basin, Egypt", Catena, 149, pp.73-85. DOI: https://doi.org/10.1016/j.catena.2016.08.032

Hussien, H.M., Yousif, M., and El Sheikh, A., 2020. "Investigation of groundwater occurrences in structurally controlled terrain, based on geological studies and remote sensing data: Wadi El Morra, South Sinai, Egypt", NRIAG Journal of Astronomy and Geophysics, vol. 9, no. 1, pp.512-531. Doi: https://doi.org/10.1080/20909977.2020.1788270

Ibrahim, T. M., 2022. "Quality and Hydraulic Assessment of Groundwater for Drinking and Irrigation Purpose in Al-Mahaweel District -Iraq", M.Sc. Thesis, College of Engineering, University of Babylon, Iraq.

Ifediegwu, S.I., 2022. "Assessment of groundwater potential zones using GIS and AHP techniques: A case study of the Lafia district, Nasarawa State, Nigeria", Applied Water Science, vol. 12, no. 10. DOI: https://doi.org/10.1007/s13201-021-01556-5

IMOS, 2021. Iraqi Meteorological Organization and Seismology for temperature and rainfall records, Baghdad, Iraq.

Ishaq, A. M. and Ajward, M. H., 1993. "Investigation of chloroform plumes in a porous media", Proceedings of the Symposium on Engineering Hydrology, pp. 970-976.

Jacek, G. and Maciek, L., 2004. "Modeling of complex multi-aquifer systems for groundwater resources evaluation—Swidnica study case (Poland)", Hydrogeology Journal vol.13, pp. 627–639.

Jalut, Q. H., 2021. "Groundwater Modeling of Ahmed Taher and Bakhtiari areas southwest of Khanaqin Basin in Diyala Governorate – Iraq", IOP Conference Series: Materials Science and Engineering, vol. 1076, 2nd International Scientific Conference of Engineering Sciences (ISCES 2020), Diyala, Iraq. DOI: https://doi.org/10.1088/1757-899X/1076/1/012093

Jassim, R.Z. and Al-Jiburi, B.S., 2009. "Geology of Iraqi southern desert", Iraq Bull, Geol. Min, pp. 53-76.

Jassim, S.Z., and Goff, J.C. eds., 2006. "Geology of Iraq", DOLIN, sro, (1st) distributed by Geological Society of London. URL: https://books.google.iq/books?id=LLH8aygMJFwC

Jha, M.K., Chowdary, V.M., and Chowdhury, A., 2010. "Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques", Hydrogeology Journal, vol. 18, no. 7, pp.1713-1728. DOI: https://doi.org/10.1007/s10040-010-0631-z

Jha, M.K., Chowdhury, A., Chowdary, V.M., and Peiffer, S., 2007. "Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints", Water Resources Management, vol. 21, pp.427-467. DOI: https://doi.org/10.1007/s11269-006-9024-4.

Jiao, S.R., Song, J., and Liu, B., 2020. "A review of decision tree classification algorithms for continuous variables", In Journal of Physics: Conference Series, vol. 1651, no. 1, p. 012083. IOP Publishing. DOI: https://doi.org/10.1088/1742-6596/1651/1/012083

Kaliraj, S., Chandrasekar, N., and Magesh, N.S., 2014. "Identification of potential groundwater recharge zones in Vaigai upper basin, Tamil Nadu, using GIS-based analytical hierarchical process (AHP) technique", Arabian Journal of Geosciences, vol. 7, pp.1385-1401. DOI: https://doi.org/10.1007/s12517-013-0849-x

Kanevski, M., Timonin, V., and Pozdnukhov, A., 2009. "Machine Learning for Spatial Environmental Data: Theory, Applications, and Software", CRC Press.

Kanyama, Y., Ajoodha R., Seyler,H., Makondo, N. and Tutu, H., 2020. "Application of Machine Learning Techniques In Forecasting Groundwater Levels in the Grootfontein Aquifer," 2nd International Multidisciplinary

Information Technology and Engineering Conference (IMITEC), Kimberley, South Africa, pp. 1-8, Doi: https://doi.org/10.1109/IMITEC50163.2020.9334142

Kareem, H., 2018. "Study of water resources by using 3d groundwater modelling in Al-Najaf Region, Iraq", Doctoral dissertation, Cardiff University. URL: https://orca.cardiff.ac.uk/id/eprint/111826

Karim, H. A., Al-Manmi, D. A., 2019."Integrating GIS-based and geophysical techniques for groundwater potential assessment in Halabja Said Sadiq sub-basin, Kurdistan, NE Iraq", Tikrit Journal of Pure Science, vol. 24, no. 6, pp. 81-92. DOI: http://dx.doi.org/10.25130/tjps.24.2019.112

Karim Ibrahim, Yuk Feng Huang, Ali Najah Ahmed, Chai Hoon Koo, Ahmed El-Shafie, 2022, "A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow Forecasting", Alexandria Engineering Journal, vol. 61, pp.279–303. https://doi.org/10.1016/j.aej.2021.04.100

Keith, D.J., Yoder, J.A., and Freeman, S.A., 2002. "Spatial and temporal distribution of coloured dissolved organic matter (CDOM) in Narragansett Bay, Rhode Island: implications for phytoplankton in coastal waters", Estuarine, Coastal and Shelf Science, vol. 55, no. 5, pp. 705-717. DOI: https://doi.org/10.1006/ecss.2001.0922

Kotsiantis, S.B., 2007. "Supervised machine learning: a review of classification techniques", vol. 31, pp. 249-268.

Kuhn, M., and Johnson, K., 2013. "Applied predictive modelling", vol. 26, p. 13. New York: Springer. DOI: https://doi.org/10.1007/978-1-4614-6849-3

Kumar, C.P., 2019. "An overview of commonly used groundwater modelling software", International Journal of Advanced Research in Science, Engineering and Technology Journal, vol. 6, no. 1, pp. 7854-7865. URL: http://117.252.14.250:8080/jspui/handle/123456789/7106

Kumar, T., Gautam, A.K., and Kumar, T., 2014. "Appraising the accuracy of GIS-based multi-criteria decision making technique for delineation of groundwater potential zones", Water Resources Management, vol. 28, pp.4449-4466. DOI: https://doi.org/10.1007/s11269-014-0663-6

Kumari, R., and Srivastava, S.K., 2017. "Machine learning: A review on binary classification", International Journal of Computer Applications, vol. 160, no. 7.

Kuruvilla, J. and Gunavathi, K., 2014. "Lung cancer classification using neural networks for CT images", Computer methods and programs in biomedicine, vol. 113, no. 1, pp. 202-209. DOI: https://doi.org/10.1016/j.cmpb.2013.10.011

LeCun, Y., Bengio, Y. and Hinton, G., 2015. "Deep learning", Nature, vol.52, no.7553, pp.436-444. DOI: https://doi.org/10.1038/nature14539

Lemacha, G., 2008. "Groundwater potential for upper Tumet catchment, Merge and Komosha woredas, Benishangul-Gumuz region", Guide for GIS developers, WaterAid Ethiopia and Ripple www. rippleethiopia. org.

Lentswe, G.B., and Molwalefhe, L., 2020. "Delineation of potential groundwater recharge zones using analytic hierarchy process-guided GIS in the semi-arid Motloutse watershed, eastern Botswana", Journal of Hydrology: Regional Studies, vol. 28, p.100674. DOI: https://doi.org/10.1016/j.ejrh.2020.100674

Liang, J., Qin, Z., Xiao, S., Ou, L., and Lin, X., 2019. "Efficient and secure decision tree classification for cloud-assisted online diagnosis services", IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 4, pp.1632-1644. DOI: https://doi.org/10.1109/TDSC.2019.2922958.

Loudyi, D., 2005. "2D finite volume model for groundwater flow simulations: integrating non-orthogonal grid capability into mudflow", Cardiff University, United Kingdom.

Luijendijk, E., 2022. "Transmissivity and groundwater flow exert a strong influence on drainage density. Earth Surface Dynamics", vol. 10, no. 1, pp.1-22. DOI: https://doi.org/10.5194/esurf-10-1-2022

Luo, D., Guo, Q., and Wang, X., 2003. "Simulation and prediction of underground water dynamics based on RBF neural network", Acta Geoscientia Sinica, vol. 24, no. 5; ISSU 78, pp. 475-478.

MacKay, D.J., and Mac Kay, D.J., 2003. "Information theory, inference and learning algorithms". Cambridge University Press.

Madani, A., and Niyazi, B., 2015. "Groundwater potential mapping using remote sensing techniques and weights of evidence GIS model: a case study from Wadi Yalamlam basin, Makkah Province, Western Saudi Arabia", Environmental Earth Sciences, vol. 74, pp. 5129-5142. DOI: https://doi.org/10.1007/s12665-015-4524-2

Mahesh, B., 2020. “Machine learning algorithms-a review. International Journal of Science and Research (IJSR)”, vol. 9, pp.381-386. DOI: https://doi.org/10.21275/ART20203995

Maiti, S., and Tiwari, R.K., 2010. “Neural network modeling and an uncertainty analysis in Bayesian framework: a case study from the KTB borehole site”, Journal of Geophysical Research: Solid Earth, vol. 115, no. B10. DOI: https://doi.org/10.1029/2010JB000864

Mallick, J., Khan, R.A., Ahmed, M., Alqadhi, S.D., Alsubih, M., Falqi, I., and Hasan, M.A., 2019. “Modeling groundwater potential zone in a semi-arid region of Aseer using fuzzy-AHP and geoinformation techniques”, Water, vol. 11, no. 12, p. 2656. DOI: https://doi.org/10.3390/w11122656

Marinósdóttir, H., 2019. “Applications of different machine learning methods for water level predictions”, M.Sc. thesis, Reykjavík University.

Markstrom, S.L., Niswonger, R.G., Regan, R.S., Prudic, D.E., and Barlow, P.M., 2008. “GSFLOW-Coupled Ground-water and Surface-water FLOW model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005)”, US Geological Survey techniques and methods, vol. 6, p. 240. URL: https://pubs.usgs.gov/tm/tm6d1/

Marsland, S., 2015. “Machine learning: an algorithmic perspective”, CRC press.

MathWorks, Inc., 2020. Deep Learning Toolbox: User's Guide (R2020b).

Meena, S.R., Mishra, B.K., and Tavakkoli Piralilou, S., 2019. “A hybrid spatial multi-criteria evaluation method for mapping landslide susceptible areas in kullu valley, Himalayas”, Geosciences, vol. 9, no. 4, p.156. DOI: https://doi.org/10.3390/geosciences9040156

Melgani, F., and Bruzzone, L., 2004. “Classification of hyperspectral remote sensing images with support vector machines”, IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 8, pp.1778-1790. DOI: https://doi.org/10.1109/TGRS.2004.831865

Milevski, I., and Dragićević, S., 2019. “Landslides susceptibility zonation of the territory of north macedonia using analytical hierarchy process approach”, Contributions, Section of Natural, Mathematical and Biotechnical Sciences, vol. 40, no. 1, pp.115-126. DOI: http://dx.doi.org/10.20903/csnmbs.masa.2019.40.1.136

Moayedi, H. M., Khari, M., Foong, L. K., Bahiraei, M., and Bui, D. T., 2020. "Hybridizing four wise neural-metaheuristic paradigms in predicting soil shear strength," Measurement, vol. 156, Article ID 107576.

Moghaddam, D.D., Rahmati, O., Panahi, M., Tiefenbacher, J., Darabi, H., Haghizadeh, A., Haghighi, A.T., Nalivan, O.A., and Bui, D.T., 2020. "The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers", Catena, vol. 187, p.104421. DOI: https://doi.org/10.1016/j.catena.2019.104421.

Mohr, F., Wever, M., and Hüllermeier, E., 2018. "ML-Plan: Automated machine learning via hierarchical planning", Machine Learning, vol. 107, pp.1495-1515. DOI: https://doi.org/10.1007/s10994-018-5735-z

Mondal, N.C., Adike, S., Singh, V.S., Ahmed, S., and Jayakumar, K.V., 2017. "Determining shallow aquifer vulnerability by the DRASTIC model and hydrochemistry in granitic terrain, southern India", Journal of Earth System Science, 126, pp. 1-23.

Morgan, H., Hussien, H.M., Madani, A., and Nassar, T., 2022. "Delineating Groundwater Potential Zones in Hyper-Arid Regions Using the Applications of Remote Sensing and GIS Modeling in the Eastern Desert, Egypt", Sustainability, vol. 14, no. 24, p.16942. DOI: https://doi.org/10.3390/su142416942

Mukherjee, P., Singh, C.K., and Mukherjee, S., 2012. "Delineation of groundwater potential zones in arid region of India—a remote sensing and GIS approach", Water Resources Management, 26, pp.2643-2672. DOI: https://doi.org/10.1007/s11269-012-0038-9

Naghibi, S.A., Moghaddam, D.D., Kalantar, B., Pradhan, B., and Kisi, O., 2017. "A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping", Journal of Hydrology, vol. 548, pp. 471-483. DOI: https://doi.org/10.1016/j.jhydrol.2017.03.020, 2017a.

Naghibi, S.A., Pourghasemi, H.R., and Dixon, B., 2016. "GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran", Environmental Monitoring and Assessment, vol. 188, pp.1-27. DOI: https://doi.org/10.1007/s10661-015-5049-6.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., and Williams, J. R. "Soil and water assessment tool theoretical documentation version 2009", Texas Water Resources Institute 2011.

Nguyen, P.T., Ha, D.H., Jaafari, A., Nguyen, H.D., Van Phong, T., Al-Ansari, N., Prakash, I., Le, H.V., and Pham, B.T., 2020. "Groundwater potential mapping combining artificial neural network and real AdaBoost ensemble technique: the DakNong province case-study, Vietnam", International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2473. DOI: https://doi.org/10.3390/ijerph17072473.

Nikam, S.S., 2015. "A comparative study of classification techniques in data mining algorithms", Oriental Journal of Computer Science and Technology, vol. 8, no. 1, pp.13-19.

Ouma, Y.O., and Tateishi, R., 2014. "Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment", Water, vol. 6, no. 6, pp. 1515-1545. DOI: https://doi.org/10.3390/w6061515.

Olayinka, S., 2013. "Assessing the importance of geo-hydrological data acquisition in the development of sustainable water resources framework in Nigeria," Journal of Environmental and Earth Science, vol., no.14, pp.1-10.

Oswald, S. E., and W. Kinzelbach, 2004. "Three-dimensional physical benchmark experiments to test variable-density flow models", Journal of Hydrology, Vol. 290, Issues 1–2, pp. 22-42, DOI: https://doi.org/10.1016/j.jhydrol.2003.11.037

Ozdemir, A., 2011. "Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey)", Journal of Hydrology, vol. 405, no. 1-2, pp.123-136. DOI: https://doi.org/10.1016/j.jhydrol.2011.05.015.

Pereira, A.I., Fernandes, F.P., Coelho, J.P., Teixeira, J.P., Pacheco, M.F., Alves, P., and Lopes, R.P., 2021. "International Conference on Optimization, Learning Algorithms and Applications: book of abstracts", In OL2A 2021-International Conference on Optimization, Learning Algorithms and Applications, Instituto Politécnico de Bragança.

Pham, B. T., Qi, C., Ho, L. S. et al., 2020. "A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil," Sustainability, vol. 12, no. 6, p. 2218.

Phong, T.V., Pham, B.T., Trinh, P.T., Ly, H.B., Vu, Q.H., Ho, L.S., Le, H.V., Phong, L.H., Avand, M., and Prakash, I., 2021. "Groundwater potential mapping using GIS-based hybrid artificial intelligence methods". Groundwater, vol. 59, no. 5, pp. 745-760. DOI: https://doi.org/10.1111/gwat.13094

Prasad, R.K., Mondal, N.C., Banerjee, P., Nandakumar, M.V., and Singh, V.S., 2008. "Deciphering potential groundwater zone in hard rock through the application of GIS", Environmental Geology, vol. 55, pp. 467-475. DOI: https://doi.org/10.1007/s00254-007-0992-3

Quang, H. N., Hai-Bang, L., Lanh, S. H., Nadhir, A., Hiep, V. L., Van, Q. T., Indra, P., Binh, T. P., 2021. "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil", Mathematical Problems in Engineering, vol. 2021, Article ID 4832864, p.15. DOI: https://doi.org/10.1155/2021/4832864

Rahaman, S.A., Ajeez, S.A., Aruchamy, S., and Jegankumar, R., 2015. "Prioritization of sub watershed based on morphometric characteristics using fuzzy analytical hierarchy process and geographical information system–A study of Kallar Watershed, Tamil Nadu", Aquatic Procedia, vol. 4, pp. 1322-1330. DOI: https://doi.org/10.1016/j.aqpro.2015.02.172

Rahman, A., 2008. "A GIS based DRASTIC model for assessing groundwater vulnerability in shallow aquifer in Aligarh, India", Applied Geography, vol. 28, no. 1, pp.32-53. DOI: https://doi.org/10.1016/j.apgeog.2007.07.008

Rahmati, O., Nazari Samani, A., Mahdavi, M., Pourghasemi, H.R., and Zeinivand, H., 2015. "Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS", Arabian Journal of Geosciences, 8, pp.7059-7071. DOI: https://doi.org/10.1007/s12517-014-1668-4.

Rahnama, M. B., and Zamzam, A., 2014. "Quantitative and qualitative simulation of groundwater by mathematical models in Rafsanjan aquifer using MODFLOW and MT3DMS", Arabian Journal of Geosciences, vol. 6, no. 3, pp. 901–912. DOI: https://doi.org/10.1007/S12517-011-0364-X

Rapantova, N., Tylcer, J. and Vojtek, D., 2017. "Numerical modelling as a tool for optimisation of groundwater exploitation in urban and industrial areas", Procedia Engineering, vol. 209, pp. 92-99. DOI: https://doi.org/10.1016/j.proeng.2017.11.134

Raschka, S., and Mirjalili, V., 2017. "Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2", Packt Publishing Ltd.

Ratnakumari, Y., Rai, S.N., "Thiagarajan, S., and Kumar, D., 2012. "2D Electrical resistivity imaging for delineation of deeper aquifers in a part of the Chandrabhaga River basin, Nagpur District, Maharashtra, India", Current Science, pp. 61-69. URL: https://www.jstor.org/stable/24080387

Rejani, R., Jha, M.K., Panda, S.N., and Mull, R., 2008. "Simulation modeling for efficient groundwater management in Balasore coastal basin, India",

Ripley, B.D., 2007. "Pattern recognition and neural networks", Cambridge University Press.

Roy, S., Hazra, S., Chanda, A., and Das, S., 2020. "Assessment of groundwater potential zones using multi-criteria decision-making technique: a micro-level case study from red and lateritic zone (RLZ) of West Bengal, India", Sustainable Water Resources Management, 6, pp.1-14. DOI: https://doi.org/10.1007/s40899-020-00373-z

Russo, T.A., Fisher, A.T., and Lockwood, B.S., 2015. Assessment of managed aquifer recharge site suitability using a GIS and modeling. Groundwater, vol. 53, no. 3, pp.389-400. DOI: https://doi.org/10.1111/gwat.12213

Saaty, T.L., 1980. "The analytic hierarchy process: planning, priority setting, resource allocation", McGraw: New York, NY, USA

Saha, D., Dhar, Y.R., and Vittala, S.S., 2010. "Delineation of groundwater development potential zones in parts of marginal Ganga Alluvial Plain in South Bihar, Eastern India", Environmental Monitoring and Assessment, vol. 165, pp. 179-191. DOI: https://doi.org/10.1007/s10661-009-0937-2

Saha, S., Gayen, A., Mukherjee, K., Pourghasemi, H.R., and Santosh, M., 2021. "Spatial Prediction of Groundwater Potentiality Mapping Using Machine Learning Algorithms", PREPRINT (Version 1), available at Research Square: https://doi.org/10.21203/rs.3.rs-303246/v1

Sarkar, S.K., Talukdar, S., Rahman, A., and Roy, S.K., 2022. "Groundwater potentiality mapping using ensemble machine learning algorithms for sustainable groundwater management", Frontiers in Engineering and Built

Environment, vol. 2, no. 1, pp. 43-54. DOI: https://doi.org/10.1108/FEBE-09-2021-0044

Sarker, I.H., 2022. "AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems", SN COMPUT. SCI, vol. 3, no. 158. DOI: https://doi.org/10.1007/s42979-022-01043-x

Seeyan S., & Merkel, B., 2015. "Groundwater Modeling of Harrir plain and Mirawa valley in Shaqlawa-Harrir basins, Kurdistan Region, Iraq", FOG-Freiberg Online Geoscience.

Shahin, M.A., Jaksa, M.B., and Maier, H.R., 2008. "State of the art of artificial neural networks in geotechnical engineering", Electronic Journal of Geotechnical Engineering, vol. 8, no. 1, pp. 1-26.

Singh, A., Yadav, A., and Rana, A., 2013. "K-means with Three different Distance Metrics", International Journal of Computer Applications, vol. 67, no. 10.

Sissakian, V.K. and Fouad, S.F., 2015. "Geological map of Iraq, scale 1: 1000 000, 2012", Iraqi Bulletin of Geology and Mining, vol. 11, no. 1, pp. 9-16.

Sissakian, V.K., Youkhana, R.Y., and Zainal, E.M., 1994. "The geology of Al-Birreet Quadrangle. Sheet NH–38–1, Scale 1: 250 000", GEOSURV. Baghdad, Iraq.

Sule, B.F., and Ayenigba, S.E., 2017. "Application of GMS-MODFLOW to investigate groundwater development potential in River Meme catchment, Kogi State, Nigeria", International Journal of Sciences, vol. 6, no. 09, pp.39-51. DOI: https://doi.org/10.18483/ijSci.1332.

Suresh, A., Udendhran, R., and Balamurgan, M., 2020. "Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers", Soft Computing, vol. 24, no. 11, pp. 7947-7953. DOI: https://doi.org/10.1007/s00500-019-04066-4.

Tamiru, H., Wagari, M., and Tadese, B., 2022. "An integrated Artificial Intelligence and GIS spatial analyst tools for Delineation of Groundwater Potential Zones in complex terrain: Fincha Catchment, Abay Basi, Ethiopia", Air, Soil and Water Research, vol. 15, p.11786221211045972. DOI: https://doi.org/10.1177/11786221211045972

Tiwari, K., Goyal, R., and Sarkar, A., 2017. "GIS-based spatial distribution of groundwater quality and regional suitability evaluation for drinking

water," Environmental Processes, vol. 4, pp.645-662. DOI: https://doi.org/10.1007/s40710-017-0257-4

Todd, D.K., 2007. "Groundwater hydrology," Jhon Wiley and Sons, Third Reprint. Inc. India.

Trunfio, T.A., Scala, A., Vecchia, A.D., Marra, A., and Borrelli, A., 2021. "Multiple Regression Model to Predict Length of Hospital Stay for Patients Undergoing Femur Fracture Surgery at "San Giovanni Di Dio e Ruggi d'Aragona" University Hospital," In 8th European Medical and Biological Engineering Conference: Proceedings of the EMBEC 2020, November 29–December 3, 2020, Portorož, Slovenia, pp. 840-847, Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-64610-3_94

UN-Water, 2021 "Water Scarcity," [Online], Available: https://www.unwater.org/water-facts/water-scarcity.

Vapnik, V., 2013. "The Nature of Statistical Learning Theory," Springer Science and Business Media, Berlin, Heidelberg, Germany.

Vihinen, M., 2012. "How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis", In BMC Genomics, Vol. 13, no. 4, pp. 1-10. BioMed Central. DOI: https://doi.org/10.1186/1471-2164-13-S4-S2

Waikar, M.L. and Nilawar, A.P., 2014. "Identification of groundwater potential zone using remote sensing and GIS technique," International Journal of Innovative Research in Science, Engineering and Technology, vol. 3, no. 5, pp.12163-12174.

Wang, S., Phillips, P., Liu, A. and Du, S.,2017. "Tea category identification using computer vision and generalized eigenvalue proximal SVM. Fundam", Inform, vol.151, pp.325–339.

Wang, W.C., Chau, K.W., Cheng, C.T., and Qiu, L., 2009. "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," Journal of Hydrology, vol. 374, no. 3-4, pp. 294-306. DOI: https://doi.org/10.1016/j.jhydrol.2009.06.019.

Węglarczyk, S., 2018. "Kernel density estimation and its application", In ITM Web of Conferences, vol. 23, p. 00037. EDP Sciences. DOI: https://doi.org/10.1051/itmconf/20182300037

Wibowo, F. W. and Wihayati W., 2021. "Multi-classification of Alcohols using Quartz Crystal Microbalance Sensors based-on Artificial Neural Network

Single Layer Perceptron", International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), Yogyakarta, Indonesia, pp. 37-41, DOI: https://doi.org/10.1109/ICAICST53116.2021.9497839.

Witten, I.H., and Frank, E., 2002. "Data mining: practical machine learning tools and techniques with Java implementations", Acm Sigmod Record, 31(1), pp.76-77. DOI: https://doi.org/10.1145/507338.507355

Xu, X., Yan, Z. and Xu, S., 2015. "Estimating wind speed probability distribution by diffusion-based kernel density method". Electric Power Systems Research, vol. 121, pp. 28-37. DOI: https://doi.org/10.1016/j.epsr.2014.11.029

Yadav, H., Keshari, A. K., 2017." Assessment of Groundwater Pollution from Industries in Agra Region", International Journal of Renewable Energy Research-IJRER, Vol.7, no.4. DOI: https://doi.org/10.20508/ijrer.v7i4.6417.g7254

Yang, F.J., 2019. "An extended idea about decision trees", In 2019 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 349-354. IEEE. DOI: https://doi.org/10.1109/CSCI49370.2019.00068

Ye, J., 2015. "Using machine learning for exploratory data analysis and predictive modelling", Master's thesis, University of Stavanger, Norway.

Yehia, T.M., 2020. "Quality and Hydraulic Assessment of Groundwater for Drinking and Irrigation Purpose in Al-Mahaweel District -Iraq", Master's thesis, University of Babylon, Iraq.

Yin, H., Shi, Y., Niu, H., Xie, D., Wei, J., Lefticariu, L., and Xu, S., 2018. "A GIS-based model of potential groundwater yield zonation for a sandstone aquifer in the Juye Coalfield, Shangdong, China," Journal of Hydrology, vol. 557, pp. 434-447. DOI: https://doi.org/10.1016/j.jhydrol.2017.12.043.

Zheng, C. and Wang, P.P., 1999. "MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems", Documentation and User's Guide, Contract Report SERDP-99-1, US Army Corps of Engineers-Engineer Research and Development Center, 220.

Zghibi, A., Mirchi, A., Msaddek, M.H., Merzougui, A., Zouhri, L., Taupin, J.D., Chekirbane, A., Chenini, I., and Tarhouni, J., 2020. "Using analytical

hierarchy process and multi-influencing factors to map groundwater recharge zones in a semi-arid Mediterranean coastal aquifer", Water, vol. 12, no. 9, p.2525. DOI: https://doi.org/10.3390/w12092525.

Zhou, T., Wang, F., and Yang, Z., 2017. "Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction," Water, vol. 9, no. 10, p. 781. DOI: https://doi.org/10.3390/w9100781

Table (A-1): Characteristics of the studied wells for developing the machine learning models (GCGW, 2021).

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
1	W1	44° 07' 47.7594" E	32° 09' 51.2866" N	206	82.466	2
2	W2	44° 11' 31.4411" E	32° 09' 52.7699" N	230	64.631	4
3	W3	44° 07' 49.4971" E	32° 06' 48.2303" N	260	89.116	3
4	W4	44° 00' 25.0362" E	32° 02' 40.8720" N	230	41.328	2
5	W5	44° 09' 06.2708" E	32° 02' 44.6573" N	230	16.000	3
6	W6	43° 53' 00.5076" E	31° 59' 34.1077" N	230	90.356	2
7	W7	44° 04' 10.2699" E	31° 59' 39.5105" N	208	34.991	3
8	W8	44° 14' 07.1707" E	31° 56' 40.4227" N	170	16.000	2
9	W9	43° 54' 17.8206" E	31° 55' 30.6910" N	230	88.042	3
10	W10	44° 04' 12.7304" E	31° 55' 35.4326" N	220	47.271	3
11	W11	43° 45' 38.1271" E	31° 54' 24.8939" N	195	132.958	3
12	W12	43° 49' 21.9352" E	31° 53' 25.9999" N	260	122.694	5
13	W13	44° 07' 57.5484" E	31° 52' 33.9479" N	252	47.239	3
14	W14	43° 43' 13.6601" E	31° 49' 18.3535" N	252	166.072	8
15	W15	44° 06' 46.7248" E	31° 46' 27.3059" N	254	86.949	3
16	W16	44° 19' 09.1982" E	31° 46' 31.8758" N	210	36.000	3
17	W17	44° 06' 50.2102" E	31° 40' 21.1710" N	200	108.276	2
18	W18	43° 29' 56.3521" E	31° 28' 49.2213" N	145	289.800	8
19	W19	43° 38' 35.2204" E	31° 27' 53.9895" N	292	251.147	8
20	W20	44° 19' 20.6934" E	31° 20' 05.1706" N	180	140.000	11
21	W21	44° 00' 56.0305" E	31° 13' 51.8394" N	250	215.385	8
22	W22	44° 23' 06.3581" E	31° 09' 55.9798" N	148	150.808	7
23	W23	44° 59' 59.5847" E	31° 06' 58.1599" N	142	56.002	6
24	W24	45° 08' 35.7405" E	31° 06' 57.8732" N	145	38.030	9
25	W25	43° 59' 48.0070" E	31° 04' 42.0359" N	200	258.000	8
26	W26	44° 39' 06.5169" E	31° 04' 54.3960" N	195	138.839	11
27	W27	45° 04' 54.4265" E	31° 04' 55.9950" N	180	64.067	6
28	W28	44° 13' 19.1939" E	31° 03' 46.6057" N	240	208.182	13
29	W29	45° 14' 43.6367" E	31° 01' 52.1397" N	230	42.226	17
30	W30	44° 29' 18.1610" E	31° 00' 48.2940" N	220	174.976	11
31	W31	44° 20' 43.0047" E	30° 59' 44.9323" N	220	197.908	11
32	W32	45° 12' 16.0350" E	30° 59' 50.3241" N	175	63.494	7
33	W33	45° 22' 04.7183" E	30° 57' 46.9469" N	176	36.000	19

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
34	W34	44° 52' 37.9490" E	30° 56' 47.5859" N	180	131.917	13
35	W35	45° 07' 21.2218" E	30° 56' 47.5867" N	180	97.242	12
36	W36	45° 15' 56.4609" E	30° 56' 46.8129" N	150	69.933	13
37	W37	44° 48' 57.2486" E	30° 55' 46.2868" N	180	143.947	12
38	W38	44° 56' 18.8061" E	30° 55' 46.7060" N	157	125.048	14
39	W39	45° 01' 13.1786" E	30° 55' 46.7528" N	262	122.000	12
40	W40	45° 29' 25.4823" E	30° 54' 42.3744" N	168	22.651	11
41	W41	45° 11' 01.6890" E	30° 53' 44.2135" N	162	89.000	10
42	W42	43° 07' 15.3229" E	30° 50' 53.5171" N	180	317.128	7
43	W43	44° 50' 11.2564" E	30° 51' 42.2360" N	230	158.761	11
44	W44	45° 15' 55.6183" E	30° 51' 41.6282" N	240	93.297	11
45	W45	45° 01' 13.1142" E	30° 50' 41.5639" N	232	138.696	17
46	W46	45° 19' 36.0316" E	30° 50' 40.0865" N	225	93.194	11
47	W47	44° 35' 29.2929" E	30° 49' 38.2127" N	262	194.951	13
48	W48	43° 40' 22.4627" E	30° 48' 15.0313" N	265	289.638	17
49	W49	44° 45' 17.5591" E	30° 48' 38.6585" N	250	173.912	11
50	W50	44° 55' 05.6277" E	30° 47' 38.3616" N	180	152.899	9
51	W51	45° 07' 20.4461" E	30° 46' 37.2080" N	180	125.516	12
52	W52	45° 24' 29.1051" E	30° 46' 35.1042" N	180	74.501	8
53	W53	45° 34' 16.5310" E	30° 45' 31.8486" N	160	56.071	14
54	W54	44° 48' 58.6447" E	30° 43' 33.8306" N	265	176.824	11
55	W55	44° 18' 23.2042" E	30° 42' 26.5911" N	220	252.267	14
56	W56	44° 41' 38.2139" E	30° 42' 31.9607" N	180	200.542	11
57	W57	44° 46' 32.0523" E	30° 41' 31.5217" N	220	198.763	9
58	W58	45° 35' 28.1277" E	30° 40' 26.3388" N	220	74.355	7
59	W59	44° 44' 05.7291" E	30° 38' 28.1273" N	293	206.050	12
60	W60	45° 50' 07.7858" E	30° 38' 19.4357" N	250	30.000	13
61	W61	44° 35' 32.3784" E	30° 37' 25.7598" N	250	237.670	8
62	W62	44° 41' 39.5588" E	30° 35' 24.6868" N	220	229.972	10
63	W63	44° 35' 33.4000" E	30° 33' 21.6037" N	210	254.541	12
64	W64	44° 53' 53.0367" E	30° 33' 23.7541" N	230	190.108	12
65	W65	45° 31' 45.6164" E	30° 33' 20.0247" N	244	125.676	11
66	W66	44° 29' 27.5009" E	30° 31' 18.2365" N	255	268.134	11
67	W67	45° 53' 42.8287" E	30° 29' 08.6695" N	234	75.047	9
68	W68	44° 36' 47.9164" E	30° 28' 16.6278" N	276	258.149	18
69	W69	45° 07' 18.9143" E	30° 26' 16.4026" N	289	182.459	12
70	W70	45° 12' 11.5445" E	30° 24' 13.9542" N	234	196.659	8
71	W71	44° 35' 35.9387" E	30° 23' 11.2023" N	250	283.313	13

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
72	W72	44° 27' 04.3606" E	30° 21' 07.2491" N	262	311.410	19
73	W73	45° 23' 09.0975" E	30° 19' 07.2612" N	220	193.549	13
74	W74	44° 40' 30.4774" E	30° 14' 02.6426" N	232	304.807	13
75	W75	44° 44' 10.0094" E	30° 12' 01.0483" N	280	305.643	10
76	W76	45° 23' 07.4280" E	30° 11' 59.9653" N	276	204.753	12
77	W77	45° 38' 56.1614" E	30° 09' 54.1386" N	245	190.782	13
78	W78	46° 02' 00.8032" E	30° 05' 41.1017" N	234	160.382	11
79	W79	46° 04' 26.7121" E	30° 05' 39.9315" N	300	148.511	11
80	W80	46° 20' 15.0568" E	30° 05' 31.2282" N	233	153.346	19
81	W81	46° 26' 18.8887" E	30° 04' 26.3488" N	276	125.219	19
82	W82	46° 05' 38.3222" E	30° 03' 37.2611" N	223	154.705	11
83	W83	45° 06' 04.1282" E	30° 00' 50.3683" N	280	266.022	10
84	W84	45° 21' 50.5832" E	29° 54' 42.4209" N	268	237.289	13
85	W85	45° 55' 47.4845" E	29° 50' 28.2563" N	220	216.000	19
86	W86	45° 37' 34.3257" E	29° 46' 30.5125" N	200	216.973	12
87	W87	44° 47' 52.3627" E	29° 45' 34.2455" N	260	331.184	7
88	W88	44° 33' 22.9221" E	29° 33' 19.5439" N	250	384.000	14
89	W89	45° 20' 31.5408" E	29° 24' 11.1720" N	274	324.000	9
90	W90	44° 02' 53.3259" E	32° 03' 43.0324" N	270	26.456	2
91	W91	44° 12' 49.1431" E	32° 03' 47.1183" N	270	43.010	3
92	W92	44° 20' 16.0263" E	32° 03' 49.6721" N	300	41.000	5
93	W93	44° 05' 22.8813" E	32° 02' 43.1079" N	280	23.480	2
94	W94	44° 07' 51.8071" E	32° 02' 44.1530" N	212	18.649	3
95	W95	44° 00' 25.6948" E	32° 01' 39.8543" N	240	39.615	3
96	W96	44° 15' 19.0909" E	32° 01' 45.9751" N	300	12.930	5
97	W97	44° 22' 45.8145" E	32° 01' 48.3799" N	240	37.886	6
98	W98	44° 06' 38.5235" E	32° 00' 41.5975" N	200	25.019	3
99	W99	44° 09' 07.3955" E	32° 00' 42.6170" N	200	21.674	3
100	W100	44° 19' 02.9033" E	32° 00' 46.2095" N	280	55.867	6
101	W101	44° 04' 10.2699" E	31° 59' 39.5105" N	300	34.991	3
102	W102	44° 10' 22.3803" E	31° 59' 42.0877" N	280	22.982	2
103	W103	44° 21' 32.2093" E	31° 59' 45.9631" N	234	47.000	6
104	W104	44° 04' 11.5012" E	31° 57' 37.4718" N	300	41.353	3
105	W105	44° 07' 54.6838" E	31° 57' 39.0525" N	245	32.044	3
106	W106	44° 11' 37.8711" E	31° 57' 40.5242" N	260	16.063	3
107	W107	44° 24' 01.8584" E	31° 57' 44.6429" N	280	40.852	5
108	W108	44° 11' 38.4040" E	31° 56' 39.5027" N	346	19.195	2
109	W109	44° 25' 16.6420" E	31° 56' 43.9637" N	282	37.329	6

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
110	W110	44° 06' 41.4636" E	31° 55' 36.4971" N	211	40.910	3
111	W111	44° 09' 10.1990" E	31° 55' 37.5132" N	225	31.926	3
112	W112	44° 15' 22.0462" E	31° 55' 39.8419" N	280	14.371	2
113	W113	44° 21' 33.9046" E	31° 55' 41.8683" N	300	17.158	6
114	W114	44° 11' 39.4685" E	31° 54' 37.4591" N	250	27.093	2
115	W115	44° 20' 19.9686" E	31° 54' 40.4634" N	287	16.000	5
116	W116	44° 24' 03.0463" E	31° 54' 41.5696" N	245	18.982	6
117	W117	44° 25' 17.4063" E	31° 54' 41.9142" N	298	17.447	5
118	W118	44° 06' 42.6362" E	31° 53' 34.4557" N	320	47.410	2
119	W119	44° 12' 54.3422" E	31° 53' 36.9022" N	300	25.612	3
120	W120	44° 14' 08.6849" E	31° 53' 37.3552" N	195	23.269	2
121	W121	44° 21' 34.7501" E	31° 53' 39.8199" N	190	18.470	5
122	W122	44° 07' 57.5484" E	31° 52' 33.9479" N	180	47.239	3
123	W123	44° 12' 54.8597" E	31° 52' 35.8796" N	185	29.617	2
124	W124	44° 16' 37.8483" E	31° 52' 37.2017" N	145	18.829	3
125	W125	44° 24' 03.8366" E	31° 52' 39.5200" N	195	20.777	5
126	W126	44° 11' 41.0618" E	31° 51' 34.3924" N	194	39.876	3
127	W127	44° 22' 49.9125" E	31° 51' 38.1390" N	190	19.004	5
128	W128	44° 27' 47.1897" E	31° 51' 39.4905" N	192	22.910	5
129	W129	44° 05' 30.0926" E	31° 50' 30.8680" N	238	57.477	2
130	W130	44° 12' 55.8934" E	31° 50' 33.8341" N	200	44.309	2
131	W131	44° 09' 13.5477" E	31° 49' 31.3832" N	250	54.744	3
132	W132	44° 15' 24.9867" E	31° 49' 33.7028" N	209	35.443	2
133	W133	44° 19' 07.8555" E	31° 49' 34.9500" N	157.5	28.301	3
134	W134	44° 25' 19.3116" E	31° 49' 36.7877" N	248	19.946	4
135	W135	44° 19' 08.3034" E	31° 48' 33.9254" N	246	30.000	2
136	W136	44° 12' 57.4408" E	31° 47' 30.7646" N	245	48.151	3
137	W137	44° 26' 34.3366" E	31° 47' 35.0670" N	245	20.457	6
138	W138	44° 09' 15.2157" E	31° 46' 28.3160" N	286	81.988	3
139	W139	44° 15' 26.4513" E	31° 46' 30.6310" N	210	49.104	3
140	W140	44° 16' 40.6999" E	31° 46' 31.0579" N	272	43.902	2
141	W141	44° 24' 06.1996" E	31° 46' 33.3671" N	200	21.000	2
142	W142	44° 21' 38.1178" E	31° 45' 31.6199" N	230	30.485	3
143	W143	44° 14' 13.2047" E	31° 44' 28.1440" N	280	57.731	3
144	W144	44° 20' 24.7467" E	31° 43' 29.1901" N	287	31.943	3
145	W145	44° 19' 10.9831" E	31° 42' 27.7745" N	245	44.295	3
146	W146	44° 24' 07.7683" E	31° 42' 29.2619" N	257	19.939	3
147	W147	44° 16' 43.0652" E	31° 41' 25.9336" N	240	56.332	3

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
148	W148	44° 19' 11.4284" E	31° 41' 26.7487" N	260	45.333	3
149	W149	44° 21' 39.7932" E	31° 41' 27.5159" N	257	30.470	3
150	W150	44° 24' 08.1596" E	31° 41' 28.2352" N	215	22.295	2
151	W151	44° 08' 04.3744" E	31° 40' 21.6800" N	257	107.090	2
152	W152	44° 17' 59.5347" E	31° 36' 21.2174" N	235	80.937	2
153	W153	44° 11' 49.4921" E	31° 35' 18.0120" N	276	101.389	4
154	W154	44° 13' 06.1353" E	31° 30' 13.3426" N	300	120.321	5
155	W155	43° 38' 37.8574" E	31° 24' 50.9434" N	220	263.000	8
156	W156	44° 18' 04.9871" E	31° 24' 08.8892" N	220	119.983	8
157	W157	44° 36' 35.5980" E	31° 19' 08.4420" N	220	72.211	6
158	W158	44° 21' 50.1374" E	31° 16' 01.8069" N	293	131.530	7
159	W159	44° 05' 50.7139" E	31° 14' 55.0326" N	250	196.000	9
160	W160	44° 03' 24.2673" E	31° 12' 51.9198" N	250	208.161	7
161	W161	44° 09' 33.2878" E	31° 12' 54.4789" N	220	201.000	9
162	W162	44° 40' 18.7555" E	31° 12' 01.8298" N	210	98.825	8
163	W163	44° 48' 55.4834" E	31° 11' 01.8245" N	230	70.227	5
164	W164	44° 56' 18.2170" E	31° 11' 02.2480" N	244	55.000	6
165	W165	45° 07' 22.1614" E	31° 09' 00.0200" N	255	25.758	7
166	W166	45° 24' 34.5638" E	31° 07' 56.8513" N	234	10.066	11
167	W167	44° 52' 37.1653" E	31° 06' 57.9486" N	276	81.968	13
168	W168	43° 31' 32.2339" E	31° 05' 26.7696" N	289	316.023	9
169	W169	45° 04' 54.4788" E	31° 05' 57.0308" N	234	54.191	3
170	W170	45° 15' 57.9865" E	31° 05' 56.1354" N	250	21.019	8
171	W171	44° 46' 28.7725" E	31° 04' 55.3796" N	262	110.876	19
172	W172	44° 15' 46.5723" E	31° 03' 47.4712" N	220	210.000	11
173	W173	45° 29' 27.9809" E	31° 02' 50.6516" N	232	10.000	12
174	W174	43° 59' 50.5667" E	31° 00' 37.9283" N	280	260.831	12
175	W175	45° 11' 02.5076" E	31° 00' 51.4712" N	276	62.043	9
176	W176	45° 22' 05.4213" E	31° 00' 50.0536" N	245	20.249	11
177	W177	45° 23' 18.8299" E	30° 59' 48.8024" N	234	24.542	18
178	W178	45° 23' 18.5826" E	30° 58' 47.7670" N	300	29.616	19
179	W179	45° 29' 26.7295" E	30° 58' 46.5142" N	233	20.107	10
180	W180	45° 25' 45.5689" E	30° 57' 46.2656" N	276	24.887	12
181	W181	45° 25' 45.2960" E	30° 56' 45.2302" N	223	32.147	12
182	W182	45° 28' 12.2038" E	30° 55' 43.6829" N	280	21.673	16
183	W183	45° 02' 26.7459" E	30° 54' 45.6980" N	268	117.510	12
184	W184	45° 29' 25.4823" E	30° 54' 42.3744" N	220	22.651	11
185	W185	45° 03' 40.2872" E	30° 53' 44.6314" N	200	121.065	10

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
186	W186	45° 15' 55.7866" E	30° 52' 42.6654" N	260	78.961	12
187	W187	45° 31' 51.9614" E	30° 52' 39.7231" N	250	26.320	13
188	W188	45° 11' 01.1073" E	30° 48' 39.0246" N	274	128.000	11
189	W189	45° 33' 04.1155" E	30° 48' 35.2741" N	270	27.559	7
190	W190	45° 37' 58.1057" E	30° 48' 33.9312" N	270	24.859	7
191	W191	45° 25' 42.8500" E	30° 47' 35.9042" N	300	72.543	7
192	W192	45° 42' 50.7368" E	30° 45' 29.3011" N	280	22.688	13
193	W193	44° 35' 30.5824" E	30° 44' 33.0267" N	212	196.080	13
194	W194	44° 17' 10.2352" E	30° 41' 25.1583" N	240	262.480	12
195	W195	45° 45' 15.7476" E	30° 41' 24.3333" N	300	31.527	10
196	W196	44° 37' 58.4050" E	30° 40' 29.3129" N	240	216.887	12
197	W197	46° 01' 09.3226" E	30° 40' 16.7875" N	200	6.655	12
198	W198	46° 08' 28.9039" E	30° 39' 12.0984" N	200	6.023	8
199	W199	44° 41' 39.1754" E	30° 37' 26.7658" N	280	212.990	11
200	W200	45° 53' 47.3133" E	30° 37' 16.9369" N	300	27.057	7
201	W201	44° 26' 59.2180" E	30° 36' 22.8310" N	280	256.062	14
202	W202	44° 34' 19.8245" E	30° 34' 22.4077" N	234	242.279	8
203	W203	44° 36' 46.4656" E	30° 34' 22.8667" N	300	241.000	13
204	W204	44° 35' 33.9094" E	30° 31' 19.5247" N	245	252.216	12
205	W205	44° 34' 20.8943" E	30° 30' 18.2503" N	260	245.290	19
206	W206	44° 31' 54.6484" E	30° 29' 16.7074" N	280	202.831	12
207	W207	44° 36' 47.6751" E	30° 29' 17.6680" N	346	219.768	18
208	W208	44° 39' 14.1898" E	30° 29' 18.0797" N	282	257.000	20
209	W209	44° 40' 27.6506" E	30° 28' 17.2278" N	211	204.423	11
210	W210	44° 35' 35.9387" E	30° 23' 11.2023" N	225	283.313	12
211	W211	45° 09' 45.0527" E	30° 23' 13.1171" N	280	187.369	10
212	W212	44° 31' 56.6876" E	30° 22' 09.4272" N	300	289.735	17
213	W213	44° 39' 15.6972" E	30° 22' 10.7930" N	250	229.097	11
214	W214	44° 42' 55.2048" E	30° 22' 11.3223" N	287	256.873	13
215	W215	45° 13' 24.4619" E	30° 22' 11.7506" N	245	192.963	8
216	W216	44° 46' 35.1302" E	30° 19' 08.6220" N	298	267.158	13
217	W217	44° 40' 29.6732" E	30° 18' 06.8130" N	320	279.773	11
218	W218	45° 53' 32.8493" E	30° 10' 50.0302" N	300	164.901	10
219	W219	46° 06' 52.6186" E	30° 05' 38.7162" N	206	154.368	9
220	W220	44° 53' 54.8641" E	30° 03' 53.5046" N	230	288.129	14
221	W221	45° 18' 12.6429" E	29° 57' 46.1112" N	260	226.523	13
222	W222	45° 36' 21.5957" E	29° 46' 30.8513" N	230	204.876	14
223	W223	44° 51' 30.8729" E	29° 41' 30.3352" N	230	336.720	9

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
224	W224	44° 44' 14.9906" E	29° 40' 28.6220" N	230	349.140	8
225	W225	44° 50' 19.0750" E	29° 32' 20.8075" N	208	337.919	13
226	W226	44° 56' 21.9701" E	29° 30' 19.0096" N	170	318.012	12
227	W227	45° 09' 39.8176" E	29° 29' 17.6575" N	230	320.356	12
228	W228	44° 57' 34.7038" E	29° 22' 10.6271" N	220	322.033	14
229	W229	45° 01' 30.3000" E	30° 55' 39.2000" N	120	114.501	13
230	W230	44° 40' 31.0000" E	30° 43' 19.0000" N	100	202.852	13
231	W231	45° 16' 50.5000" E	31° 13' 02.3000" N	160	14.253	7
232	W232	46° 06' 13.1000" E	30° 06' 32.8000" N	162	142.047	8
233	W233	45° 05' 39.9000" E	30° 01' 51.8000" N	150	266.848	10
234	W234	44° 53' 20.6000" E	30° 33' 47.1000" N	150	189.065	11
235	W235	44° 48' 39.2000" E	30° 43' 40.4000" N	174	171.742	8
236	W236	44° 45' 48.2000" E	30° 41' 46.4000" N	154	196.614	8
237	W237	44° 34' 32.9000" E	30° 33' 49.1000" N	120	248.908	7
238	W238	44° 36' 20.4000" E	30° 28' 39.5000" N	105	217.741	19
239	W239	44° 43' 41.1000" E	30° 38' 50.0000" N	150	205.732	10
240	W240	44° 38' 09.1000" E	31° 05' 25.5000" N	170	143.466	9
241	W241	44° 41' 03.3000" E	30° 58' 19.4000" N	200	153.821	13
242	W242	44° 26' 19.2000" E	30° 22' 27.4000" N	152	297.272	11
243	W243	44° 51' 57.2000" E	30° 56' 47.5000" N	180	128.165	13
244	W244	44° 40' 12.0000" E	30° 36' 27.0000" N	116	223.186	12
245	W245	44° 54' 56.9000" E	30° 48' 57.5000" N	157	151.610	7
246	W246	44° 29' 29.9000" E	30° 31' 37.8000" N	115	263.408	12
247	W247	44° 34' 50.7000" E	30° 23' 23.4000" N	100	224.000	14
248	W248	45° 07' 08.4000" E	30° 47' 35.2000" N	180	127.049	11
249	W249	45° 00' 45.3000" E	30° 50' 44.4000" N	200	138.383	11
250	W250	44° 44' 13.5000" E	30° 48' 15.9000" N	134	167.193	12
251	W251	44° 49' 51.4000" E	30° 52' 27.9000" N	200	154.240	11
252	W252	46° 20' 08.8427" E	30° 06' 33.3467" N	158	154.039	19
253	W253	46° 26' 22.4000" E	30° 06' 29.4000" N	162	124.683	16
254	W254	45° 05' 09.5000" E	30° 55' 17.8000" N	135	108.000	11
255	W255	45° 10' 23.0000" E	30° 53' 58.3000" N	145	94.770	13
256	W256	45° 16' 06.0000" E	30° 51' 25.9000" N	176	102.702	10
257	W257	44° 49' 06.5000" E	30° 48' 16.9000" N	160	160.819	9
258	W258	44° 35' 23.1000" E	30° 50' 12.1000" N	230	192.887	10
259	W259	45° 06' 55.2000" E	30° 55' 45.6000" N	180	105.477	10
260	W260	44° 41' 31.7000" E	30° 59' 51.8000" N	246	140.924	13
261	W261	44° 55' 35.9000" E	30° 56' 27.8000" N	200	118.678	12

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
262	W262	44° 55' 34.6000" E	30° 56' 55.6000" N	200	116.611	12
263	W263	44° 14' 08.0016" E	32° 09' 57.0996" N	180	54.452	7
264	W264	44° 13' 02.8992" E	32° 09' 58.6008" N	180	58.519	9
265	W265	44° 09' 07.5996" E	32° 07' 47.2008" N	200	81.768	2
266	W266	43° 40' 01.5600" E	30° 48' 35.6400" N	85	292.095	19
267	W267	44° 19' 59.9988" E	31° 00' 00.0000" N	150	196.129	10
268	W268	44° 16' 51.8016" E	32° 02' 27.3012" N	160	61.615	6
269	W269	44° 28' 34.5000" E	31° 01' 31.0008" N	150	174.314	11
270	W270	44° 23' 54.0492" E	31° 57' 32.2992" N	160	43.237	5
271	W271	44° 17' 35.5308" E	32° 06' 41.5692" N	160	42.113	4
272	W272	44° 22' 41.5614" E	31° 10' 28.8755" N	90	148.000	7
273	W273	44° 10' 51.3012" E	31° 54' 01.5984" N	75	31.120	3
274	W274	44° 18' 35.7984" E	31° 47' 09.8988" N	160	36.022	3
275	W275	44° 03' 42.6996" E	32° 03' 11.4984" N	150	26.426	2
276	W276	44° 12' 33.9012" E	31° 57' 06.8004" N	150	16.541	2
277	W277	44° 17' 25.5012" E	31° 43' 01.7004" N	110	47.000	2
278	W278	45° 19' 59.9988" E	29° 25' 00.0012" N	200	320.933	12
279	W279	45° 37' 00.0012" E	29° 46' 59.9988" N	120	212.014	10
280	W280	43° 04' 59.9988" E	30° 52' 59.9988" N	147	321.359	8
281	W281	46° 07' 00.0012" E	30° 07' 00.0012" N	158	143.346	11
282	W282	45° 22' 59.9988" E	30° 19' 59.9988" N	132	187.910	7
283	W283	43° 43' 48.0000" E	31° 56' 24.0000" N	130	132.000	2
284	W284	43° 49' 12.0000" E	31° 53' 24.0000" N	115	145.000	5
285	W285	43° 53' 45.6000" E	31° 56' 24.0000" N	80	96.000	3
286	W286	44° 04' 12.0000" E	31° 55' 48.0000" N	85	47.000	2
287	W287	43° 38' 11.0004" E	31° 27' 30.9996" N	100	260.000	9
288	W288	44° 01' 12.0000" E	31° 14' 24.0000" N	150	209.000	9
289	W289	44° 44' 24.0000" E	31° 22' 48.0000" N	180	35.000	2
290	W290	44° 12' 36.0000" E	31° 04' 48.0000" N	120	204.000	13
291	W291	43° 43' 12.0000" E	31° 49' 12.0000" N	80	294.000	8
292	W292	44° 17' 24.0000" E	30° 42' 36.0000" N	80	255.000	10
293	W293	45° 03' 00.0000" E	31° 09' 00.0000" N	85	45.000	5
294	W294	44° 34' 48.0000" E	30° 37' 48.0000" N	83	240.000	7
295	W295	45° 08' 24.0000" E	31° 07' 12.0000" N	122	41.000	7
296	W296	45° 16' 48.0000" E	31° 05' 24.0000" N	280	27.000	7
297	W297	45° 15' 36.0000" E	31° 04' 12.0000" N	190	33.000	12
298	W298	45° 21' 36.0000" E	30° 58' 12.0000" N	200	45.000	18
299	W299	45° 32' 24.0000" E	30° 55' 12.0000" N	220	26.000	11

Appendix A ***Table: (A-1) Continued….***

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
300	W300	45° 19' 48.0000" E	30° 51' 00.0000" N	240	99.000	12
301	W301	44° 40' 00.9984" E	30° 14' 00.9996" N	230	303.000	10
302	W302	45° 24' 36.0000" E	30° 47' 45.6000" N	245	74.000	8
303	W303	44° 43' 12.0000" E	30° 12' 36.0000" N	260	308.000	13
304	W304	45° 36' 00.0000" E	30° 48' 00.0000" N	274	32.000	13
305	W305	45° 35' 24.0000" E	30° 40' 48.0000" N	234	73.000	8
306	W306	45° 49' 48.0000" E	30° 39' 36.0000" N	80	26.000	8
307	W307	45° 31' 12.0000" E	30° 33' 36.0000" N	96	125.000	8
308	W308	45° 11' 49.2000" E	30° 24' 36.0000" N	180	194.000	8
309	W309	44° 32' 24.0000" E	29° 34' 12.0000" N	227	380.000	10
310	W310	44° 47' 24.0000" E	29° 46' 12.0000" N	238	331.000	13
311	W311	45° 22' 12.0000" E	30° 12' 36.0000" N	200	204.000	17
312	W312	45° 21' 57.9996" E	29° 55' 00.9984" N	260	238.000	12
313	W313	45° 38' 58.9992" E	30° 10' 00.9984" N	240	191.000	8
314	W314	45° 07' 12.0000" E	30° 26' 24.0000" N	280	180.000	12
315	W315	46° 07' 12.0000" E	30° 21' 36.0000" N	300	87.000	11
316	W316	46° 07' 12.0000" E	30° 07' 48.0000" N	180	146.000	8
317	W317	45° 54' 36.0000" E	29° 51' 36.0000" N	200	215.000	18
318	W318	43° 58' 56.2754" E	31° 35' 41.1816" N	185.2	146.028	5
319	W319	43° 24' 57.7358" E	30° 54' 48.9036" N	192.21	364.485	7
320	W320	44° 15' 30.1231" E	30° 26' 34.1984" N	237.3	317.790	10
321	W321	45° 53' 04.2182" E	30° 29' 40.1829" N	109.5	77.896	13
322	W322	45° 32' 06.8596" E	29° 44' 02.6956" N	201.7	199.184	11
323	W323	46° 26' 06.2870" E	29° 21' 28.6247" N	221.4	283.532	18
324	W324	45° 12' 48.8061" E	30° 56' 28.1578" N	115.8	69.769	10
325	W325	45° 15' 12.7971" E	30° 13' 22.0477" N	218.3	225.302	12
326	W326	44° 46' 13.8069" E	29° 42' 14.1101" N	260.7	342.364	9
327	W327	46° 19' 11.2395" E	29° 55' 07.0393" N	240.6	204.469	17
328	W328	44° 27' 25.1437" E	31° 04' 25.4341" N	166.1	173.214	12
329	W329	43° 54' 44.8533" E	30° 44' 46.9633" N	264.1	318.716	9
330	W330	44° 52' 29.8399" E	30° 29' 40.5681" N	238.4	217.031	9
331	W331	43° 43' 20.9153" E	30° 15' 26.6924" N	193.1	387.265	8
332	W332	43° 55' 48.6398" E	31° 13' 09.4988" N	236.3	237.664	9
333	W333	43° 26' 49.9841" E	31° 32' 33.9049" N	264.4	270.425	9
334	W334	43° 16' 28.6251" E	30° 29' 19.1704" N	212.1	385.890	9
335	W335	44° 21' 07.0150" E	29° 50' 45.6878" N	201.4	393.150	12
336	W336	45° 15' 49.3164" E	29° 20' 21.2537" N	160.2	329.000	7
337	W337	46° 05' 32.1760" E	29° 34' 17.2911" N	264.1	263.023	12

Well No.	Well Code	Coordinate		Depth (m)	Ground Elevation (m.a.s.l)	Pumping Rate (ℓps)
		Longitude	Latitude			
338	W338	44° 38' 15.5435" E	30° 41' 59.6365" N	231.2	215.194	13
339	W339	45° 47' 48.1820" E	30° 03' 19.7133" N	192.3	203.753	8
340	W340	46° 27' 17.0022" E	30° 17' 29.9661" N	150.2	102.123	17
341	W341	45° 16' 57.8083" E	29° 53' 00.2789" N	146.1	234.979	12
342	W342	44° 02' 55.0063" E	31° 57' 52.3314" N	174.2	47.950	3
343	W343	44° 05' 34.8613" E	31° 45' 10.2818" N	128.1	98.321	2
344	W344	43° 48' 00.7140" E	30° 42' 04.2410" N	264	345.000	7
345	W345	44° 34' 07.6289" E	31° 09' 16.0495" N	160.5	138.661	8
346	W346	45° 23' 52.5506" E	30° 58' 20.2802" N	61.4	30.870	16
347	W347	44° 23' 38.7233" E	29° 43' 27.8103" N	226.1	402.424	13
348	W348	45° 07' 43.2404" E	30° 05' 08.8888" N	235.2	250.558	8
349	W349	45° 02' 54.5451" E	29° 29' 21.5182" N	246	325.000	10

Tectonic and Geological Maps of Iraq

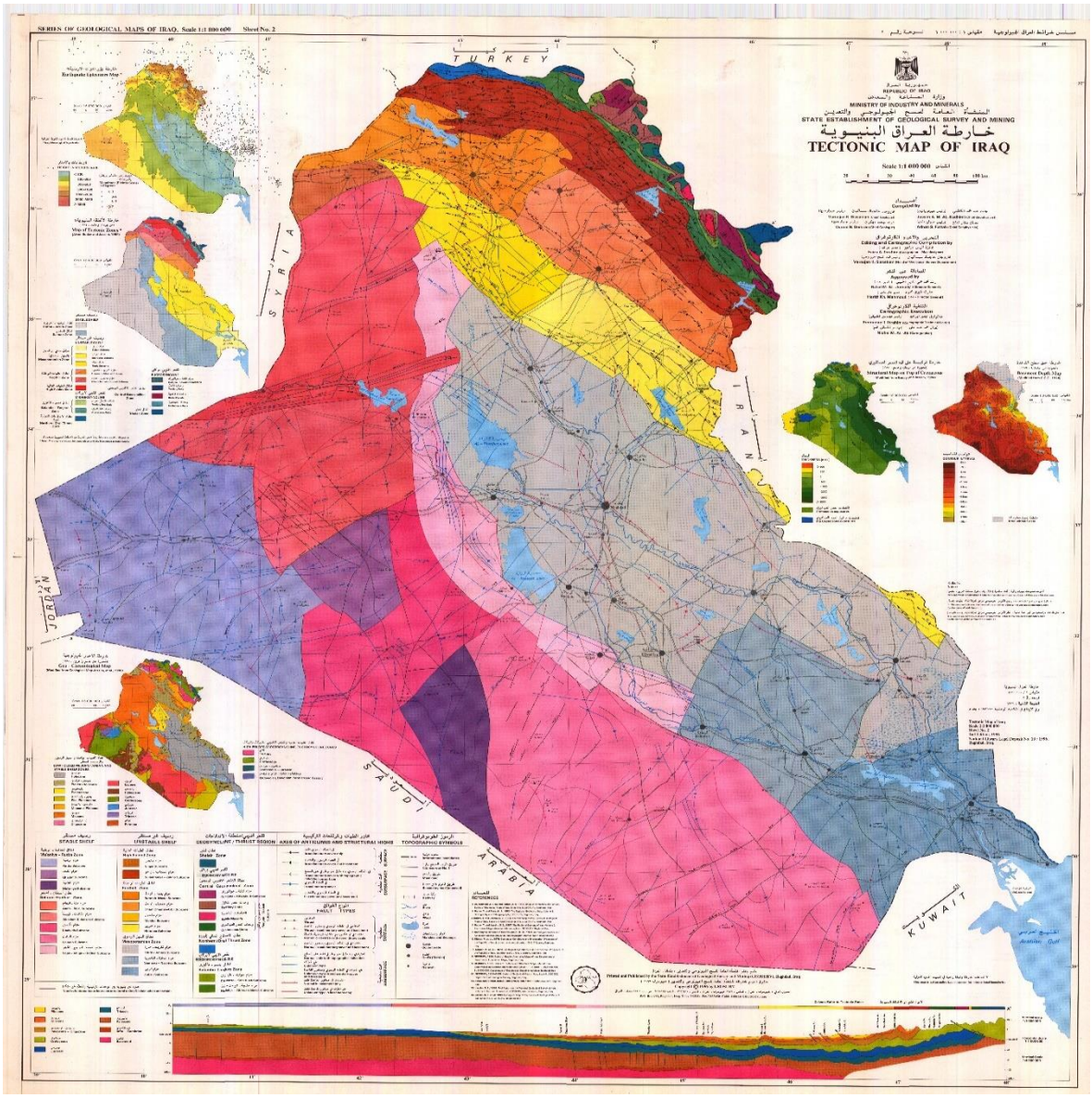


Figure (B-1): Tectonic map of Iraq (GEOSURV, 1996).

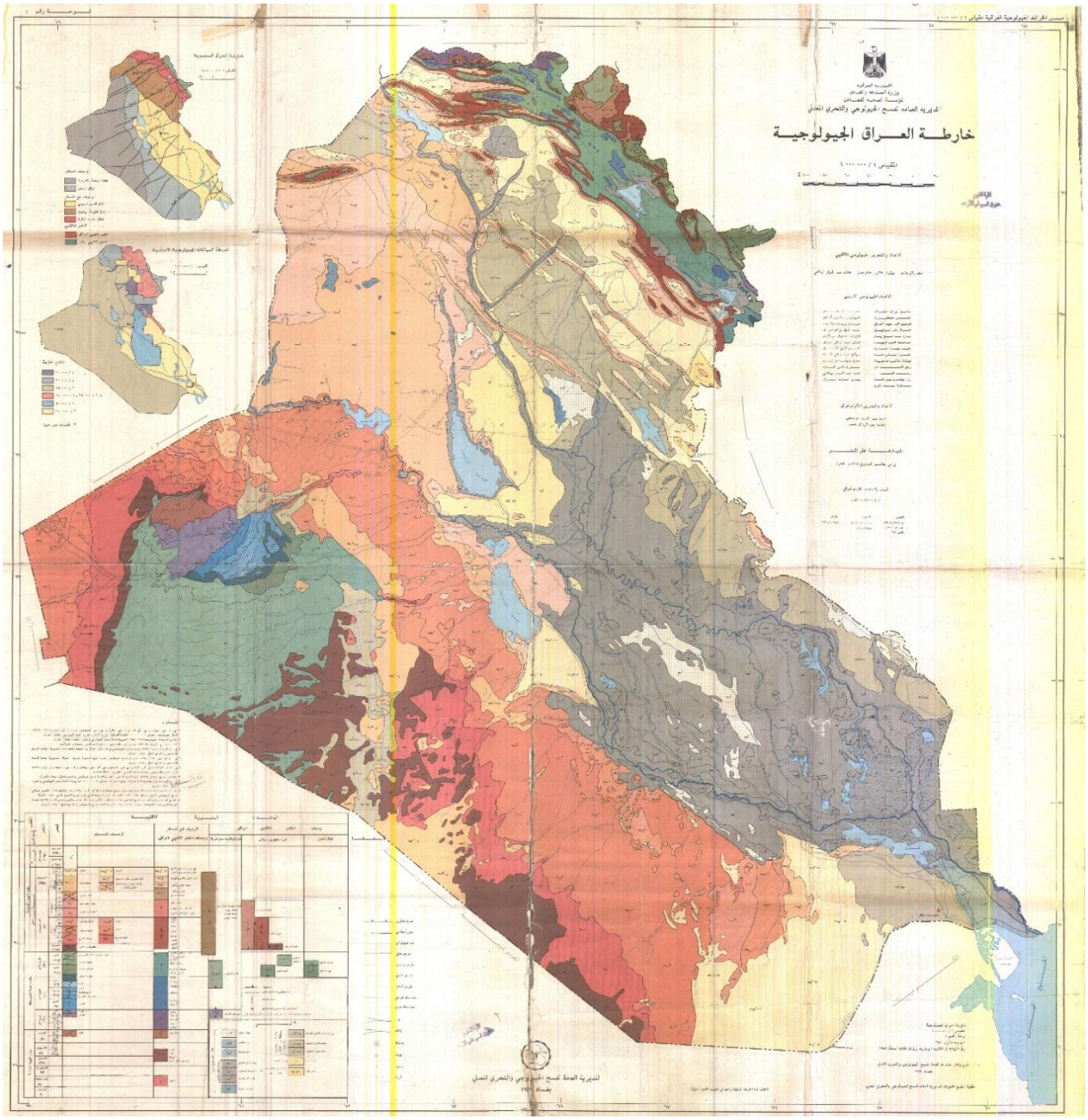


Figure (B-2): Geological map of Iraq (GEOSURV, 1986).

MATLAB Codes for Groundwater Potential Zones Detection

C-1 multiclass classification of groundwater potential zones

C-1.1 Parameters that control groundwater in the study area

1. Hydrogeological Aquifer Units

```
function [V] = Units(x,y)
A=imread('thematic maps\aquifer unit.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[255 187 0;0 97 0;164 196 0;255 38 0;255 255 255;0 0 0];
T=[4;3;2;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

2. Transmissivity (T)

```
function [V] = Trans(x,y)
A=imread('thematic maps\Transmissivity.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[255 153 0;0 97 0;255 34 0;255 255 0;122 171 0;255 255
255;0 0 0];
T=[5;4;3;2;1;0;0];

for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

3. Lineament Density (LD)

```
function [V] = LD(x,y)
A=imread('thematic maps\lineament density.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
```

```
B=[255 255 0;0 97 0;255 34 0;255 153 0;122 171 0;255 255
255;0 0 0];
T=[5;4;3;2;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

4. Slope (SLO)

```
function [V] = Slop(x,y)
A=imread('thematic maps\slope.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[0 97 0;255 34 0;255 255 0;122 171 0;255 153 0;255 255
255;0 0 0];
T=[5;4;3;2;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

5. Soil (SL)

```
function [V] = Soil(x,y)
A=imread('thematic maps\soil.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[255 217 0;255 38 0;255 132 0;197 219 0;0 97 0;97 153
0;255 255 255;0 0 0];
T=[5;4;4;3;2;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

6. Land use Land cover (LU/LC)

```
function [V] = LULC(x,y)
A=imread('thematic maps\LULC.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[0 77 168;255 235 190;0 97 0;233 255 190;76 230 0;130
130 130;255 255 255;0 0 0];
```

```
T=[5;5;5;4;3;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

7. Drainage Density (DD)

```
function [V] = DD(x,y)
A=imread('thematic maps\drainage density.bmp');
for i=1:length(x)
    P(i,:)=[A(x(i),y(i),1) A(x(i),y(i),2) A(x(i),y(i),3)];
end
B=[255 255 0;0 97 0;255 34 0;255 153 0;122 171 0;255 255
255;0 0 0];
T=[5;4;3;2;1;0;-1];
for i=1:length(x)
    V(i)=VG(double(P(i,:)),[B T]);
end
```

8. Wells pumping rate multiclass classification

```
function [V] = Wells(Flow)
    if Flow <= 3
        V=1;
    elseif (Flow >= 4 && Flow <= 6)
        V=2;
    elseif (Flow >= 7 && Flow <= 9)
        V=3;
    elseif (Flow >= 10 && Flow <= 14)
        V=4;
    else
        V=5;
    end
```

C-1.2 Create Features

```
A=imread('thematic maps\study area.bmp');
[n,m,~]=size(A);
for i=1:n
    X1(i,:) = Units(ones(1,m)*i,1:m);
    X2(i,:) = Trans(ones(1,m)*i,1:m);
    X3(i,:) = LD(ones(1,m)*i,1:m);
    X4(i,:) = Slop(ones(1,m)*i,1:m);
    X5(i,:) = Soil(ones(1,m)*i,1:m);
```

```
    X6(i,:) = LULC(ones(1,m)*i,1:m);
    X7(i,:) = DD(ones(1,m)*i,1:m);
    i
end
```

C-1.3 Create the Database

```
clc
clear
[num,txt,raw] = xlsread('wells349final.xls');
load('Features.mat')
x1=num(:,1);
y1=num(:,2);
y=0.001701124647527.*x1-3.025444143231974e+02;
x=-0.001702545318960.*y1+6.127384649127625e+03;
x=round(x);
y=round(y);
for i=1:length(x)
    D(i,1) = X1(x(i),y(i));
    D(i,2) = X2(x(i),y(i));
    D(i,3) = X3(x(i),y(i));
    D(i,4) = X4(x(i),y(i));
    D(i,5) = X5(x(i),y(i));
    D(i,6) = X6(x(i),y(i));
    D(i,7) = X7(x(i),y(i));
    D(i,8) = Wells(num(i,7));
end
save('Data.mat','D')
```

C-1.4 Mapping of multiclass groundwater potential map

```
clc
clear; close all
A=imread('thematic maps\study area.bmp');
load('Features.mat')
load('Data.mat')
[n,m,~]=size(A);
P=[122 171 0;0 97 0;255 34 0;255 255 0;255 153 0];
for i=1:n
    for j=1:m
        if X1(i,j)>0
            X(1)=X1(i,j);
            X(2)=X2(i,j);
            X(3)=X3(i,j);
```

```
            X(4)=X4(i,j);
            X(5)=X5(i,j);
            X(6)=X6(i,j);
            X(7)=X7(i,j);
        Ax(i,j)=VG(X,D);
        else
        Ax(i,j)=0;
        end
    end
end

save('Map_class.mat','Ax')
load('Map_class.mat')
for i=1:n
    for j=1:m
        if Ax(i,j)>0
            A(i,j,1)=P(Ax(i,j),1);
            A(i,j,2)=P(Ax(i,j),2);
            A(i,j,3)=P(Ax(i,j),3);
            end
    end
end

imshow(A)
L={{'Very poor'} {'Poor'} {'Moderate'} {'Good'} {'Very
good'}};
imlegend(P,L)
imwrite(A,'Final.bmp')
```

C-1.5 Area calculation

```
clc
clear
load('Map_class.mat')
[n,m]=size(Ax);
a=n*m-length(find(Ax<1));
a1=length(find(Ax==1))/a*100
a2=length(find(Ax==2))/a*100
a3=length(find(Ax==3))/a*100
a4=length(find(Ax==4))/a*100
a5=length(find(Ax==5))/a*100

a1+a2+a3+a4+a5
```

c-2 Binary classification of groundwater potential zones

c-2.1 Wells pumping rate classification

```
function [V] = Wells(Flow)
    if Flow <= 9
        V=1;
    else
        V=2;
    end
```

C-2.2 Mapping of binary groundwater potential map

```
clc
clear; close all
A=imread('thematic maps\study area.bmp');
load('Features.mat')
load('Data.mat')
% D=[D;zeros(1,8)];
[n,m,~]=size(A);
P=[122 171 0;255 153 0];
for i=1:n
    for j=1:m
        if X1(i,j)>0
            X(1)=X1(i,j);
            X(2)=X2(i,j);
            X(3)=X3(i,j);
            X(4)=X4(i,j);
            X(5)=X5(i,j);
            X(6)=X6(i,j);
            X(7)=X7(i,j);
        Ax(i,j)=VG(X,D);
        else
        Ax(i,j)=0;
        end
    end
end
save('Map_class.mat','Ax')
load('Map_class.mat')
for i=1:n
    for j=1:m
        if Ax(i,j)>0
```

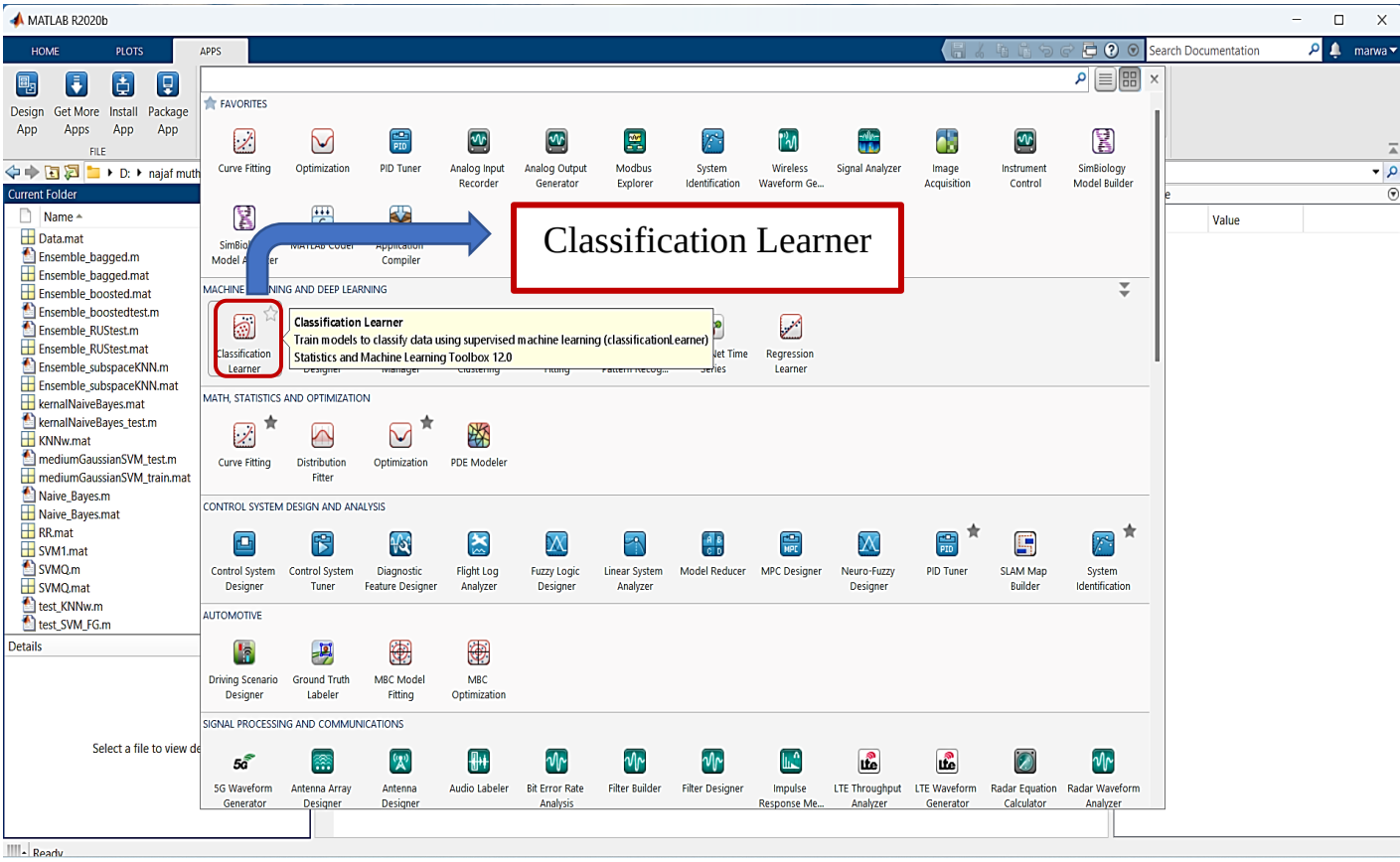
```
            A(i,j,1)=P(Ax(i,j),1);
            A(i,j,2)=P(Ax(i,j),2);
            A(i,j,3)=P(Ax(i,j),3);

            end
      end
end

imshow(A)
L={{'Low- Yield Groundwater'} {'High- Yield Groundwater'}
};
imlegend(P,L)
imwrite(A,'Final.bmp')
```

C-2.3 Area calculation

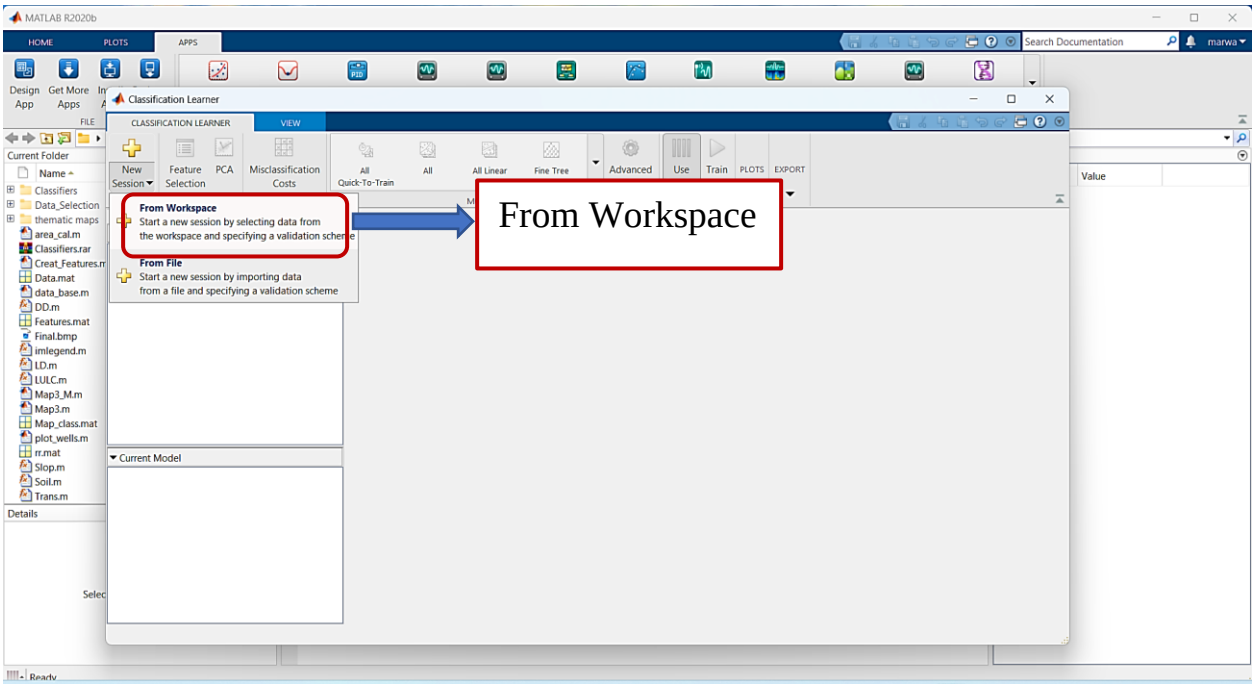
```
 clc
clear
load('Map_class.mat')
[n,m]=size(Ax);
a=n*m-length(find(Ax<1));
a1=length(find(Ax==1))/a*100
a2=length(find(Ax==2))/a*100

a1+a2
```

Classification Learner APP.

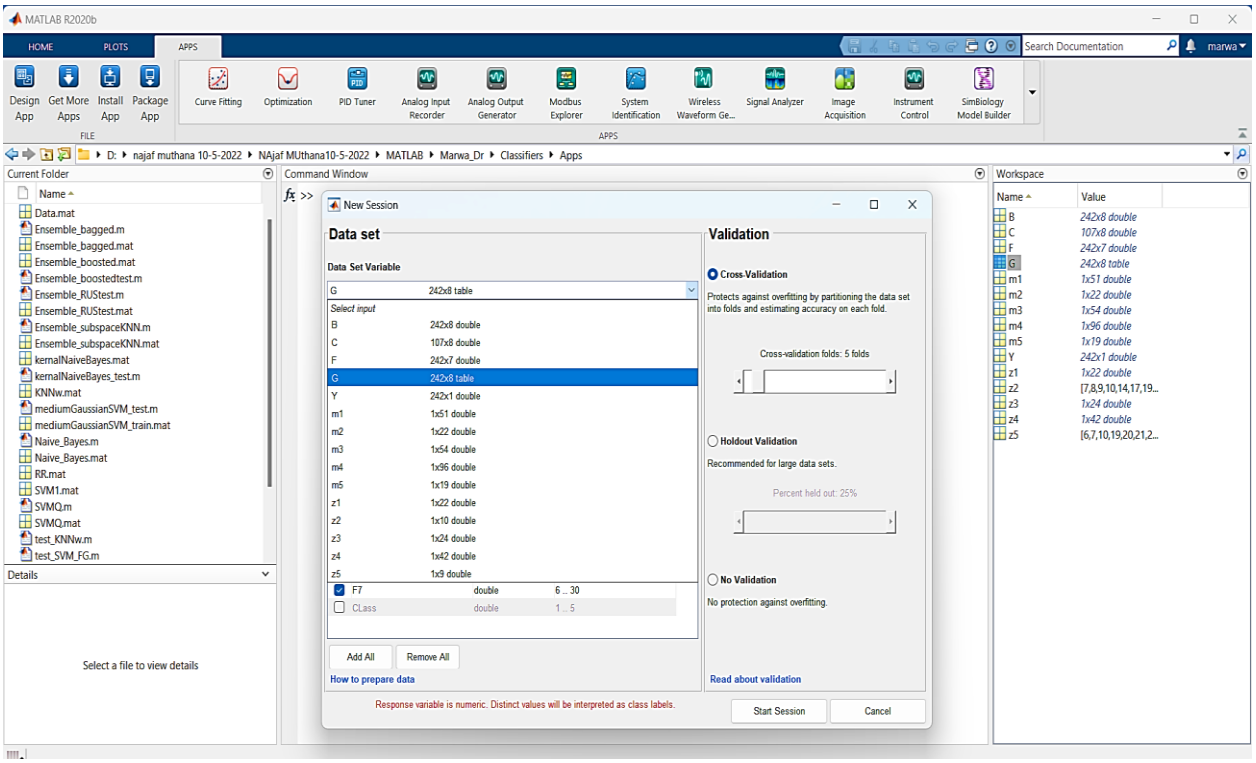
1. Select Classification Learner From the App section of the machine learning toolbox.



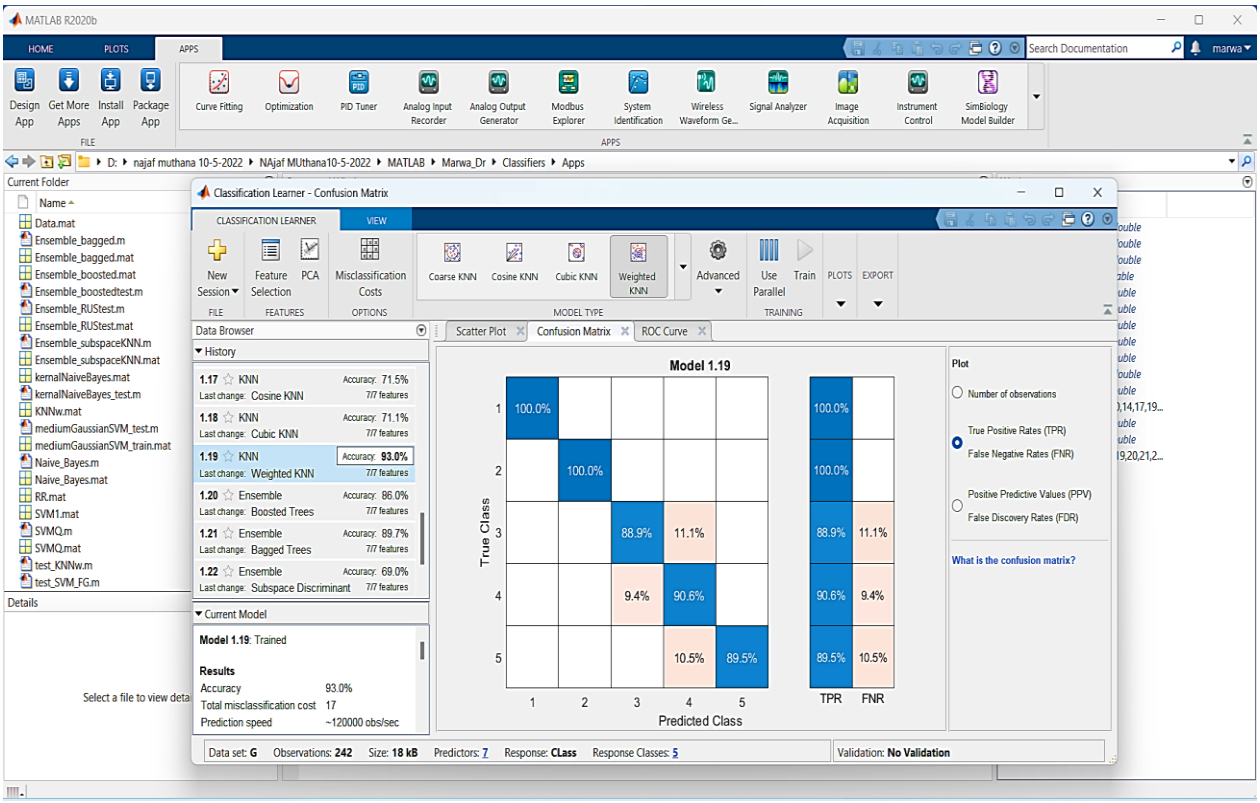
2. Start a new session by selecting data from the workspace and specify a validation scheme.



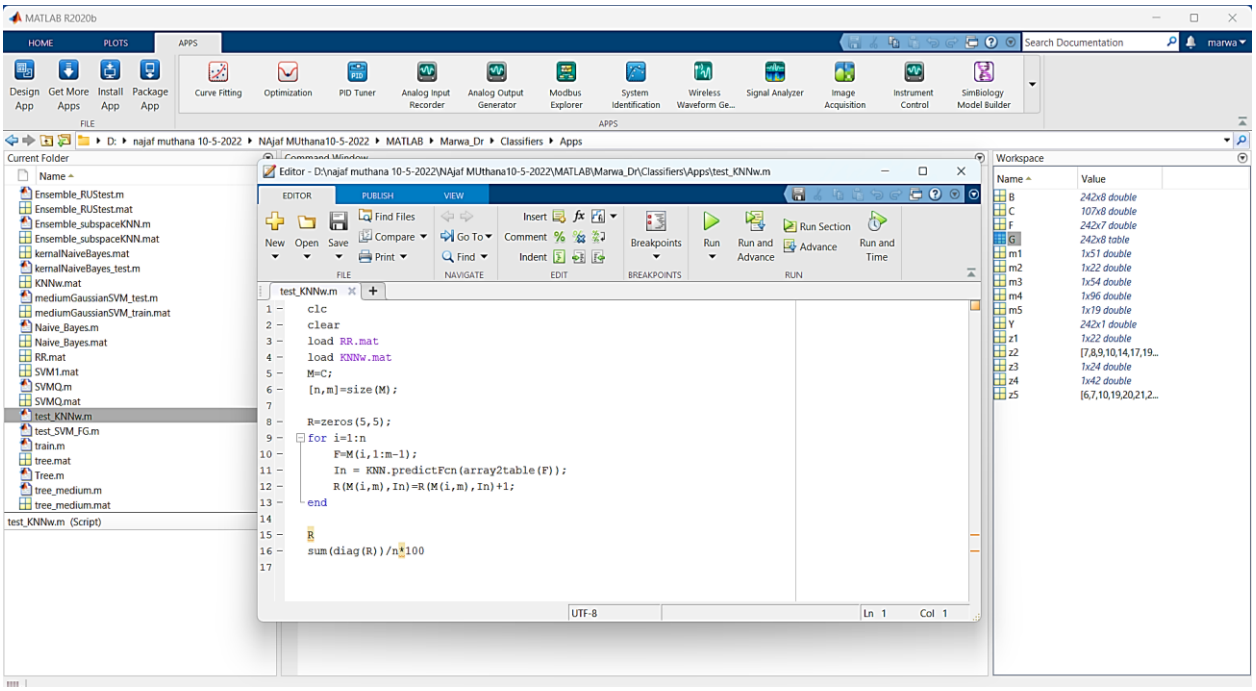
3. Select the training dataset to train the classifiers.



4. Train all available classifiers to select the optimal classifier performance.



5. Testing of classifier performance



الخلاصة

تعتبرالمياه الجوفية مورداً قيماً لتوفير المياه وخصوصاً في المناطق القاحلة وشبه القاحلة؛ إذ من المتوقع أن يؤدي تغير المناخ إلى انخفاض معدل هطول الأمطار وزيادة تواتر حالات الجفاف؛ ذلك بأن طول مدة الجفاف يؤثر تأثيراً سلبياً على الموارد المائية في العراق؛ فالجفاف هو السبب الرئيس لاستمرار استهلاك المياه الجوفية على مدى العقد الماضي، من هنا تبحث هذه الدراسة في امكانية تحديد مناطق الاجهاد والنضوب في المياه الجوفية ضمن طبقة الدمام المحصورة الممتدة على طول محافظتي النجف والمثنى حيث ينقسم العمل في هذه الدراسة على قسمين: يقدم الاول محاكاة عددية باستخدام برنامج نظام نمذجة المياه الجوفية (GMS v10.4) لتحليل المياه الجوفية في خزان الدمام الجوفي ضمن منطقة الدراسة ؛ حيث جرت معايرة النموذج في كلتا الحالتين المستقرة (في حالة السكون وعند عدم وجود سحب)، وغير المستقرة (في حالة السحب من الآبار). بالإضافة إلى ذلك، تم تقديم سيناريوهين تشغيليين واقعيين لتوفير تصور أفضل للمناطق التي قد يحدث فيها إجهاد للمياه الجوفية، إذ اوضحت النتائج ان تشغيل 89 بئراً بكامل طاقتها في آن معاً كسيناريو اول، قد تسبب بانخفاض منسوب المياه الجوفية بمقدار (0.4-5.8) متر، ومن جهة أخرى فإن السيناريو الثاني الذي تضمن استخدام 228 بئراً تسبب بالانخفاض في منسوب المياه الجوفية بمقدار (1-22) متر نتيجةً لزيادة معدلات الضخ الإجمالية من 582.3 لتر/ ثانية إلى 1534.1 لتر/ثانية. اضافة لذلك فقد وجد ان الموصلية الهيدروليكية لمنطقة الدراسة تتراوح ما بين 1.47 الى 20.0 م/يوم وان معدل إعادة التغذية يساوي1.66×10^{-6} م/ يوم؛ ولم تختلف هذه القيم المحسوبة من عملية المحاكاة اختلافاً كبيراً عن تلك التي تم الحصول عليها من البيانات الميدانية؛ وقد أظهر التطابق المُرضي بين ارتفاعات المياه الجوفية المُقاسة والمُتوقعة خلال فترة التشغيل من 1 يوليو 2021 حتى 1 يونيو 2022 بأن معامل الخزن لمنطقة الدراسة يتراوح من 5×10^{-5} إلى 6.5×10^{-4}. علاوة على ذلك، تم إنشاء خريطة المناطق المحتملة للمياه الجوفية (GPM) والتي تسهم في ادارة المياه الجوفية من خلال استخدام نهج صنع القرار متعدد المعايير(MCDM) المتمثلة بطريقة التحليل الهرمية (AHP)؛ إذ تم اعتماد العوامل التي تؤثربشكل كبيرعلى مدى توافر المياه الجوفية المتمثلة بـ (انواع وحدات الخزان الجوفي ، والنفاذية، وكثافة الصدوع، والميل ، ونوع التربة ، واستخدام الأراضي والغطاء الأرضي ، وكثافة الصرف). وتم عرض خريطة المناطق المحتملة للمياه الجوفية على شكل خريطة متعددة الفئات متكونة من خمس فئات هي: (جيدة جدا، وجيدة، ومتوسطة، وضعيفة، وضعيفة جدا)؛ حيث شغلت هذه المناطق مساحة 15.91٪، 39.02٪، 33.01٪، 6.49٪، و5.57٪ من اجمالي المساحة الكلية على التوالي؛ إذ تم تحقيق دقة (82.5٪) عن طريق دمج نُظُم المعلومات الجغرافية بعملية التسلسل الهرمي التحليلي (AHP). اما في الجزء الثاني من الدراسة

فقد تم استخدام تقنيات الذكاء الاصطناعي لانتاج نماذج هجينة تتضمن استخدام خمس خوارزميات للتعلم الآلي (MLAs) خاضعة للإشراف مدمجة مع التقنيات القائمة على نُظُم المعلومات الجغرافية لتقييم إمكانات المياه الجوفية التي تعالج قضية نضوب المياه؛ ذلك بان الخوارزميات المستخدمة هي EBT وهي من المُعَلِّمات المنطقية، وخوارزميات التعلم الإحصائي Naive Bayes (NB)، وآلات ناقلات الدعم (SVM)، والشبكات العصبية الاصطناعية متعددة الطبقات (MLP)، والجيران الاقرب لنقطة الاختبار (kNN) وهي من المعلمات القائمة على مقياس التشابه لنقاط الاختبار مع اقرب نقاط تدريب لها ، حيث تم استخدام هذه التقنيات الخمسة ذات الاسس المختلفة لإنشاء خريطة إمكانات المياه الجوفية (GPM). وأوجدت الدراسة الحالية أن المصنفات من نوع (kNN) قد أبدت اداء افضل مع دقة تحقق بنسبة 90.70٪ و AUC تساوي 1؛حيث تم تصنيف 322 بئر بدقة من اصل 349 بئر؛ وتأسيسا عليه تم إنشاء خريطة تمثل خمس مناطق مختلفة لإمكانات المياه الجوفية (GWPZ) وهي: 12.37 ٪ (جيد جدا)،42.12 ٪ (جيد) ،35.69 ٪ (معتدل) ،4.24 ٪ (ضعيف)، و5.58 ٪ (ضعيف جدا) من إجمالي المساحة الكلية على التوالي، بينما في التصنيف الثنائي تم إجراء تحليل مقارَن لتقييم فعالية وقدرات المصنفات التنبؤية إذ تم استخدام مصنف kNN لانشاءGPM ثنائية التصنيف وقد اظهر هذا الموديل دقة اعلى من بين الموديلات الاخرى بلغت قيمتها93.3 ٪، حيث قسمت الخريطة النهائية منطقة الدراسة على فئتين من GWPZs وهي منخفضة العائد وعالية العائد والتي غطت حوالي 45.51 ٪ و54.49 ٪ من المساحة الإجمالية على التوالي، علاوة على ذلك تم تطوير واجهة مستخدم رسومية (GUI) في التطبيقات القائمة على MATLAB , والتي يسهل استخدامها من قبل الجهات المعنية لتقدير وإدارة موارد المياه الجوفية المحتملة من دون الحاجة الى الكفاءة في استخدام البرامج او لغات البرمجة المعقدة لتحديد مناطق المياه المحتملة قبل عملية الحفر؛ حيث تتضمن العملية فقط إدخال الإحداثيات الجغرافية للبئر باستخدام إسقاط UTM ، بما في ذلك خطوط الطول والعرض، مما يسهل تحديد تصنيفها وموقعها على خريطة منطقة الدراسة.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة بابل/ كلية الهندسة

قسم الهندسة المدنية

النمذجة باستخدام الذكاء الاصطناعي الهجين للمياه الجوفية في مكمن الدمام المحصور/ العراق

اطروحة

مقدمة الى كلية الهندسة / جامعة بابل

كجزء من متطلبات نيل درجة دكتوراه فلسفة في الهندسة المدنية / هندسة الموارد المائية

من قبل

مروة عبد الله شلاش

(ماجستير في الهندسة المدنية/ المنشأت الهيدروليكية والموارد المائية/ 2012)

بأشراف

أ.م.د. عماد حبيب عبيد

ذو القعدة 1444 ايار 2023