TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PASHCHIMANCHAL CAMPUS

A FINAL REPORT ON

"Integrated Approach for Delineation of Groundwater Potential Zones in Chitwan District: Statistical, and Machine Learning Analysis"

SUBMITTED TO:

RESEARCH MANAGEMENT CELL
IOE PASHCHIMANCHAL CAMPUS
TRIBHUVAN UNIVERSITY
POKHARA, NEPAL

UNDER THE SUPERVISION OF:
Er. Netra Bahadur Katuwal

SUBMITTED BY:

INDRA PAUDEL	(PAS077BGE022)
SACHIN RAYAMAJHI	(PAS077BGE036)

DATE: 03/04/2025

Table of Content

LIST OF ABBREVIATION

- DEM - Digital Elevation Model
- EIA - Environmental Impact Assessment
- GIS - Geographic Information System
- GWP - Groundwater Potential
- GWPZ- Ground Water Potential Zone
- LR – Logistics Regression
- MLR – Multiple Linear Regression
- MIF - Multi-Influencing Factor
- ML – Machine Learning
- NGO - Non-Governmental Organization
- RF – Random Forest
- RS - Remote Sensing

LIST OF FIGURES

LIST OF TABLE

LIST OF EQUATION

Abstract

Groundwater is one of the important natural resources that are available to meet the demand of water for domestic, agricultural, and other purposes. Availability of ground water depends upon the spatial and temporal nature of terrain. In this research, the delineation of groundwater potential was studied using three different models. : Random Forest (RF), Logistic Regression (LR), and Multiple Linear Regression (MLR) are statistical approaches. The groundwater well points for training and testing data were collected from the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board, and using OpenStreetMap. The total 600 well points were collected using those sources: 300 present well points and 300 absent well points. The data was sampled for training and testing points using Random Sampling technique with a 75:25 ratio respectively. We analyzed 11 key factors to map Ground Water Potential Zones (GWPZs). A database was created using these factors and well points. To refine the factors, we applied three screening methods: Information Gain (IG), Pearson Correlation Coefficient (PCC), and VIF & Tolerance, using training inventory points. Each factor was weighted based on different models, and the final GWPZ map was generated using ArcGIS 10. The resulting GWPZs were classified into 5 zones using Natural Breaks Classification methods as Very Low, Low, Moderate, High and Very High zones. The resulting GWPZs maps were validated through the Area-Under-the curve (AUC) analysis using the wells data sampled for testing the model performance. The obtained AUC values for LR, MLR and RF model are 81.3%, 84.6% and 87.2% respectively. Hence, Random Forest (RF) method is best for delineation of GWPZs maps and ground water resource planning by comparing the AUC value between different models. Therefore, the study can play a vital role in watershed management and identifying appropriate locations for wells in the future.

Keywords: Groundwater Potential Zone (GWPZ); Logistic Regression (LR); Geographic Information System (GIS); Remote Sensing; Multiple Linear Regression (MLR); Random Forest (RF); AUC-ROC

Chapter 1 Introduction

1.1 General Background

Groundwater is the crucial water resource that is stored in the cracks and spaces of soil, sand, and rock. Groundwater is water that is found underground in saturated zones below the land's surface. It lies in aquifers, which are permeable rock and/or sediment that holds the water. . It provides approximately half of the accessible freshwater used for daily cooking, drinking, and cleaning (Pande et al., 2021). According to Global Environment Facility (GEF), today groundwater is estimated globally to provide 36% of potable water, 42% of water for irrigated agriculture, and 24% of direct industrial supply. At present, only about 22% of the available dynamic groundwater recharge in Terai is being utilized. The quality of groundwater is generally suitable for irrigation as well as drinking purpose. The groundwater in Kathmandu Valley is overexploited (Shrestha et al., 2018). Rapidly increasing population, unequal distribution of water resources, industrialization, and global warming have contributed to a significant rise in freshwater demand, resulting in water scarcity worldwide (Jasrotia et al., 2016). This increasing demand and supply of groundwater has caused excessive and unplanned groundwater extraction, which is responsible for the degradation of these resources both in terms of quality and quantity (Treidel et al., 2012). Hence, identifying areas with the potential for groundwater is equally crucial for the continued improvement of irrigation systems and the sustainable utilization of this resource (Pathak, 2017).

Groundwater in Terai originates from rainwater infiltration (water movement across the ground surface into the soil) and lateral recharge from the Bhabar zone, along the Siwalik foothills. Water-bearing soil, sand and rock formations are known as aquifers, which can absorb and transmit water. Rainfall naturally recharges most of the aquifers. In addition, sub-surface inflow and seepage losses from streams and rivers also contribute to the recharging process (Panday, 2023). Chitwan is a NNW–SSE-trending dun valley (an alluvial basin) surrounded by the Mahabharata and Churia hills. It is also called Bhitrimadesh (inner plain land of the Inner Terai) and lies in the Narayani Zone of the Central Development Region (Malla & Karki, n.d.). According to a study published in the International Journal of Environmental Research and Public Health, the groundwater level in the district has been declining at a rate of 0.5 to 1 meter per year due to over extraction, particularly in the dry season (Shiwakoti et al., n.d.). In Chitwan, groundwater levels are depleting due to population growth and urbanization, land use/cover (LULC) changes and excessive water extraction. While reducing water extraction is challenging, enhancing groundwater recharge can help balance the extraction. Identifying locations with high infiltration and groundwater storage capacity is crucial, yet such areas remain unknown. Our study identifies these high-potential recharge zones, facilitating the preservation of these areas to mitigate future drinking water scarcity in Chitwan district.

Numerous studies of the literature suggests that researchers across the globe have used various methods to delineate Ground Water Potential Zones. Among them Analytical Hierarchy Process (Rahman et al., 2022), Logistic regression (Park et al., 2017), Frequency Ratio (Razandi et al., 2015) are very commonly used for this purpose. Besides, various machine learning is also being used more in advanced to delimit the Groundwater Potential Zones. Those approach include Random Forest (Sameen et al., 2019), Naïve Bayes (Arabameri et al., 2021), SVM (Support vector machine) (Lee et al., 2018)), classification and regression tree (Arabameri et al., 2021) and artificial neural network (Lee et al., 2018). All those methods are being used for the determination of the GWPZs in the planet but every method and techniques follows some drawbacks. Identification of groundwater potential zones based on one single method is now not justifying the study (Hasanuzzaman et al., 2022).

The machine learning (ML) approach uses statistical models and optimization algorithms to identify hidden relationships between the inputs and outputs in large datasets (Martínez-Santos & Renard, 2020). Random Forest is a widely used machine-learning algorithm, which combines the output of multiple decision trees to reach a single result. While comparing with other machine learning approaches, RF has out performance in both classification and prediction, unlike other methods it uses the combination of multiple decision trees and produces the result based on majority voting. Similarly, many studies employed a statistical approach for Ground Water Potential Zones mapping. Logistic Regression is the most widely used empirical model in the fields of science, in particular for environmental studies. The LR model is a type of multivariate regression used to explain the relationships among a dichotomous dependent variable coded into 0 and 1, and one or more categorical or numerical independent variables (Hosmer et al., n.d.). This approach uses statistical algorithms and ground truth data in place of expert opinions to compute the weights of influencing variables. The statistical approach is considered a hybrid of the expert decision and ML approaches because it uses similar methods to the expert decision (i.e., overlay analysis to calculate Ground Water Potential Index and reclassification of GWPI to create

GWPZs). A multiple linear regression (MLR) equation where the computed recharge rates (dependent variable) and the interactive model regression between the geoelectric parameters (layer resistivity and layer thick ness) (independent variables) will form the underlying recharge model (Mogaji et al., 2015). With different GWP mapping approaches, there are trade-offs in terms of costs, accuracy, and complexity associated (Thanh et al., 2022).

Although many researches has been conducted on ground water recharge potential in Chitwan, a comprehensive study has not been done yet. Similarly, the studies of the literature shows up, there is vast gap of those three models in a same study area to delineate the Ground Water Potential Zones. Therefore, our study intends to map out the Ground Water Potential Zones of the Chitwan using LR, MLR and RF models with the consideration of the influencing factor such as Elevation, Slope, Aspect, Geology, Precipitation, Temperature, Topographical Wetness Index (TWI), and Land use and Land cover (LULC). Similarly, the findings will present the recommendations for preventive measures to address potential issues while providing valuable insights for assessing and developing groundwater resources.

1.2 Objectives

1.2.1 Primary Objective:

The main objective of this research is to delineate groundwater potential zones in the Chitwan district using Geographic Information System (GIS), Remote Sensing (RS), and Logistic Regression (LR), Multiple Linear Regression (MLR) and Random Forest (RF) techniques.

1.2.2 Secondary Objective

The Sub-objectives of this research are:

1. To analyze the influence of hydrogeological factors such as slope, drainage density, land use, geology, lineament density, elevation, rainfall, and soil on groundwater potential.
2. To apply and compare the effectiveness of the Logistic Regression, Multiple Linear Regression (MLR) and Random Forest methods in mapping groundwater potential zones.
3. To evaluate and categorize the study area into regions of high and low groundwater potential.

1.3 Study Area

Chitwan district in Nepal is located in the central part of the country and is situated in the Himalayan orogenic belt, which is characterized by complex geological structures and processes. It has an area of 2,238.39 Sq.km (864.25 Sq. mi) lies within latitude 27°36′21.60″ north, longitude 84°22′47.28″ east.

The district takes its name from the Chitwan Valley, one of Nepal's Inner Terai valleys between the Mahabharat and Siwalik ranges, foothills of the Himalayas. Chitwan district in Nepal features a diverse landscape with elevations ranging from approximately 100 meters above sea level in the southern Terai region to over 2,000 meters in the northern hills. This variation encompasses low-lying areas to mountainous terrains. The district's topography reflects significant elevation changes from south to north.

Chitwan, a district in Nepal, boasts several significant water bodies and river systems. The Narayani River, the largest in the region, serves as a vital water source for both irrigation and hydropower generation. Another major river, the Rapti, originates in the Siwalik Hills and flows through the picturesque Chitwan valley. Meanwhile, the Reu River, though smaller, plays an essential role, originating in the Mahabharata Range and winding its way through the eastern part of Chitwan. Notably, within the Chitwan National Park, lies Beeshazar Tal—a wetland complex teeming with migratory bird species. This pristine habitat not only supports local tourism but also sustains fishing activities, contributing to the area's economy. Ground water being the dominantly used source of water in this area there are several artificial recharge structures in the districts which includes ponds, dug wells, recharge trenches, and recharge wells.

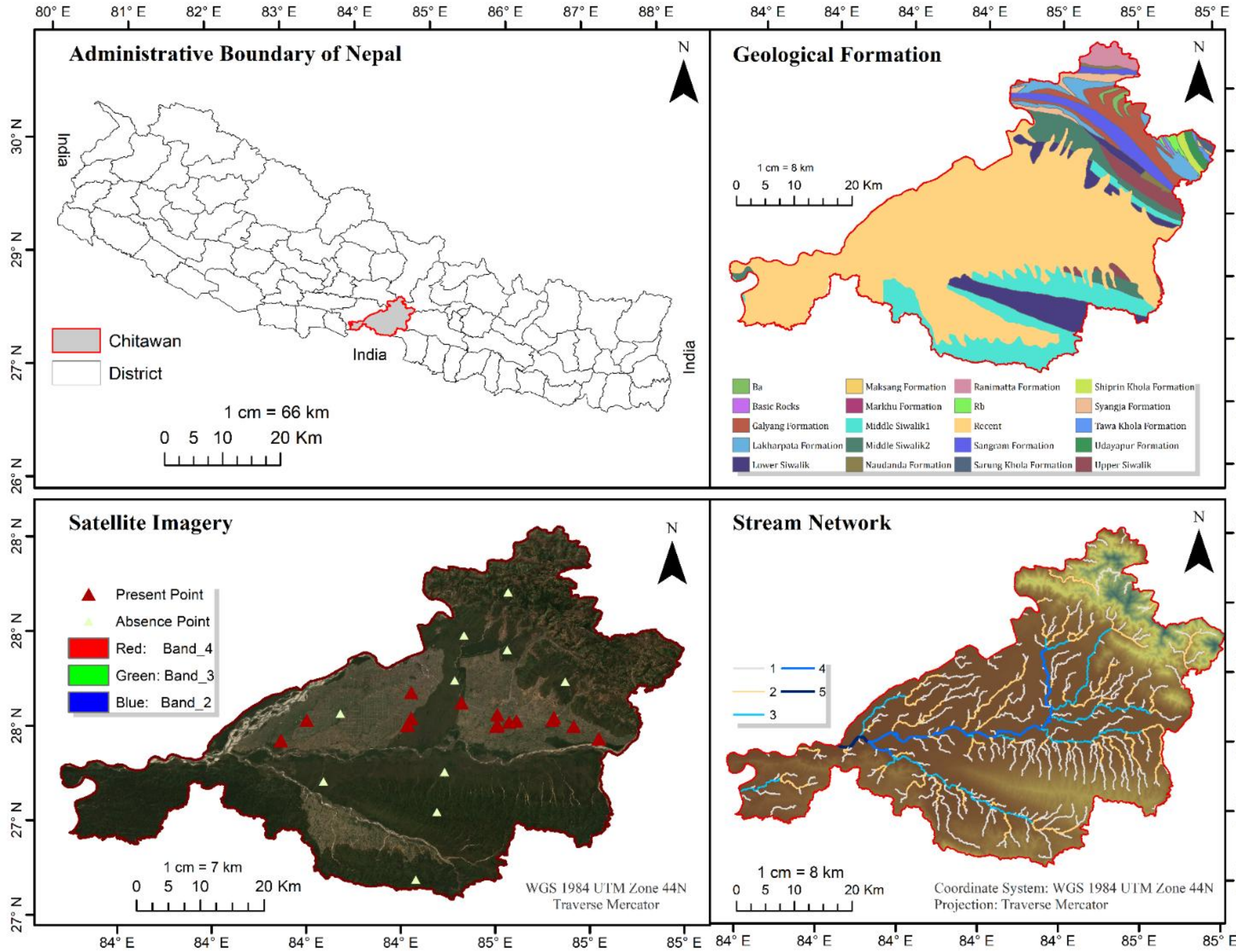

Figure 1: The study area map

a) The administrative boundary of Nepal with Chitwan Districtl.

b) The map representing Landsat 8 true composite multi-spectral image of Chitwan District with Presence and Absence of well points.

c) The geological map of the area.

Chapter 2 Literature Review

Groundwater is among the most indispensable resources of the earth that takes place below the surface of the earth (Naghibi & Pourghasemi, 2015). Groundwater is an essential part of the hydrological cycle and is a valuable natural resource providing the primary source of water for agriculture, domestic, and industrial uses in many countries. Groundwater is the renewable reserve of freshwater in many cases is stored in aquifers (Shiwakoti et al., n.d.). Groundwater can be recharged with the precipitation, streams, rivers and other sources in the water-bearing soil, rock formation known as aquifers that absorbs and transmits the water. The groundwater is also the main source of water that serves for different purposes and the utilizations.

Large numbers of countries of the earth are facing the problem of water scarcity at the societal level (Manap et al., 2014). Over the last few decades, areas with varying demographics and economic reliance have been forced into the 'water crisis', which has to be topmost among the many crises staring us in the face (Guru et al., 2017). The prospects of global population growth that will lead to an increase of approximately 3 billion by 2050 and the continued deterioration of water quality suggest that, even if everything else remains equal, water scarcity will intensify over the coming decades (Vaux, 2011). Groundwater varies spatially in both quality and quantity; however, it is very important for socio-economic development because groundwater meets certain demands of mankind, namely water for drinking, for irrigation, for forestry, for industrial purpose and to support livestock (Naghibi et al., 2016). Groundwater is less vulnerable than surface water sources to climate fluctuations in an undisturbed aquifer system and therefore, acts as a key buffer against drought and normal variations in rainfall. Reliable supply of groundwater lead input for increasing yields, reducing agricultural risk, stabilizing farm incomes and thus leading to higher levels of social and economic security (Moench, 2003). The use of groundwater has particular relevance to the availability of many potable water supplies because groundwater has a capacity to balance large swings in precipitation and associated increased demands during drought and when surface water resources reach the limits of sustainability (Treidel et al., 2012). In Nepal, there are 1,337,000 ha of irrigable agricultural land in the Terai. But, only 1,121,000 ha (84 percent) are under irrigation at present, out of this, only 206,000 ha (or 18 percent of the total irrigated area) are under irrigation by groundwater (Kansakar, 2016). Under the Agricultural Perspective Plan (APP) and Community Ground Water Irrigation Project (CGISP) large number of shallow tube well (STW) and deep tube well (DTW) have been installed to irrigate a vast agricultural land through the groundwater (Pathak, 2017). Intra-mountain valleys such as Kathmandu, Dang and other similar valleys have isolated groundwater basins whereas groundwater in the southern Terai plain is a part of the larger system in the Gangetic basin (Kansakar, 2016).

Groundwater mapping has been carried out with direct field surveys in recent past in expensive and time-consuming manner (Prasad et al., 2020). Groundwater exploration has traditionally relied on drilling, geophysical, geological, and hydrological methods, but these methods are time-consuming and expensive (Murmu et al., 2019; Razandi et al., 2015). The delineation of the groundwater potential zone through conventional field-based methods is time-consuming and expensive and requires a high degree of professionalism (Rahmati et al., 2015). Also, most of the conventional methods did not consider the combined interactions of different groundwater controlling variables (Acharya et al., 2019). Alternatively, the geospatial technology that combines the wide utilization of remote sensing and geographic information system is a rapid and cost-effective tool that helps in divination of Ground Water Potential Zones. Groundwater occurrences can be recognized as the outcome of a complex interaction between various factors including hydro geomorphology, lithology, topography, and land use/land cover (LULC), slope, drainage pattern, and hydrological conditions within the area (Murmu et al., 2019). The occurrence of ground water varies over place to place in accordance with hydrology, climate, topography, geology, ecology, soil, slope, etc., of the region (Karimi-Rizvandi et al., 2021). Therefore, the geospatial technology incorporates those influencing factor that are mentioned above to determine the suitable GWPZs. However, the systematic integration of these different thematic factors using geospatial technology accompanied by the hydrogeological investigation provides a rapid and cost-effective method of groundwater potential delineation (Champak et al., 2018). By combining the remote sensing information with adequate field data, particularly well inventory and yield data, it is possible to arrive at prognostic models to predict the ranges of depth, the yield, the success rate and the types of wells suited to different litho units under different hydrogeological domains (Kumar & Kumar, 2011).

For the integration of the different thematic layers, different probabilistic models including frequency ratio, certainty factor, logistic regression, artificial neural network model, Shannon’s entropy, different machine learning models, etc., are being practiced, among which using the Analytical Hierarchy Process (AHP) has been considered the powerful multi-criteria decisionmaking tools, especially for a groundwater regime (Khadka & Pathak, 2021; Sarkar et al., 2022). In this context Logistic Regression is widely used statistical approach in the field of GIS. The LR model is useful for predicting whether groundwater exists in a particular location, based on the predictor variables (Kim et al., 2019). The primary reason for using the LR model is to explain the relationship between dependent variables and independent variables (Lee & Talib, 2005). In LR, the probability of occurrence of a phenomenon is estimated within the range of 0 to 1, and it is not necessary to assume the normality of the predictor variables. Binomial LR analysis is used when the dependent variable is at the binomial nominal level, and it is used to predict the presence or absence of an attribute based on a set of independent variables (Nguyen et al., 2020). The advantage of the LR model is that variables need not have normal distribution, regardless of whether they are continuous, discrete, or a combination of both types (Lee et al., 2020). Multiple linear regression (MLR) is a linear approach for modeling the relationship between input parameters and resulting metrics (Huang et al., 2019). Multiple regressions have the general purpose of learning more about the relationship between one or more independent or predictor variables and a dependent or observed variable(Pan et al., 2019). In other words, the multiple linear regression can predict the value of the dependent variable for the given set of predictors. The multiple regression model will match the observed dependent variable with a measured variable by changing the coefficients linearly relating to the predictors (Nathan et al., 2017). Random forest is one of the popular machine learning approach that builds multiple decision trees from different subsets of data which uses the bagging technique to randomly select samples from the training dataset for the classification and regression tree construction (Karki, 2022). The random forest model is an ensemble classification technique that was developed as an extension of classification and regression trees (CART) to improve the prediction performance of the model (Cutler et al., 2012). The random forest does not require any assumptions about the relationship between explanatory rvariables and response variables. This is an appropriate way to analyze hierarchical interactions and nonlinearities in large data sets. RF model creates thousands of trees, forming a ‘forest’ based on the decision rule. Each tree in the RF model depends on a sample of bootstrapped of data using a CART process with a random subset of variables selected at each node (Arabameri et al., 2019). The RF model is also a very fast machine learning solution, allowing a highly accurate classification with internal unbiased generalizability estimation during the process of forest construction (Reif et al., 2006). Random Forest have gained popularity in GWPZs mapping. These algorithms enable automated analysis and classification of large datasets, improving the accuracy and efficiency of mapping.

Groundwater resources occur in various natural settings in Nepal due to diversity in its geology, geomorphology and physiography (Kansakar, 2016). Among the different physiographic divisions of Nepal, the groundwater resources in the Indo-Gangetic Plains and the inter montane basins are relatively well exploited (Silwal et al., 2023). Pathak (2016) conducted the study on water availability and hydrogeological condition in siwalik foothill of east Nepal. In the study the author used geological and hydrogeological data to delineate the groundwater potential (Champak et al., 2018). Chitwan Valley is one of Nepal's Inner Terai valleys between the Mahabharata and Siwalik ranges, foothills of the Himalayas (Malla & Karki, n.d.). The inner Terai valleys (e.g. Dang, Deukhuri, and Chitwan) are drained by larger rivers and have coarse fluvial sediment deposits, allowing quick aquifer recharge (Kansakar, 2016). Most of the study suggests that the consideration of the influencing factor for the determination of the GWPZs mapping based upon the natural climate, and the location of study area. The study of the different literature suggests that the influencing factors includes elevation, rainfall, lineament density, drainage density, slope, lithology, geomorphology, soil, topographical wetness index (TWI), distance from the river, Normalized Difference Vegetation Index (NDVI), and Land use and Land cover (LULC).

Identification of groundwater occurrence location using remote sensing data is based on indirect analysis of directly observable terrain features like geological structures, geomorphology, and their hydrologic characteristics (Champak et al., 2018). Groundwater potential zones in the Siwalik of Eastern Nepal have not been assessed even using any costly and time-consuming techniques, especially based on geophysical and other hydrogeological methods (Silwal et al., 2023). The growing number of wells, uncontrolled pumping and unregulated disposal of pollutants are all proximate causes of emerging groundwater problems in Kathmandu. This calls for the more research that will address an integrated approach on availability and quality of groundwater for past, current and future scenarios. (Pradhananga et al., 2012).

In conclusion, Groundwater is a critical natural resource, essential for agriculture, domestic use, and industrial activities, particularly in regions facing water scarcity(Naghibi & Pourghasemi, 2015). Its spatial variability in quality and quantity makes it a vital component for socio-economic development, especially in drought-prone areas where it serves as a buffer against climate fluctuations(Moench, 2003). In Nepal, groundwater plays a significant role in irrigation, with projects like the Community Ground Water Irrigation Project (CGISP) enhancing agricultural productivity through shallow and deep tube wells (Khadka & Pathak, 2021)However, traditional groundwater exploration methods, such as drilling and geophysical surveys, are costly and time-intensive, prompting the need for more efficient approaches (Murmu et al., 2019). Geospatial technologies, integrating remote sensing and GIS, have emerged as cost-effective tools for delineating groundwater potential zones by analyzing factors like topography, lithology, and land use (Champak et al., 2018). Advanced techniques such as logistic regression, random forest, and machine learning models are increasingly being employed to improve the accuracy of groundwater mapping (Thanh et al., 2022). Despite these advancements, regions like the Siwalik foothills in Nepal remain understudied, highlighting the need for further research to address groundwater availability and quality challenges (Silwal et al., 2023; Pradhananga et al., 2012).

Chapter 3 Research Methodology

3.1 Data and Software

Data: The project incorporates different kinds of datasets from several trustworthy sources. The elevation data were extracted using the SRTM DEM provided by USGS Earth Explorer. The groundwater inventory data, such as wells and springs, were collected using the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board, and OpenStreetMap (OSM). The geological data of the study area were extracted from the website of ICIMOD. The soil condition of the study area was acquired using raster images distributed by NAARC. Similarly, the precipitation data was obtained from a raster image of the CHIRPS using GEE. The lineament density of the study area was derived using Landsat 8 SWIR bands. Other factors used in the study were derived using these primary datasets.

Table 1 The table below shows the description about the data can be used for the study

S.N	Data	Resolution /Scale	Data Type	Source
1	Administrative Boundary	1:250,000	Vector	Survey Department Nepal
2	DEM (SRTM DEM)	30 m	Raster	USGS Earth Explorer
3	Geology	1:1000000	Vector	ICIMOD
4	Precipitation	5 km	Raster	CHIRPS
5	LULC	30 m	Raster	Landsat 8 using Random Classifier
6	Lineament Density	30 m	Raster	USGS Earth Explorer
7	Soil	100 m	Raster	NARC
8	Well point	Random Points	Vector	Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board; OSM.

Software: To visualize, analyze, process, manage, monitor, calculate, and study, different softwares and tools were utilized. The cloud-based processing environment, Google Earth Engine (GEE) was used to visualize and process multi-petabyte archives of satellite imagery and geospatial datasets. ArcGIS10@ESRI was used as an instrument in managing, utilizing, and visualizing the spatial data. For the management and storage of statistical and tabular data, various Microsoft products were used. Anaconda is a distribution of the Python and other programming languages for scientific computing that aims to simplify package management. Python, with its extensive libraries, was crucial in examining and analyzing the collinearity relationships between different factors. Together, these software tools enabled the effective implementation of the Logistics Regression (LR), Multiple Linear Regression (MLR) and Random Forest (RF) models for the delineation of Groundwater Potential Zones (GWPZs) maps.

3.2 Preparation of Wells Inventory Points

The collection of the wells inventory points was done using different trustworthy sources. The well-presented points were collected using the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board and OpenStreetMap (OSM). The dashboard of the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board contains the datasheet of spatial information of well points. Additional data was collected from OSM using Tags: man_made=water_well. Altogether 600 inventory points were collected, including 300 present and the other 300 as absent well points data in a study area. Moreover, those inventory points were classified as 0 for absent and 1 for present of well points. The total point data was resampled in the proportion of 75:25 for respective training and testing datasets. This structured approach ensures a balanced and reliable dataset delineation of the GWPZs map.

3.3 Preparation of Influencing Factors

For this purpose, all influencing factors for groundwater potential zones map were converted into raster and reclassified with the same pixel size (30 m × 30 m) and the same projection using GIS tools under the Arc toolbox in conversion as well as a spatial analysis tool. All the raster contains the same shape and geometry as the same rows and columns for raster data layers.

3.4 Reliability test

Reliability test is statistical methods that allows one to assess the consistency and stability of the estimated model. Such methods provided the confidence in the performance of the model on new data and its robustness when rerun on different samples to confirm that any identified relationships are not due to chance variations or outliers that could affect reliability.

3.4.1 Correlation matrix/Heat map

Correlation is a statistical measure that describes the extent to which two or more variables are related to each other. It indicates the strength and direction of a relationship between variables When variables are correlated, it means that changes in one variable are associated with changes in another—either positively or negatively: Here, 1 - Indicates a perfect positive correlation. When two variables increase or decrease together, they are positively correlated. Similarly, -1 - Indicates a perfect negative correlation. When one variable increases while the other decreases, they are negatively correlated.

$$\textit{Equation 1}: \textit{Pearson Correlation Coefficient}\ (r) = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\left\{\sqrt{\{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2\}}\right\}}$$

Where X and Y are the two variables and $\bar{X}, \bar{Y}$ are the **mean** of X and Y, respectively.

3.4.2 Variance Inflation Factor (VIF) and Tolerance

VIF quantifies multicollinearity in regression models by measuring how much the variance of a predictor variable is inflated due to its correlation with other variables. A high VIF (typically > 5 or 10) indicates strong multicollinearity, meaning the variable is highly redundant and may need to be removed. Conversely, low VIF values (< 5) suggest minimal correlation among predictors, which is ideal for regression models. Tolerance, defined as the inverse of VIF (Tolerance = 1/VIF), provides an additional measure of multicollinearity. A tolerance value below 0.2 or 0.1 indicates high collinearity, suggesting that the variable contributes little unique information and should be reconsidered for inclusion in the model.

3.4.3 Information Gain

Information Gain (IG) was used in decision trees and feature selection to measure how much a feature reduces uncertainty in predicting a target variable. It was calculated using entropy, where a feature with high information gain effectively splits the data, leading to a better classification. The higher the IG, the more useful the feature is for model performance.

3.5 Implication of Model

3.5.1 Logistic Regression (LR)

A logistic regression model is a statistical model that was used to predict the probability of a binary outcome based on one or more predictor variables. It is a generalization of the classical linear regression model and is commonly used in practice for interpreting the relationship between predictors and the outcome variable.

Equation 2 $Y = \{Logit\}(P) = ln(\frac{P}{1-P}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + e$

Equation 3 $LR\{\, logistic(y)\} = \frac{1}{1+e^{-y}}$

3.5.2 Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a linear approach for modeling the relationship between input parameters and resulting metrics. In recent studies, this method was applied to model and analyze groundwater recharge (Huang et al., 2019). Multivariate statistical analysis is widely used to test the reliability of different processes that affect the mineralization of the groundwater aquifer system.

Equation 4 $MLR = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n$

3.5.3 Random Forest

Random Forest is a highly powered ML algorithm for the classification and regression analysis. The algorithm was applied using the Python library and class in the Jupyter Notebook platform. The algorithm built 100 decision trees for delineation of ground water potential zones weightage. In addition, the weight of each influencing factor was calculated using the feature importance function in the Random Forest algorithm which was multiplied by their respective factors to gain the final GWPZs map.

The model was run by creating a Bootstrap dataset from original data by Randomly choosing data by allowing repetition. Then, created a randomized decision tree from these Bootstrap Datasets. Finally, the output of the random forest was the class selected by most trees.

3.6 Validation

3.6.1 AUC-ROC Curve

Area under the Curve (AUC) represents one of the most important measures of performance analysis for a binary classifier. The Receiver Operating Characteristic (ROC) curve is a graphical plot of the TPR against the FPR at different classification thresholds. It provides a visualization of the model's power of discrimination between positive and negative classes. A higher AUC value means better classification performance.

Equation 5

$TPR = \frac{TP}{FN+TP}$, $FPR = \frac{FP}{TN+FP}$, $TNR = \frac{TN}{TN+FP}$ where, TP= True Positive, FN= False Negative, TN= True Negative, FP= False Positive

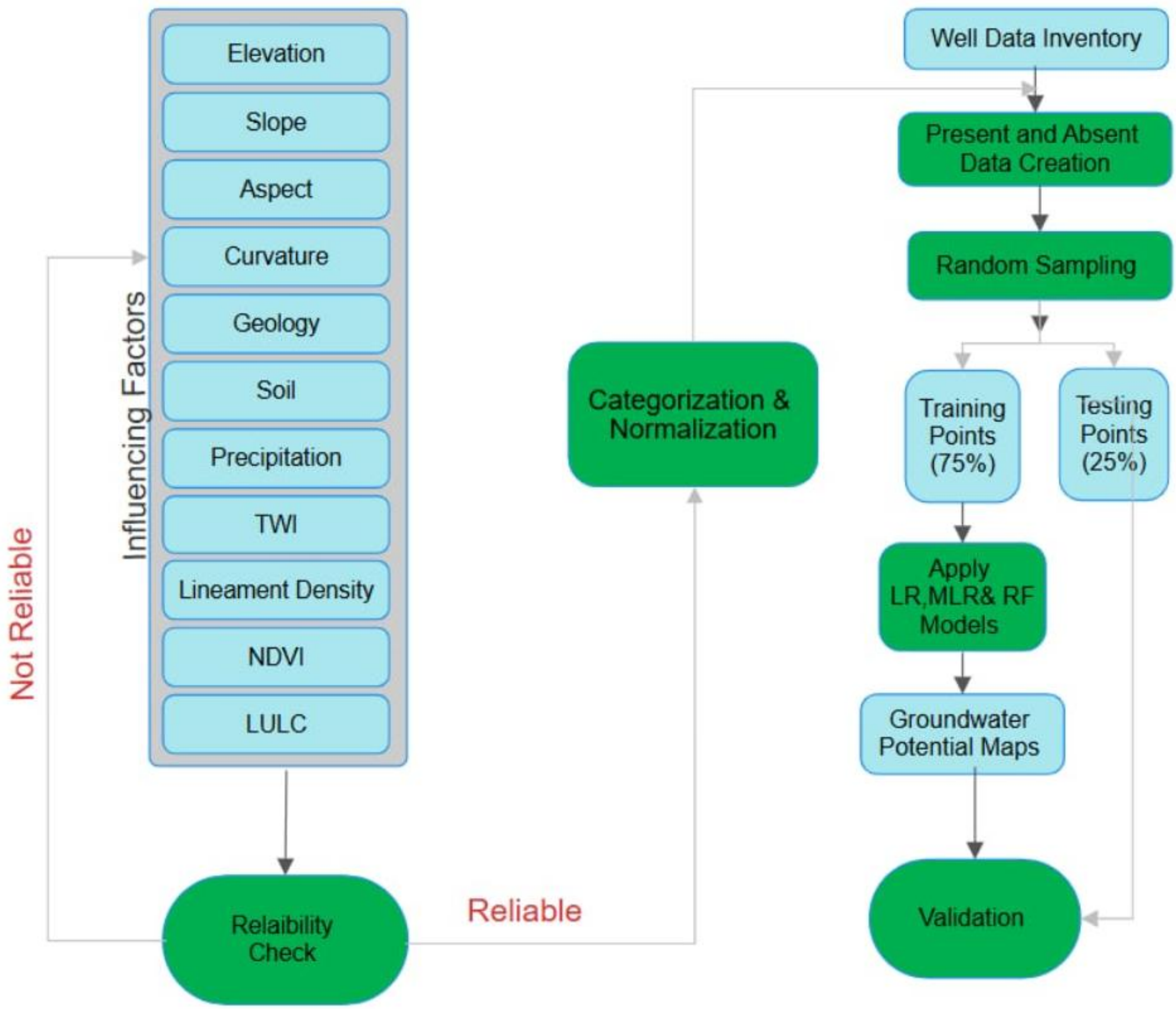

Figure 2: Workflow Diagram of the study

Chapter 4 Result and Analysis

4.1 Influencing Factors

All together 11 influencing factors were prepared using different available datasets and software. The influencing factors were prepared using ArcGIS 10@ESRI. The map is shown in below as:

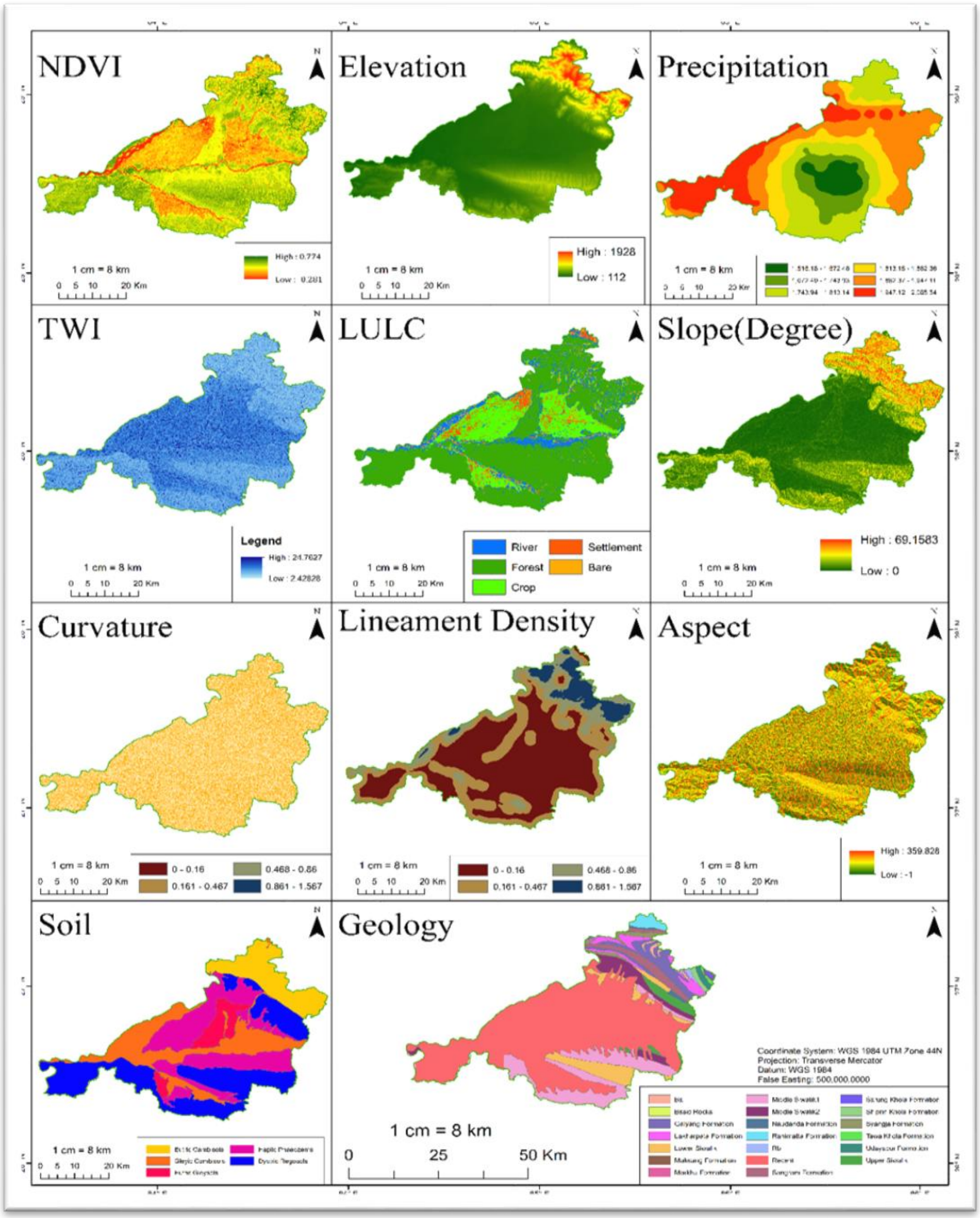

Figure 3: All influencing Factors

4.2 Reliability Test

4.2.1 Correlation matrix/Heat map

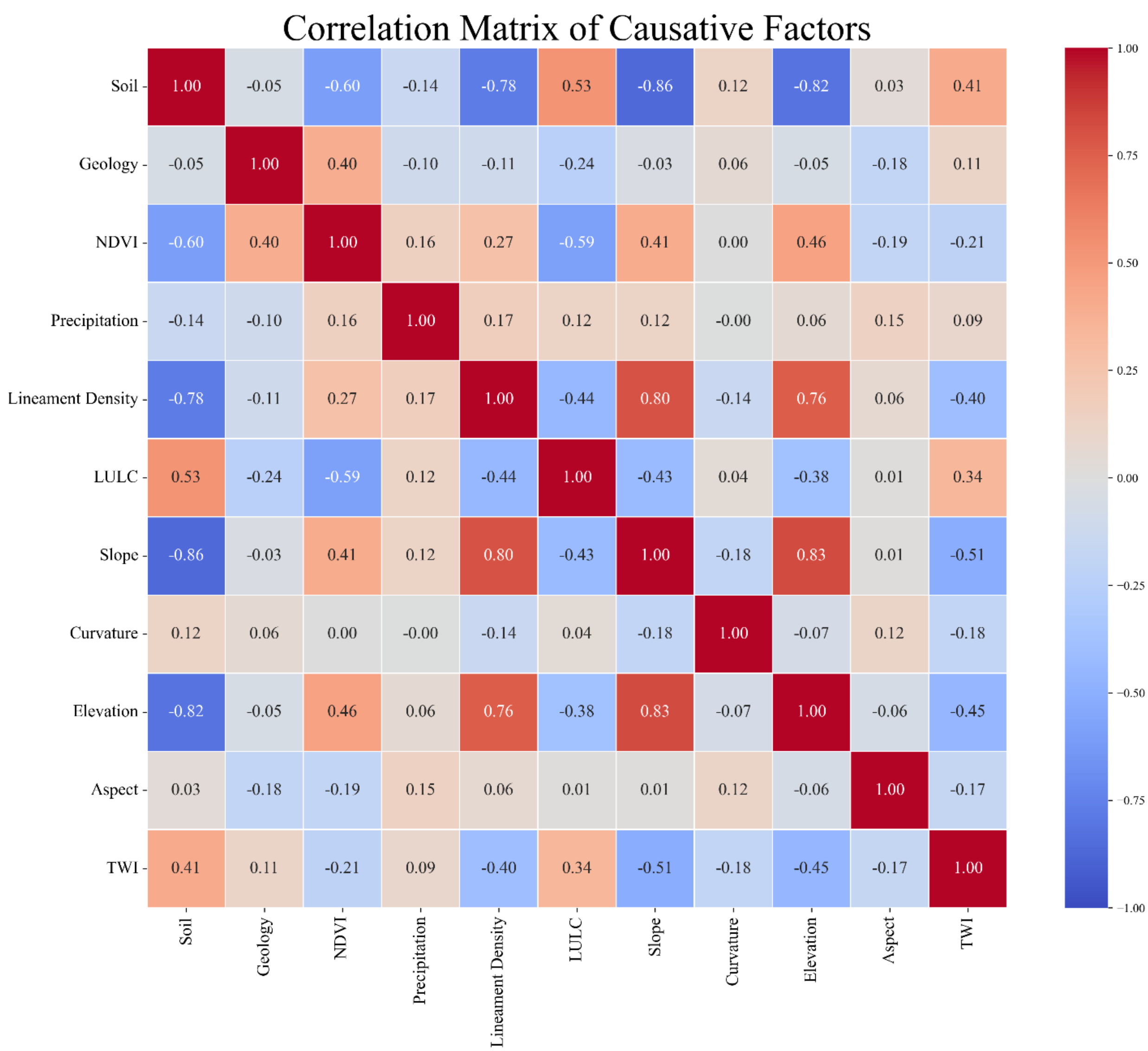

Figure 4: Pearson Correlation Coefficient/ Heat Map

The obtained Pearson correlation matrix highlights the relationships between various environmental and geological factors. Soil shows strong negative correlations with Lineament Density (-0.78), Slope (-0.86), and Elevation (-0.82), indicating that as these factors increase, soil characteristics tend to decrease. Conversely, Soil has a moderate positive correlation with LULC (0.53) and TWI (0.41), suggesting that land use and topography influence soil properties. Geology, on the other hand, has weak correlations with most factors, except for a moderate positive correlation with NDVI (0.40), implying that vegetation cover is somewhat influenced by geological conditions.

NDVI, a measure of vegetation health, is positively correlated with Slope (0.41) and Elevation (0.46), indicating that steeper and higher areas tend to have healthier vegetation. Precipitation shows weak correlations with most factors, suggesting it has a limited direct influence on the other variables in this dataset. Lineament Density and Slope are strongly positively correlated (0.80), indicating that areas with higher lineament density tend to have steeper slopes.

4.2.2 VIF and Tolerance

A VIF of more than 5 or a Tolerance of less than 0.2 typically suggests high multicollinearity. For this analysis, all of the variables exhibit moderate multicollinearity with VIF from 1.18 (Aspect) to 4.55 (Elevation). Elevation is the highest VIF, so it has a high variance with other predictors, and Aspect has the minimum multicollinearity. The constant term (const) has a very high VIF of 236.62, as is to be expected and not of concern because it is the intercept.

Overall, the results show that even in the presence of multicollinearity, its seriousness is not such that the variables should be removed since most VIF values are less than the critical value of 5. However, there is a probable requirement for re-checking Elevation and Slope (VIF = 3.41) with comparatively higher values of VIF since they would make regression models unreliable.

Table 2: VIF and Tolerance of Influencing Factors

Variable	VIF	Tolerance
const	236.6174579	0.00147794
Soil	3.925630887	0.144391178
Geology	1.407770676	0.710342968
NDVI	3.326939574	0.300576544
Precipitation	1.401077857	0.713736211
Lineament Density	2.337361203	0.230554928
LULC	2.277086463	0.439157676
Slope	3.405009184	0.156127801
Curvature	1.222552926	0.817960498
Elevation	4.54786435	0.219883427
Aspect	1.183224354	0.845148257
TWI	1.760466881	0.568031135

4.2.3 Information Gain

The Information Gain analysis reveals the relative importance of each feature in contributing to the predictive power of the model. LULC (Land Use/Land Cover) has the highest Information Gain (0.797), indicating it is the most influential feature, followed closely by Soil (0.732) and NDVI (0.612). These results suggest that land use, soil characteristics, and vegetation health are critical factors in the context of the analysis. Slope (0.447) and Geology (0.442) also show moderate importance, while TWI (Topographic Wetness Index) and Elevation contribute relatively less but still hold significance.

On the other hand, features such as Curvature, Precipitation, and Aspect exhibit low Information Gain values, with Aspect being the least influential (0.016). This implies that these factors have minimal impact on the model's predictive capability. The findings highlight the dominance of land use, soil properties, and vegetation indices in explaining the variability in the dataset, while topographic and climatic factors play a secondary role. These insights can guide feature selection for modeling, ensuring a focus on the most impactful variables.

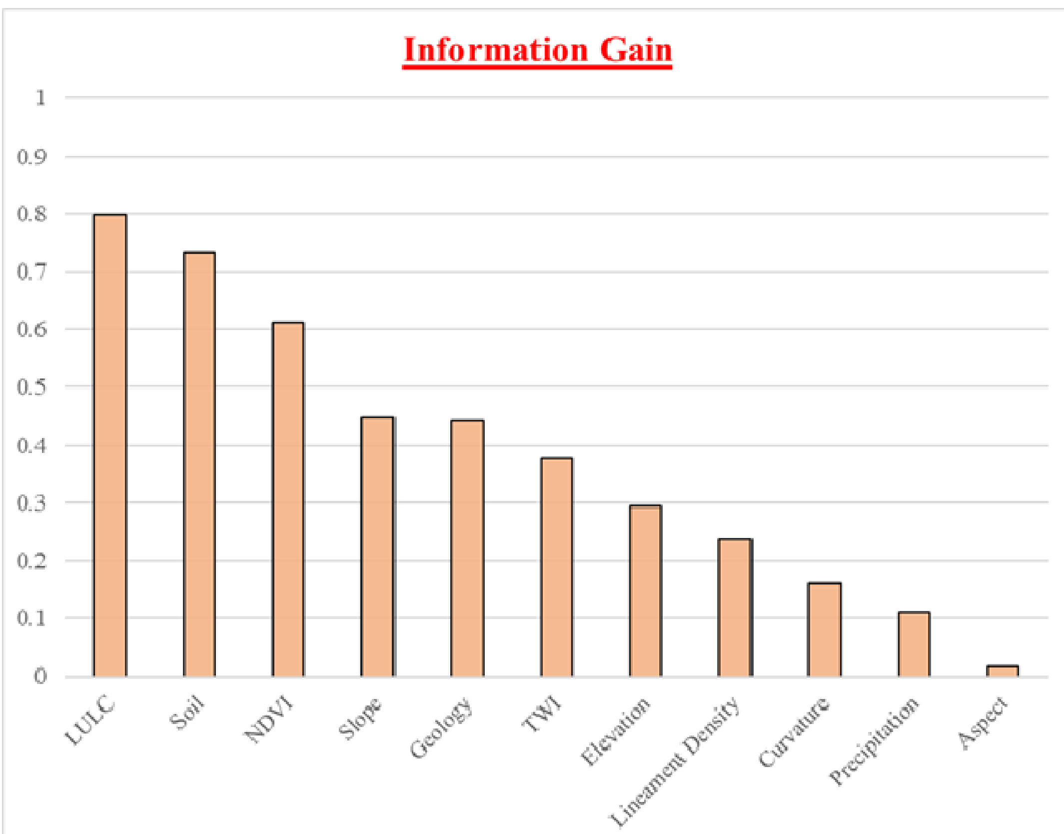

Figure 5: Feature Importance of all influencing factors

4.3 Implication of Model: FR

4.3.1 Logistic Regression (LR)

The logistic regression (LR) model result shows the impact of various factors on the outcome, for which there are statistically significant variables and non-statistically significant ones. Geology (p = 0.028), Precipitation (p = 0.042), and Slope (p = 0.035) are statistically significant, and negative coefficients show their negative association with the outcome. For instance, an increase in Slope or Precipitation lowers the chance of the predicted event. Although Soil, NDVI, Lineament Density, LULC, and others are not significant statistically (p > 0.05), they were included in the model for their hypothesized physical relevance from domain expertise. LULC and Lineament Density, for example, have positive coefficients, suggesting a potential positive impact on the outcome. Despite including non-significant variables in the model, it was used to develop a predictive map, capitalizing on both the statistical and physical data to provide an integrated picture of the determining factors. The approach guarantees that the model utilizes data-driven as well as theoretically relevant relationships even when some of the variables are not statistically significant.

LR={1 / (1 + exp(-(63.512 + (-0.197285785 * "Soil") + (-0.709 * "Geology") + (-12.52669181 * "NDVI") + (-0.0274 * "Precipitation") + (10.64 * "Lineament Density") + (3.513 * "LULC") + (-0.5135 * "Slope") + (-1.70e-10 * "Curvature") + (-0.004 * "Elevation") + (-0.0029 * "Aspect") + (-0.421* "TWI"))))}

Table 3: Summary of Logistic Regression

	Coefficient	Standard Error	z-value	p-value	Lower 95% CI	Upper 95% CI	Lower 97% CI	Upper 97% CI
const	63.5123	25.3042	2.5100	0.0121	13.9170	113.1075	8.6000	118.4246
Soil	-0.1973	0.9922	-0.1988	0.8424	-2.1420	1.7474	-2.3504	1.9559
Geology	-0.7091	0.3229	-2.1958	0.0281	-1.3420	-0.0762	-1.4099	-0.0083
NDVI	-12.5267	10.5684	-1.1853	0.2359	-33.2404	8.1871	-35.4612	10.4078
Precipitation	-0.0278	0.0136	-2.0363	0.0417	-0.0545	-0.0010	-0.0574	0.0018
Lineament Density	10.6415	5.8190	1.8288	0.0674	-0.7634	22.0465	-1.9862	23.2692
LULC	3.5133	1.9028	1.8463	0.0648	-0.2162	7.2428	-0.6160	7.6426
Slope	-0.5136	0.2441	-2.1042	0.0354	-0.9920	-0.0352	-1.0432	0.0161
Curvature	0.0000	0.0000	-1.0309	0.3026	0.0000	0.0000	0.0000	0.0000
Elevation	-0.0045	0.0049	-0.9199	0.3576	-0.0141	0.0051	-0.0151	0.0061
Aspect	-0.0029	0.0064	-0.4590	0.6462	-0.0154	0.0095	-0.0167	0.0109
TWI	-0.4219	0.4402	-0.9585	0.3378	-1.2846	0.4408	-1.3771	0.5333

4.3.2 Multiple Linear Regression (MLR)

The results of the multiple linear regression (MLR) model indicate the relationship between the target variable and various environmental variables. Precipitation and Geology are the significant variables since they possess p-values of 0.0281 and 0.0417, respectively, and have a substantial impact on the target. Lineament Density and LULC also show borderline significance (p-values of 0.0674 and 0.0648, respectively), that is, these could have some influence, but certainly not at very high levels of confidence. Whereas, aspects like NVDI, Curvature, Elevation, and Aspect have no noticeable effect on the target variable since their p-values are above the conventional 0.05.

$$MLR = 63.5123+(-0.1973\times Soil)+(-0.7091\times Geology)+(-12.5267\times NDVI)+(-0.0278\times Precipitation)+(10.6415\times Lineament\ Density)+(3.5133\times LULC)+(-0.5136\times Slope)+(0\times Curvature)+(-0.0045\times Elevation)+(-0.0029\times Aspect)+(-0.4219\times TWI)$$

4.3.3 Random Forest (RF)

Random Forest (RF) algorithm was used for establishing Groundwater Potential Zones (GWPZs) using training and influencing factor spatial datasets. The model was executed in Python inside Jupyter Notebook, where 100 decision trees were built. Groundwater Potential Zones map was prepared by considering the weightage of each influencing factor and multiplying it by its respective Feature Importance value, thus enabling one to identify areas with varying groundwater potential. This approach efficiently integrates different factors of influence to predict and map the groundwater potential in the study area.

4.4 Model Validation: AUC-ROC

Model validation is crucial in this study to assess the accuracy of the Groundwater Potential Zones (GWPZs) prediction. The main goal is to determine the likelihood of groundwater occurrence in specific areas, which is valuable for resource management, land use planning, and sustainable development. This evaluation involves comparing the predicted zones with real-world data to assess the model's effectiveness and accuracy. The result depicted that the AUC for the LR, MLR and RF models is 0.813 (81.3%), 0.846(84.6%) and 0.872 (87.22%) respectively. Therefore, evaluating the performance of the Random Forest (RF) model is essential to ensure the reliability of the predicted GWPZs.

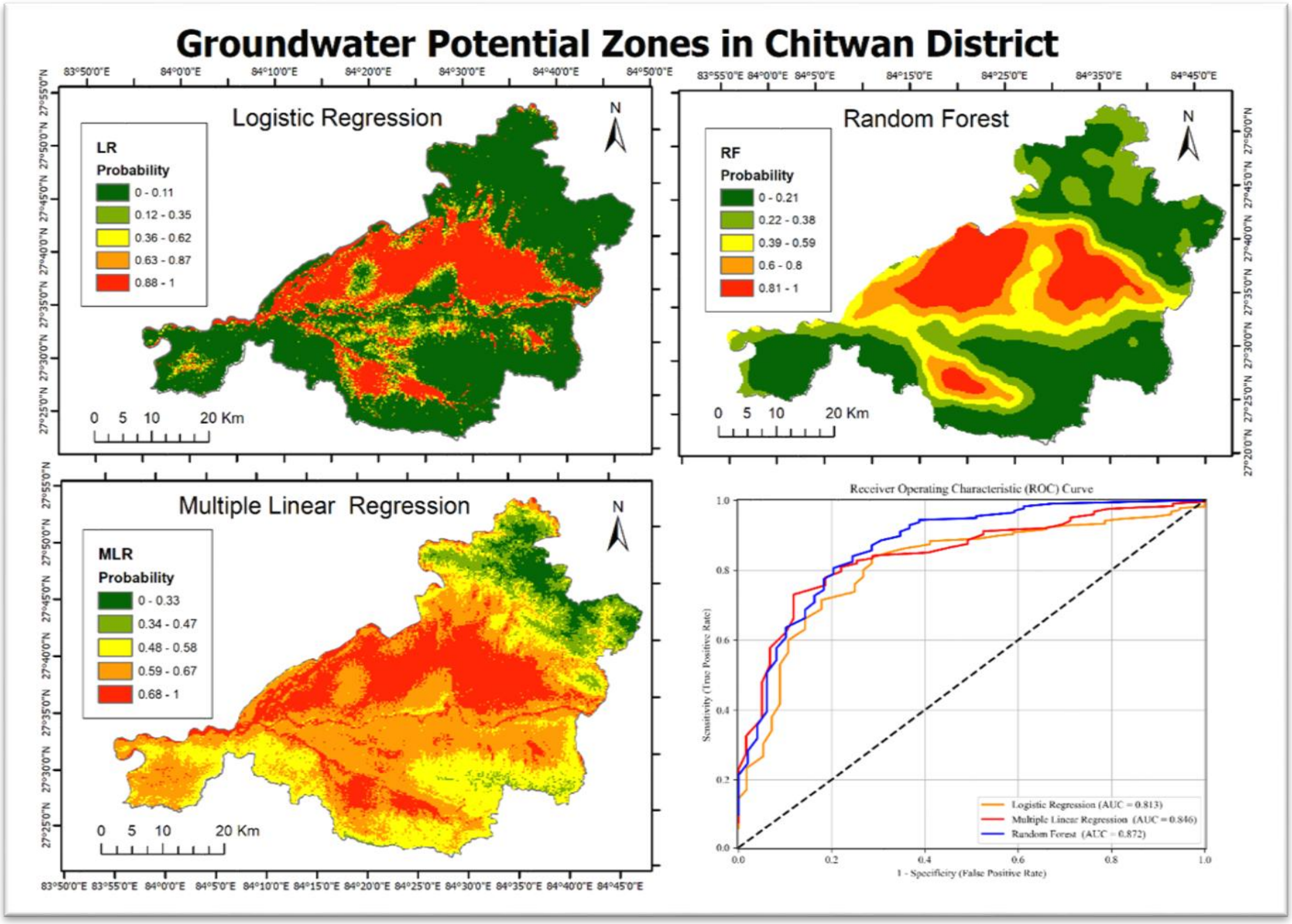

Figure 6: Ground Water Potential Zones Mapping using LR, MLR and RF model and AUC-ROC curve of respective models.

Chapter 5 Discussion

In the present study, the delineation of Groundwater Potential Zones (GWPZs) was achieved by applying three diverse models: Random Forest (RF), Logistic Regression (LR), and Multiple Linear Regression (MLR). To train and test the models, 600 well points (300 present and 300 absent) were collected from the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board and OpenStreetMap. The data was sampled using Random Sampling technique in the ratio of 80:20 for training and testing. Altogether, ten factors, i.e., soil, geology, precipitation, and land use, were considered for analysis. Three factor-screening methods, i.e., Information Gain (IG), Pearson Correlation Coefficient (PCC), and Variance Inflation Factor (VIF) along with Tolerance, were employed to select the most effective factors for groundwater potential mapping. The results indicated that the RF model outperformed the LR and MLR models, with an Area under the Curve (AUC) of 87.2%, indicating its high accuracy in delineating GWPZs. Slope is another important factor that controls rate of infiltration and run-off in any part of the globe(Hasanuzzaman et al., 2022). Higher slope adversely effects on groundwater storage; thereby, groundwater potential zones are generally associated with lower slope region (Maskooni et al., 2020). The paper (Nguyen et al., 2020) states that the Precipitation and LULC is a major factors for GWPZs mapping. Therefore, according to those authors our study also resemble the LULC is second highest factors for delineation of GWPZs and similarly, the slope has third highest feature importance value and the both LULC and slope is considered to be statistically significant influencing factors for the study.

The paper by (Todd & Mays, 2005) emphasize the importance of taking into account primary variables such as slope, geology, and precipitation, which are typically statistically significant and have direct impacts on groundwater recharge and storage. Their study identifies that the inclusion of irrelevant or weakly correlated variables can introduce noise and reduce model accuracy. However, the elimination of physically significant but statistically non-significant variables has been questioned (Freeze & Cherry, 1979) argue that groundwater systems are complex and are influenced by a wide range of factors, some of which may not have high statistical correlations but are physically significant. So, our study also ensemble all the influencing factors considering their both statistically and physically significance.

The paper (Nguyen et al., 2020) study the ground water potential zones using LR and its four ensemble technique i.e. BLR, CGLR, DLR and RSSLR. The paper includes the influencing factors similar to our research except including STI and excluding the lineament density. The authors validated using AUC-ROC Curve. The value of AUC is low in LR model while increasing significantly in other that four-ensemble technique of LR. Our research also concluded that the value of AUC of LR model seems lower in comparison to other models. This recommend that the single LR is not solely sufficient for the delineation of GWPZs mapping.

In the study (Dargahi et al., 2022), the multiple linear regression model was used to predict groundwater quality parameters, with a 5% level of significance. The p-value of influencing factors were studied and the factors below 5% were refined for delineation of GWPZs. Meanwhile, we considered all the influencing factors for the result of GWPZs mapping considering other influencing factors could stands for their physical importance.

In the context of Groundwater Potential Zone (GWPZ) mapping, several studies have utilized various machine learning algorithms, including Random Forest (RF), to predict groundwater potential. RF is widely praised for its ability to handle complex, non-linear relationships between influencing factors and groundwater occurrence. For instance, a study by (Rahaman et al., 2022) demonstrated the application of RF to predict groundwater potential in India, using factors such as geology, rainfall, land use, and slope. Their findings showed that RF outperformed traditional statistical models in terms of accuracy, highlighting the algorithm’s robustness in dealing with high-dimensional data and its capacity to handle overfitting. Similarly, (Jeihouni et al., 2020) applied RF in predicting GWPZs in arid regions of India, reporting high accuracy and emphasizing the importance of incorporating diverse spatial datasets to capture the complex nature of groundwater distribution. These studies align with our methodology, where multiple influencing factors like precipitation, slope, geology, and land use were integrated into the model.

Chapter 6 Limitation

Despite the comprehensive approach taken in this study, several limitations must be mentioned. Firstly, the final groundwater potential zone (GWPZ) map resolution is only 30 meters, which may not capture finer-scale characteristics or localized groundwater systems that are significant for accurate assessment. This limitation can potentially underestimate groundwater potential in areas of high variability at finer scales. Second, field data collection was limited by a shortage of data, with insufficient data, relying primarily on online dashboards such as the Ministry of Water Resources, Energy and Irrigation Groundwater Resources Development Board and OpenStreetMap. This lack of ground-truth data can affect the validity and accuracy of the model because field verification is essential for predictive map validation. Third, the study was based on limited training data, which could restrict the ability of the model to generalize across diverse hydrogeological settings. Finally, while the factors employed in the analysis were aggregated at a coarse level, more detailed investigation of sub-categories (e.g., fine soil classes, specific geological units, or detailed land use categories) would enhance the accuracy of the model and provide more detailed information about groundwater potential. Furthermore, the study assumes static influencing factors, while in reality, groundwater dynamics are affected by seasonal fluctuations, climate variability, and anthropogenic influences, which were not accounted for in the present analysis. Overcoming such shortcomings in later studies, i.e., in more high-resolution data, in large field surveys, and more elaborate factor analysis, would make a significant contribution towards the accuracy and applicability of GWPZ mapping.

Acknowledgement

This research was supported by the “Research Management Unit: Research Culture in Campus”. We are thankful to acknowledge the gratitude for their financial and technical assistance, through which this research could be conducted. Finally yet importantly, we value the suggestion and encouragement from my mentors and colleagues, whose views and criticism assisted in the success of this research.

Chapter 7 Conclusion

The study successfully delineated Groundwater Potential Zones (GWPZs) in Nepal's Chitwan District using Geographic Information System (GIS), Remote Sensing (RS), and three models: Random Forest (RF), Logistic Regression (LR), and Multiple Linear Regression (MLR). 600 well points (300 existing and 300 non-existing) dataset was retrieved from authentic sources and utilized for model training and validation. Eleven critical environmental and geological factors, such as elevation, slope, and precipitation, were assessed and maximized by utilizing Information Gain (IG), Pearson Correlation Coefficient (PCC), and Variance Inflation Factor (VIF) with Tolerance. The final GWPZs map was classified into five classes—Very Low, Low, Moderate, High, and Very High—utilizing the Natural Breaks Classification method. Validation of the model employing the Area Under the Curve (AUC) method showed that RF performed better than the other models with a performance of 87.2%, followed by MLR (84.6%) and LR (81.3%). RF's superior performance is a testament to its ability to detect complex relationships between the influencing factors, thus a reliable tool for groundwater potential mapping. Despite its sucess, the study is not limitation-free. The quality and well-distribution of the well point data were low in availability, and this would have affected the model quality. Also, ground verification of the classified zones was not carried out due to time constraints, and thus, the actual-world applicability was somewhat constrained. The research also assumes the influencing factors are constant, when actually the dynamics of groundwater depend on seasonal factors, climatic fluctuation, and anthropogenic factors which have not been incorporated. Subsequent research ought to concentrate on adding high-resolution information, doing ground validation, and combining further hydrological and climatic factors, like recharge rates and aquifer properties, in order to strengthen model robustness. Investigation into ensemble modeling methods based on using different machine learning techniques together would continue to increase predictability precision and applicability to a range of different hydrogeological contexts. In conclusion, this study provides a valuable model of groundwater resource management for Chitwan that offers practical recommendations to policymakers, water resource managers, and urban planners with certain limitations and futuredirections. Utilizing cutting-edge machine learning algorithms for geospatial tools has the high potential to further the groundwater assessment and groundwater management in water-scarce regions.

References

Acharya, T., Kumbhakar, S., Prasad, R., Mondal, S., & Biswas, A. (2019). Delineation of potential groundwater recharge zones in the coastal area of north-eastern India using geoinformatics. Sustainable Water Resources Management, 5(2), 533–540. https://doi.org/10.1007/S40899-017-0206-4

Arabameri, A., Pal, S. C., Rezaie, F., Nalivan, O. A., Chowdhuri, I., Saha, A., Lee, S., & Moayedi, H. (2021). Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. Journal of Hydrology: Regional Studies, 36. https://doi.org/10.1016/J.EJRH.2021.100848

Arabameri, A., Roy, J., Saha, S., Blaschke, T., Ghorbanzadeh, O., & Bui, D. T. (2019). Application of probabilistic and machine learning models for groundwater potentiality mapping in Damghan sedimentary plain, Iran. Remote Sensing, 11(24). https://doi.org/10.3390/rs11243015

Champak, *, Silwal, B., & Pathak, D. (2018). Review on Practices and State of the Art Methods on Delineation of Ground Water Potential Using GIS and Remote Sensing. Bulletin of the Department of Geology, 7–20. https://doi.org/10.3126/BDG.V20I0.20717

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. Ensemble Machine Learning, 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5

Dargahi, P., Nasseri, S., Hadi, M., Nodehi, R. N., & Mahvi, A. H. (2022). Prediction models for groundwater quality parameters using a multiple linear regression (MLR): a case study of Kermanshah, Iran. Journal of Environmental Health Science and Engineering, 21(1), 63. https://doi.org/10.1007/S40201-022-00836-9

Freeze, R. A., & Cherry, J. A. (1979). Groundwater. Groundwater. https://books.google.com/books/about/Groundwater.html?id=feVOAAAAMAAJ

Guru, B., Seshan, K., & Bera, S. (2017). Frequency ratio model for groundwater potential mapping and its sustainable management in cold desert, India. Journal of King Saud University - Science, 29(3), 333–347. https://doi.org/10.1016/j.jksus.2016.08.003

Hasanuzzaman, M., Mandal, M. H., Hasnine, M., & Shit, P. K. (2022). Groundwater potential mapping using multi-criteria decision, bivariate statistic and machine learning algorithms:

evidence from Chota Nagpur Plateau, India. Applied Water Science, 12(4), 1–16. https://doi.org/10.1007/s13201-022-01584-9

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (n.d.). Applied logistic regression. Retrieved July 25, 2024, from https://books.google.com/books/about/Applied_Logistic_Regression.html?id=bRoxQBIZRd4C

Huang, X., Gao, L., Crosbie, R. S., Zhang, N., Fu, G., & Doble, R. (2019). Groundwater recharge prediction using linear regression, multi-layer perception network, and deep learning. Water (Switzerland), 11(9). https://doi.org/10.3390/w11091879

Jasrotia, A. S., Kumar, A., & Singh, R. (2016). Integrated remote sensing and GIS approach for delineation of groundwater potential zones using aquifer parameters in Devak and Rui watershed of Jammu and Kashmir, India. Arabian Journal of Geosciences, 9(4), 1–15. https://doi.org/10.1007/S12517-016-2326-9/METRICS

Jeihouni, M., Toomanian, A., & Mansourian, A. (2020). Decision Tree-Based Data Mining and Rule Induction for Identifying High Quality Groundwater Zones to Water Supply Management: a Novel Hybrid Use of Data Mining and GIS. Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA), 34(1), 139–154. https://doi.org/10.1007/S11269-019-02447-W

Kansakar, D. R. (2016). Understanding Groundwater for Proper Utilization and Management in Nepal.

Karimi-Rizvandi, S., Goodarzi, H. V., Afkoueieh, J. H., Chung, I. M., Kisi, O., Kim, S., & Linh, N. T. T. (2021). Groundwater-Potential Mapping Using a Self-Learning Bayesian Network Model: A Comparison among Metaheuristic Algorithms. Water 2021, Vol. 13, Page 658, 13(5), 658. https://doi.org/10.3390/W13050658

Karki, S. (2022). A Comparative Study of Landslide Susceptibility Mapping along Mountainous Road using Statistical and Machine Learning Approach. 8914, 396–405.

Khadka, G., & Pathak, D. (2021). Groundwater potential as an indicator of water poverty index in drought-prone mid-hill region of Nepal Himalaya. Groundwater for Sustainable Development, 12, 100502. https://doi.org/10.1016/J.GSD.2020.100502

Kim, J. C., Jung, H. S., & Lee, S. (2019). Spatial mapping of the groundwater potential of the Geum River basin using ensemble models based on remote sensing images. Remote Sensing, 11(19). https://doi.org/10.3390/rs11192285

Kumar, B., & Kumar, U. (2011). Ground water recharge zonation mapping and modeling using Geomatics techniques. International Journal of Environmental Science, 1(7), 1670–1681.

Lee, S., Hong, S. M., & Jung, H. S. (2018). GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. Geocarto International, 33(8), 847–861. https://doi.org/10.1080/10106049.2017.1303091

Lee, S., Jang, J., Kim, Y., Cho, N., & Lee, M. J. (2020). Susceptibility Analysis of the Mt. Umyeon Landslide Area Using a Physical Slope Model and Probabilistic Method. Remote Sensing 2020, Vol. 12, Page 2663, 12(16), 2663. https://doi.org/10.3390/RS12162663

Lee, S., & Talib, J. A. (2005). Probabilistic landslide susceptibility and factor effect analysis. Environmental Geology, 47(7), 982–990. https://doi.org/10.1007/S00254-005-1228-Z/METRICS

Malla, R., & Karki, K. (n.d.). ENVIRONMENT.

Manap, M. A., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W. N. A., & Ramli, M. F. (2014). Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. Arabian Journal of Geosciences, 7(2), 711–724. https://doi.org/10.1007/S12517-012-0795-Z/METRICS

Martínez-Santos, P., & Renard, P. (2020). Mapping Groundwater Potential Through an Ensemble of Big Data Methods. Groundwater, 58(4), 583–597. https://doi.org/10.1111/GWAT.12939

Maskooni, E. K., Naghibi, S. A., Hashemi, H., & Berndtsson, R. (2020). Application of Advanced Machine Learning Algorithms to Assess Groundwater Potential Using Remote Sensing-Derived Data. Remote Sensing 2020, Vol. 12, Page 2742, 12(17), 2742. https://doi.org/10.3390/RS12172742

Moench, M. (2003). Groundwater and poverty: exploring the connections. Intensive Use of Groundwater: Challenges and Opportunities, 441–456.

Mogaji, K. A., Lim, H. S., & Abdullah, K. (2015). Modeling of groundwater recharge using a multiple linear regression (MLR) recharge model developed from geophysical parameters: a case of groundwater resources management. Environmental Earth Sciences, 73(3), 1217–1230. https://doi.org/10.1007/s12665-014-3476-2

Murmu, P., Kumar, M., Lal, D., Sonker, I., & Singh, S. K. (2019). Delineation of groundwater potential zones using geospatial techniques and analytical hierarchy process in Dumka district, Jharkhand, India. Groundwater for Sustainable Development, 9, 100239. https://doi.org/10.1016/J.GSD.2019.100239

Naghibi, S. A., & Pourghasemi, H. R. (2015). A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resources Management, 29(14), 5217–5236. https://doi.org/10.1007/S11269-015-1114-8/METRICS

Naghibi, S. A., Pourghasemi, H. R., & 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, 188(1), 1–27. https://doi.org/10.1007/S10661-015-5049-6/METRICS

Nathan, N. S., Saravanane, R., & Sundararajan, T. (2017). Application of ANN and MLR Models on Groundwater Quality Using CWQI at Lawspet, Puducherry in India. Journal of Geoscience and Environment Protection, 05(03), 99–124. https://doi.org/10.4236/GEP.2017.53008

Nguyen, P. T., Ha, D. H., Avand, M., Jaafari, A., Nguyen, H. D., Al-Ansari, N., Phong, T. Van, Sharma, R., Kumar, R., Le, H. Van, Ho, L. S., Prakash, I., & Pham, B. T. (2020). Soft computing ensemble models based on logistic regression for groundwater potential mapping. Applied Sciences (Switzerland), 10(7). https://doi.org/10.3390/app10072469

Pan, C., Ng, K. T. W., Fallah, B., & Richter, A. (2019). Evaluation of the bias and precision of regression techniques and machine learning approaches in total dissolved solids modeling of an urban aquifer. Environmental Science and Pollution Research, 26(2), 1821–1833. https://doi.org/10.1007/S11356-018-3751-Y/METRICS

Panday, D. (2023). Recharging groundwater. The Kathmandu Post. https://kathmandupost.com/columns/2023/12/16/recharging-groundwater

Pande, C. B., Moharir, K. N., Panneerselvam, B., Singh, S. K., Elbeltagi, A., Pham, Q. B., Varade, A. M., & Rajesh, J. (2021). Delineation of groundwater potential zones for sustainable development and planning using analytical hierarchy process (AHP), and MIF techniques. Applied Water Science, 11(12), 1–20. https://doi.org/10.1007/S13201-021-01522-1/FIGURES/8

Park, S., Hamm, S. Y., Jeon, H. T., & Kim, J. (2017). Evaluation of Logistic Regression and Multivariate Adaptive Regression Spline Models for Groundwater Potential Mapping Using R and GIS. Sustainability 2017, Vol. 9, Page 1157, 9(7), 1157. https://doi.org/10.3390/SU9071157

Pathak, D. (2017). Delineation of Groundwater Potential Zone in the Indo-Gangetic Plain Through Gis Analysis. Journal of Institute of Science and Technology, 22(1), 104–109. https://doi.org/10.3126/jist.v22i1.17760

Pradhananga, S. M., Shrestha, S. Das, & Steenhuis, T. S. (2012). Comprehensive review of groundwater research in the Kathmandu valley, Nepal. Kathmandu Valley Groundwater Outlook, February, 6–16. www.smallearth.org.np

Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. GIScience and Remote Sensing, 735–752. https://doi.org/10.1080/15481603.2020.1794104

Rahaman, S., Saha, S., & Das, M. (2022). Delineation of Groundwater Potential Zones Using Analytic Hierarchy Process (AHP) Technique in Balason River Basin of West Bengal, India Bank erosion of River Ganga View project Delineation of Groundwater Potential Zones Using Analytic Hierarchy Process (. Riverine Environment and Its Dynamics: Challenges, Issues and Sustainable Managements, December, 16–47. https://www.researchgate.net/publication/367362274

Rahman, M. M., Althobiani, F., Shahid, S., Virdis, S. G. P., Kamruzzaman, M., Rahaman, H., Momin, M. A., Hossain, M. B., & Ghandourah, E. I. (2022). GIS and Remote Sensing-Based Multi-Criteria Analysis for Delineation of Groundwater Potential Zones: A Case Study for Industrial Zones in Bangladesh. Sustainability (Switzerland), 14(11). https://doi.org/10.3390/su14116667

Rahmati, O., Nazari Samani, A., Mahdavi, M., Pourghasemi, H. R., & Zeinivand, H. (2015). Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and

GIS. Arabian Journal of Geosciences, 8(9), 7059–7071. https://doi.org/10.1007/S12517-014-1668-4/METRICS

Razandi, Y., Pourghasemi, H. R., Neisani, N. S., & Rahmati, O. (2015). Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. Earth Science Informatics, 8(4), 867–883. https://doi.org/10.1007/S12145-015-0220-8/METRICS

Reif, D. M., Motsinger, A. A., McKinney, B. A., Crowe, J. E., & Moore, J. H. (2006). Feature selection using a random forests classifier for the integrated analysis of multiple data types. Proceedings of the 2006 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB'06, 171–178. https://doi.org/10.1109/CIBCB.2006.330987

Sameen, M. I., Pradhan, B., & Lee, S. (2019). Self-Learning Random Forests Model for Mapping Groundwater Yield in Data-Scarce Areas. Natural Resources Research, 28(3), 757–775. https://doi.org/10.1007/S11053-018-9416-1

Sarkar, S. K., Esraz-Ul-Zannat, M., Das, P. C., & Mohiuddin Ekram, K. M. (2022). Delineating the groundwater potential zones in Bangladesh. Water Supply, 22(4), 4500–4516. https://doi.org/10.2166/WS.2022.113

Shiwakoti, S., Aryal, S., Giri, S., Pandit, S., Timsina, S., & Agrawal, T. (n.d.). DEPARTMENT OF CIVIL ENGINEERING FINAL YEAR PROJECT REPORT on " DELINEATION OF GROUND WATER RECHARGE POTENTIAL ZONES OF CHITWAN DISTRICT ." April 2023.

Shrestha, S. R., Tripathi, G. N., & Laudari, D. (2018). Groundwater Resources of Nepal: An Overview. 169–193. https://doi.org/10.1007/978-981-10-3889-1_11

Silwal, C. B., Nepal, M., Pathak, D., Karkee, B., Dahal, K., & Acharya, S. (2023). Groundwater potential zonation in the Siwalik of the Kankai River Basin, Eastern Nepal. Water Supply, 23(6), 2332–2348. https://doi.org/10.2166/ws.2023.137

Thanh, N. N., Chotpantarat, S., Trung, N. H., Ngu, N. H., & Muoi, L. Van. (2022). Mapping groundwater potential zones in Kanchanaburi Province, Thailand by integrating of analytic hierarchy process, frequency ratio, and random forest. Ecological Indicators, 145, 109591.

https://doi.org/10.1016/J.ECOLIND.2022.109591

Todd, D. K., & Mays, L. W. . (2005). Groundwater hydrology. 636. https://www.wiley.com/en-us/Groundwater+Hydrology%2C+3rd+Edition-p-9780471059370

Treidel, H., Martin-Bordes, J. L., & Gurdak, J. J. (2012). Climate Change Effects on Groundwater Resources. INTERNATIONAL ASSOCIATION OF HYDROGEOLOGISTS.

Vaux, H. (2011). Groundwater under stress: The importance of management. Environmental Earth Sciences, 62(1), 19–23. https://doi.org/10.1007/S12665-010-0490-X/METRICS