

A GIS-Based Logistic Regression Model for Groundwater Spring Potential Mapping in the Gamri Chhu Basin, Bhutan

Ghalley, R. K.,^{1*} Tantanee, S.,¹ Nusit, K.,¹ Chidburee, P.,¹ Soonthornnonda, P.² and Tai, C. H.³

¹Department of Civil Engineering, Faculty of Engineering, Naresuan University, Phitsanulok 65000, Thailand

E-mail: rameshk67@nu.ac.th,* ORCID ID: 0009-0006-0886-0971

E-mail: sarintipt@nu.ac.th, ORCID ID: 0000-0002-4717-5209

E-mail: korakodn@nu.ac.th, ORCID ID: 0000-0002-6702-5080

E-mail: polpreechac@nu.ac.th

²Department of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology,

Ladkrabang, Bangkok 10520, Thailand, E-mail: puripus.so@kmitl.ac.th, ORCID ID: 0009-0005-6707-480X

³Department of Sustainability and Disaster Management, Tzu Chi University, Hualien, Taiwan

E-mail: chtai0109@gms.tcu.edu.tw

*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v22i5.4979>

Abstract

Groundwater springs constitute the predominant water supply for upland communities in Bhutan's Hindu Kush Himalaya (HKH) region. However, 25.1% of springs in Bhutan are currently reported as drying, underscoring the need for accurate mapping of spring potential zones to support conservation and recharge planning. This study employed a logistic regression (LR) model to delineate groundwater spring potential zones in the Gamri Chhu Basin of eastern Bhutan. An inventory of 145 spring locations was compiled and paired with an equal number of pseudo-absence points generated using a 500 m exclusion buffer around known spring locations. Eight environmental conditioning factors altitude, slope, geology, drainage density, lineament density, land use, soil type, and precipitation were evaluated as spatial predictors. Correlation and multicollinearity analyses were performed to ensure model reliability. The LR model was trained using 70% of the dataset and validated with the remaining 30%. Results indicated that altitude, slope, geology, drainage density, precipitation, and soil type significantly influenced spring occurrence ($p < 0.05$). The model achieved good predictive performance, with AUC values of 0.86 (training) and 0.85 (validation). The resulting map was classified into low (81.34%), moderate (11.80%), and high (6.87%) potential zones. SDI values increased progressively from low- to high-potential zones, with the highest values observed in the high-potential class (6.00 for training and 5.08 for validation), indicating strong spatial agreement between predicted and observed spring locations. The resulting map serves as a preliminary, regional-scale screening tool to prioritize areas for field verification, recharge zone investigation, and informed groundwater assessment in data-limited mountainous regions.

Keywords: GIS, Groundwater Springs, Logistic Regression, Spring Potential Mapping, Springshed Management, Eastern Bhutan, Hindu Kush Himalaya

1. Introduction

Natural groundwater springs serve as the primary water source for mid-hill communities across the Hindu Kush Himalayan region, as the upslope settlement pattern and rugged terrain make access to fast-flowing rivers in the valley bottoms difficult [1] [2] and [3]. These springs are often the only reliable source for domestic and irrigation purposes, thereby sustaining households, agriculture, and local ecosystems in the region [4] [5] and [6]. On the other

hand, many reports have pointed out that spring flow has decreased and seasonal drying has become more widespread, thus raising the issue of water security and climate resilience for upland communities significantly [6] [7] and [8]. In the Indian Himalayan region alone, nearly half of the 3 million springs have either dried up or exhibited reduced flow [7]. Bhutan, as part of the HKH, is also experiencing spring depletion. A recent national inventory identified

7,399 water sources in Bhutan, of which 67.5% are springs. Notably, 25.1% of these springs are currently drying, and 0.9% have already dried completely [9]. This severe decline threatens the very survival of these hill communities.

Identifying areas with a high potential for groundwater spring occurrence is a key first step in managing springsheds, especially for planning rejuvenation and protection strategies [3][10] and [11]. This aligns with the widely adopted springshed management framework proposed by the International Centre for Integrated Mountain Development (ICIMOD), which emphasizes sequential steps such as spring mapping, monitoring, governance understanding, recharge zone identification, implementation of recharge measures, and impact evaluation [12]. Among these, spring mapping and identification of recharge zones through spatial modeling provide a critical scientific basis for evidence-based planning and decision-making. Traditional field-based approaches for groundwater exploration, such as hydrogeological investigations, geophysical and geochemical surveys, and drilling, are often impractical in large, rugged mountainous regions due to high costs, time constraints, and limited accessibility [13] [14] and [15]. To address these limitations, Geographic Information Systems (GIS) and remote sensing-based spatial modeling approaches have emerged as efficient alternatives, enabling large-scale analysis of groundwater and spring potential even in data-scarce and difficult-to-access environments [16].

Within this context, groundwater potential mapping has largely been conducted using knowledge-driven Multi-Criteria Decision Analysis (MCDA) techniques, particularly the Analytic Hierarchy Process (AHP), where factor weights are assigned based on expert judgment [17]. While AHP is valuable in data-limited environments due to its ability to incorporate expert knowledge, it is inherently subjective and may introduce bias [18]. In parallel with these approaches, recent studies have increasingly adopted data-driven machine learning (ML) techniques, such as Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), and artificial neural networks (ANN), which are capable of capturing complex non-linear relationships and often achieve high predictive performance [19] [20] [21] [22] [23] and [24]. However, despite their accuracy, these models frequently operate as “black-box” systems with limited interpretability, which can constrain their applicability in environmental decision-making contexts [25].

In this study, Logistic Regression (LR) was adopted as a data-driven statistical approach for groundwater spring potential mapping. LR quantifies the relationship between multiple environmental conditioning factors and the probability of spring occurrence, allowing direct interpretation of the influence and significance of each predictor [26]. While advanced ML models are increasingly used, they typically require large training datasets and offer limited transparency in explaining model behavior. In contrast, LR provides a balance between predictive capability and interpretability, making it particularly suitable for mountainous, data-limited environments such as the Gamri Chhu Basin, where understanding the controlling hydrogeological processes is as important as prediction accuracy. Previous studies conducted in complex terrains, including the Sultan Mountains in Turkey [27], the Kashmir Himalaya [28], and the High Atlas Mountains, Morocco [29], have demonstrated the effectiveness and reliability of LR for groundwater and spring potential mapping. In Bhutan, research on groundwater and spring potential mapping remains limited. Existing studies have primarily focused on groundwater prospect zonation using knowledge-driven approaches such as AHP [30] [31] and [32]. Applications of statistical modeling frameworks for spring potential mapping in Bhutan remain limited.

Accordingly, this study presents one of the first applications of a GIS-based logistic regression model for groundwater spring potential mapping in the Gamri Chhu Basin of eastern Bhutan. The analysis focuses on identifying the key environmental factors controlling spring occurrence and evaluating the predictive performance of the LR model. The resulting groundwater spring potential map is intended as a preliminary, regional-scale screening tool to support springshed management, recharge planning, and water resource decision-making.

2. Study Area

Research was conducted in the Gamri Chhu Basin, located in Trashigang Dzongkhag (district) in eastern Bhutan (Figure 1). This basin's geographical coordinates are roughly between 27°14'–27°29' N and 91°33'–92°02' E. Gamri Chhu is one of the principal sub-basins of the Drangme Chhu river system, which eventually flows into the Brahmaputra [33]. The total area of the basin is around 734 km². Elevation ranges from below 655 m in deeply incised river valleys to over 4,398 m in the headwater regions, resulting in pronounced topographic, climatic, and hydrological gradients.

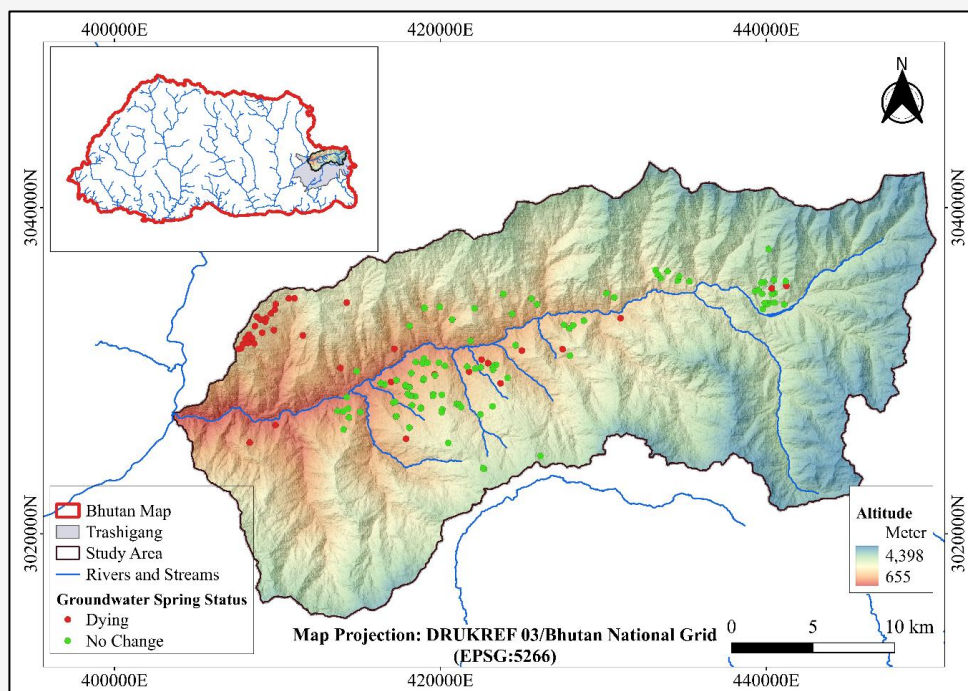


Figure 1: Gamri Chhu Basin in Trashigang Dzongkhag, Bhutan

The hydroclimate of the basin is dominated by the South Asian monsoon, with the majority of annual precipitation occurring between June and September [33] and [34]. Mean annual precipitation for the period 2000–2020, derived from interpolation of data from seven rain gauge stations (Bidung, Chenary, Khaling, Radhi, Sakteng, Thrimshing, and Wamrong), ranges from approximately 1,211 mm in rain-shadow areas to about 1,568 mm on windward slopes, reflecting strong orographic control on rainfall distribution. This pronounced seasonality governs groundwater recharge processes, with intense monsoonal rainfall providing the primary input to subsurface storage. However, steep slopes promote rapid surface runoff, limiting infiltration efficiency. As a result, recharge occurs predominantly during the monsoon period, while spring discharge during the dry season is sustained by delayed release of groundwater stored within shallow aquifer systems.

Hydrogeologically, the Gamri Chhu Basin is characterized by shallow, unconfined to semi-confined fractured aquifer systems typical of mountainous terrains in Bhutan. Groundwater occurrence and flow are primarily controlled by secondary porosity, with movement occurring through fracture networks, joints, and lithological contacts. Springs in the basin are predominantly structurally controlled fracture and contact springs, emerging where groundwater flow is concentrated along permeable zones, fracture intersections, or

lithological boundaries [35]. These systems typically have small, localized recharge areas and limited storage capacity, leading to rapid recharge during monsoon periods but relatively quick depletion during dry seasons.

Field-based investigations within the basin, particularly in the Yude Ri and Dungju Ri catchments under Gamri Chhu basin, indicate that spring discharge is mainly sustained by direct rainfall infiltration, shallow groundwater storage, and subsurface flow pathways, with precipitation as the dominant recharge source [36]. However, the combination of steep terrain, limited aquifer storage, and structurally controlled flow pathways results in highly heterogeneous and spatially discontinuous groundwater systems. Consequently, spring discharge is highly sensitive to both climatic variability and local geological conditions. Despite the presence of perennial rivers, rural communities in Gamri Chhu Basin depend heavily on groundwater springs for their daily water needs. Settlements are typically on mid-hillslopes where direct access to the main river is difficult, so springs and small streams are the primary accessible sources of drinking and irrigation water. According to the Watershed Management Division [9] inventory, there are 145 spring sources in this basin. About 70% of these springs (102) are still functioning, while the remaining 30% (43) are drying up or only flow seasonally (See Figure 1). This highlights the need for proactive springshed management in the region.

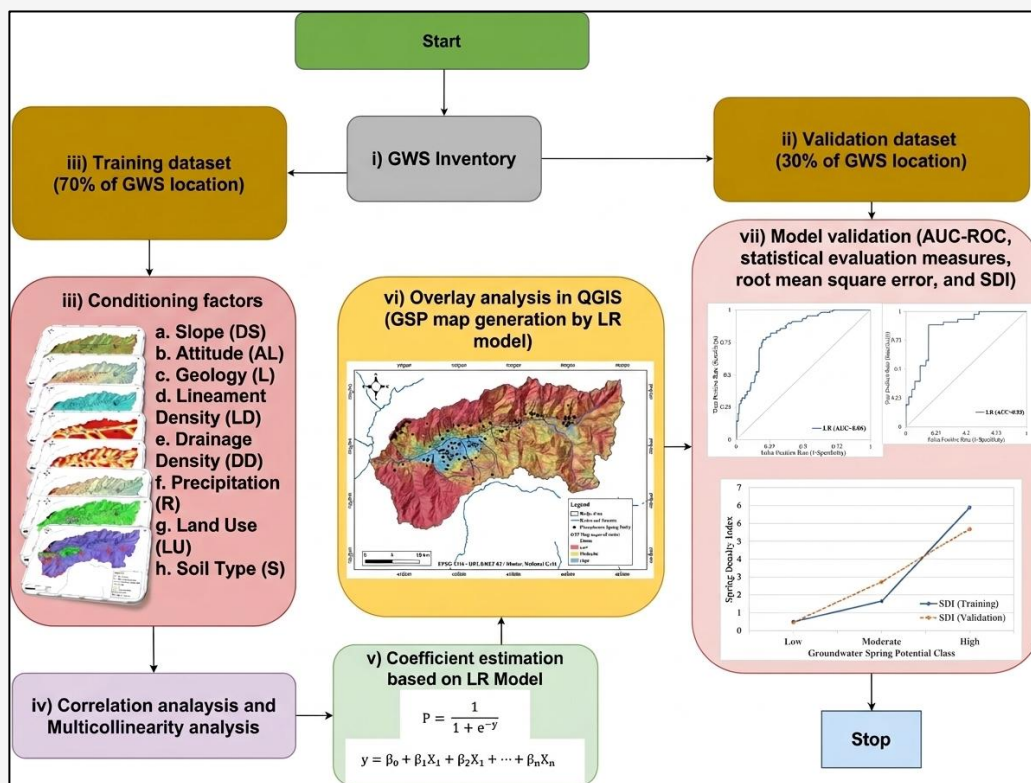


Figure 2: Ground waterspring potential mapping study workflow

3. Methodology

As shown in Figure 2, the methodology comprised: (i) preparation of a groundwater spring inventory representing observed spring locations, as well as the generation of pseudo-absence (non-spring) locations, to construct a binary dataset; (ii) division of spring presence and pseudo-absence locations into training (70%) and validation (30%) datasets for model development and testing; (iii) generation of groundwater spring conditioning factors based on topographic, geological, hydrological, and land-use data; (iv) correlation and multicollinearity analyses conducted as diagnostic checks to assess redundancy among predictors and ensure model stability; (v) estimation of model coefficients using logistic regression in JASP 0.19.3.0 and reverifying in R, incorporating all conditioning factors to quantify their influence on spring occurrence; (vi) GIS-based overlay analysis to generate the final groundwater spring potential map using only statistically significant predictors identified by the logistic regression model; and (vii) model validation using ROC–AUC, accuracy, sensitivity, specificity, root mean square error, and the spring density index to assess predictive performance.

3.1 Groundwater Spring Inventory

A comprehensive groundwater spring (GWS) inventory was obtained from the Department of Water (Ministry of Energy and Natural Resources), Bhutan. This inventory provided georeferenced point locations of 145 known springs within the Gamri Chhu Basin, referenced in the DRUKREF 03/Bhutan National Grid coordinate system (EPSG:5266), along with field observations of their current status (no change, drying, or dried up). For modeling purposes, all spring locations were treated as presence points (binary response = 1), based on the assumption that any mapped spring reflects suitable hydrogeological conditions for groundwater emergence. Spring status information was used solely for descriptive analysis and was not incorporated into the logistic regression model. It is acknowledged, however, that the inventory may not capture all existing springs in remote upland terrain; nevertheless, it represents the most comprehensive georeferenced spring dataset currently available for the basin.

A groundwater spring potential prediction model requires both spring presence and non-spring locations for binary classification [37]. Since verified absence data for groundwater springs are not available, pseudo-absence points were generated to enable binary classification. An equal number of

pseudo-absence points ($n = 145$) was selected to maintain a balanced presence–absence ratio (1:1), thereby reducing potential bias associated with class imbalance in binary classification. Previous studies in spatial modeling have shown that unbalanced pseudo-absence sampling can influence model outcomes; therefore, balanced sampling strategies are commonly adopted to improve consistency and reduce bias [38].

To minimize the likelihood of incorrectly labeling areas influenced by groundwater discharge as absence locations, a 500 m spatial exclusion buffer was applied around all mapped spring points. This buffer distance was selected to be substantially larger than the spatial resolution of the predictor variables ($30\text{ m} \times 30\text{ m}$) and to better represent the spatial extent of local recharge and discharge zones in mountainous fractured aquifer systems. Previous studies have emphasized that groundwater flow systems extend beyond immediate spring outlets and that sufficiently large exclusion zones are necessary to avoid spatial autocorrelation and misclassification of presence-influenced areas [37] and [39].

Pseudo-absence points were generated using random spatial sampling across the basin to capture a broad range of environmental variability represented in the predictor variables. Although alternative approaches such as stratified sampling can also be applied, random sampling provides an objective and reproducible method when reliable prior information on true absence locations is unavailable. This approach is widely used in environmental modeling to represent background conditions in data-driven models [40] and [41]. The combined dataset (290 points: 145 presences and 145 pseudo-absences) was randomly divided into training and validation subsets. Following common practice in environmental modeling, 70% of the data (203 points) was used for model training, while the remaining 30% (87 points) was reserved for independent validation. This 70:30 split is widely applied in machine learning as it provides sufficient data for model development while retaining an adequate portion for reliable performance evaluation [42][43] and [44].

3.2 Groundwater Spring Conditioning Factors

In this study, eight conditioning factors were selected as predictor variables based on a review of previous groundwater spring potential studies and data availability. These factors are: slope, altitude (elevation), geology, drainage density, lineament density, land use/land cover, soil type, and precipitation. They represent the key terrain, lithologic, hydrologic, and land-surface

characteristics that influence groundwater occurrence and spring formation, and each has been used in more than half of prior spring potential mapping studies [45] and [46]. All spatial data processing and analyses were performed in a GIS environment. All data layers were georeferenced to the same coordinate system (DRUKREF 03/Bhutan National Grid) and standardized to a uniform $30\text{ m} \times 30\text{ m}$ grid resolution to ensure spatial alignment. For categorical layers, including geology, land use/land cover, and soil type, nearest-neighbor resampling was applied to preserve original class values. For continuous layers, including altitude, slope, drainage density, lineament density, and precipitation, bilinear interpolation was used to preserve continuous surface characteristics during resampling. The details regarding the conditioning factors relevant to the study are presented below:

3.2.1 Slope (DS)

Slope is often described as the gradient or steepness of the land surface and is commonly expressed as the angle of inclination relative to the horizontal plane or as the percentage rise over run [47]. It is an important topographic factor that controls the percolation of water into the soil [48]. Gentler slopes favor reduced runoff and longer water residence time, which supports greater infiltration and subsurface percolation, whereas steep slopes promote faster overland flow and greater soil erosion, which reduces the time available for rainwater to infiltrate [49]. In this study, the slope layer was derived from an Advanced Land Observing Satellite–Phased Array type L-band Synthetic Aperture Radar (ALOS-PALSAR) Digital Elevation Model provided by the Japan Aerospace Exploration Agency and accessed via the Alaska Satellite Facility (ASF) data portal (<https://search.asf.alaska.edu>). The original DEM, with an approximate spatial resolution of 12.5 m, was resampled to a 30 m grid to ensure consistency with other conditioning factors. See Figure 3(a). Slope was calculated in QGIS 3.38.0 as the first derivative of elevation and incorporated into the logistic regression model as a continuous predictor representing topographic control on groundwater spring occurrence.

3.2.2 Altitude (AL)

Altitude was extracted from the same Advanced Land Observing Satellite–Phased Array type L-band Synthetic Aperture Radar (ALOS-PALSAR) Digital Elevation Model used for slope derivation and treated as a continuous variable representing land surface elevation relative to mean sea level. See Figure 3(b).

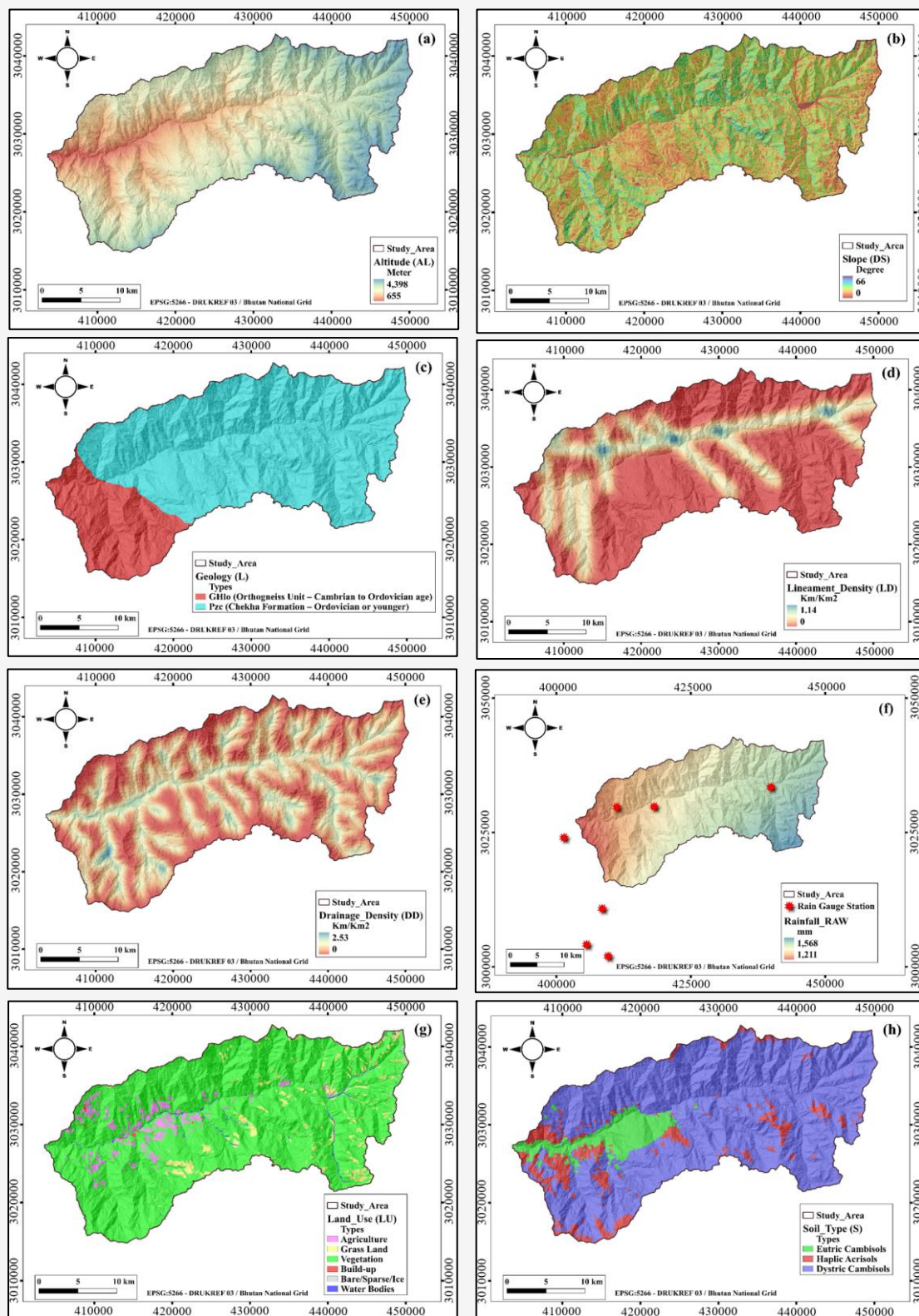


Figure 3: Groundwater conditioning factors (a) altitude, (b) slope, (c) geology, (d) lineaments density, (e) drainage density, (f) rainfall, (g) land use, and (h) soil type

It acts as a first-order control on hydro-climatic conditions and geomorphic processes influencing groundwater recharge and spring occurrence. Altitude plays a critical role in controlling groundwater occurrence, with lower elevations often exhibiting higher groundwater potential due to stronger hydraulic gradients and shallower water tables [50]. Similarly, studies have shown that groundwater potential tends to be concentrated in lower-elevation areas, with comparatively limited occurrence at higher elevations [51].

3.2.3 Geology (L)

Geology is one of the most important and widely used factors controlling groundwater occurrence [45]. Groundwater potential mapping evaluates the water-retention and transmissive characteristics of geological units, including weathered materials, source rocks, and both consolidated and unconsolidated sediments [52]. In this study, the Geological Map of Bhutan developed by Long et al. [53], at a scale of 1:500,000 and officially adopted by the Department of Geology and Mines under the Ministry of Energy and Natural Resources, Bhutan, was used and accessed through the NSDI portal (<https://nsdi.systems.gov.bt/>). According to this map, the study area is dominated by two lithological units: GHlo (Orthogneiss Unit; Cambrian–Ordovician) and Pzc (Chekha Formation; Ordovician or younger), following the original geological classification scheme. The geological layer was converted to raster format and resampled to a 30 m × 30 m spatial resolution using nearest-neighbor resampling to ensure consistency with the other conditioning factors. See Figure 3(c). It is acknowledged that the use of a 1:500,000 scale geological map introduces limitations for basin-scale analysis, as such datasets may not capture fine-scale lithological variability, small structural features, or localized fracture networks that influence groundwater occurrence. However, the objective of this study is regional-scale spring potential mapping rather than site-specific hydrogeological characterization. At this scale of analysis, generalized lithological units are considered sufficient to represent the broad geological controls on groundwater occurrence. Similar studies conducted in data-scarce mountainous regions have also used national-scale geological datasets to map regional groundwater potential [16] and [54].

3.2.4 Lineament Density (LD)

Lineament density (LD) is the total length of lineaments per unit area [55]. Mathematically, it can be expressed as in Equation 1:

$$LD = \frac{\sum L_i}{A}$$

Equation 1

Where L_i is the length of the i -th lineament, and A is the area. Areas with high lineament density are generally used as indicators of zones with higher permeability and groundwater potential in remote sensing/GIS groundwater studies [56].

In this study, lineaments were extracted from Bhutan's national geological map [53] and from the geological map compiled by Gansser [57], as reproduced in RECS International Inc. [58], at an approximate scale of 1:500,000. Since these lineaments were derived from published geological maps rather than from remote sensing imagery, the dataset primarily represents major regional structural features visible at the scale of publication. Smaller fractures, joints, and minor discontinuities, which may be important for localized groundwater flow and spring emergence in mountainous terrain, are therefore not fully captured. Nevertheless, the lineament dataset was retained to represent broad structural trends at the regional scale and to evaluate its relative contribution compared with other conditioning factors. Lineament density was calculated in QGIS 3.38 using the *Line Density* tool available in the Processing toolbox, and the resulting lineament density layer was used as a continuous predictor in the analysis. See Figure 3(d).

3.2.5 Drainage Density (DD)

Drainage density refers to the total length of streams and river channels within a defined area, divided by the surface area of that area [59]. Mathematically, it can be expressed in Equation 2:

$$DD = \frac{\sum L_d}{A}$$

Equation 2

Where L_d is the length of the channel d and A is the area considered [60]. DD represents the drainage efficiency of a basin, expressed through the cumulative length of surface water flow paths [61]. Low drainage density usually reflects permeable soils or rocks and gentle terrain that favor precipitation infiltration and groundwater recharge, whereas high drainage density is associated with steeper or less permeable surfaces where runoff dominates, and recharge is limited [62]. The drainage network was derived from the DEM using SAGA 9.1 tools in QGIS 3.38.0 (Terrain Analysis → Channels → Channel network and drainage basins), after which drainage density (km/km²) was calculated

using SAGA's Line Density tool and treated as a continuous variable for the analysis. See Figure 3(e).

3.2.6 Precipitation (*R*)

Precipitation is a significant factor affecting groundwater recharge [63] and is widely regarded as the principal source of groundwater replenishment [64]. Precipitation contributes to groundwater recharge through infiltration and percolation processes, whereby water moves through the subsurface to replenish aquifers [65]. Stable isotope studies have further shown that precipitation infiltrates at higher elevations and moves downslope through subsurface pathways before reappearing as springs, highlighting the close relationship between seasonal precipitation and spring discharge patterns [66].

In this study, annual precipitation data for the period 2000–2020 were obtained from the National Centre for Hydrology and Meteorology (NCHM), Bhutan, based on seven rain gauge stations (Bidung, Chenary, Khaling, Radhi, Sakteng, Thrimshing, and Wamrong) located within Trashigang Dzongkhag. These stations are distributed across the basin and span an elevation range of approximately 746 m to 2955 m, thereby partially representing orographic precipitation gradients. A continuous precipitation surface was generated using the inverse distance weighted (IDW) interpolation method in QGIS, with a distance coefficient (power) of 2.0. The selected QGIS interpolation tool did not include a separate user-defined search radius parameter. However, the spatial distribution of rain gauge stations remains uneven, with limited coverage in high-altitude headwater regions and steep mountainous areas, which may reduce the local accuracy of the interpolated precipitation surface. No formal cross-validation of the IDW surface was performed; therefore, the precipitation layer should be interpreted as a generalized representation of basin-scale rainfall distribution rather than a precise depiction of local rainfall variability. Despite this limitation, the dataset provides the best available long-term precipitation information for the study area and is considered adequate for regional-scale analysis of groundwater spring potential. The resulting average annual precipitation values (mm) were used as a continuous predictor in the groundwater spring potential model. See Figure 3(f).

3.2.7 Land use (*LU*)

Land use and land-cover change strongly influence watershed hydrology by altering infiltration, evapotranspiration, runoff, and groundwater recharge processes [67]. Urbanization generally increases impervious surface cover, which reduces

soil infiltration capacity, accelerates surface runoff, and consequently limits groundwater recharge [68]. Vegetated land covers such as forests and grasslands generally enhance infiltration and subsurface flow by improving soil structure and hydraulic properties, although actual groundwater recharge depends on interacting factors including vegetation characteristics, evapotranspiration demand, soil properties, and climate [69]. Agricultural land can either promote or restrict recharge depending on management practices; irrigation may increase percolation [70], whereas soil compaction and intensive tillage can reduce infiltration [71].

For this research, a 2020 land-use map was obtained from the Bhutan NSDI portal (<https://nsdi.systems.gov.bt/>). The temporal proximity of this dataset to the 2021 spring inventory supports its use for regional-scale spatial analysis, although some local land-cover changes may not be fully captured. The dataset categorizes the landscape into six classes: agriculture (4.17% area coverage; 28 spring points), grassland (4.8%; 8 spring points), forest or vegetation (90%; 109 spring points), built-up areas (0.21% area coverage; no spring points), bare or sparsely vegetated/ice surfaces (0.23% area coverage; no spring points), and water bodies (0.59% area coverage; no spring points). See Figure 3(g).

3.2.8 Soil type (*S*)

Soil characteristics play a key role in controlling groundwater infiltration processes [72] and are therefore fundamental to understanding and assessing groundwater recharge [73]. Coarse, well-drained soils (sand, gravel) typically have higher saturated hydraulic conductivity and percolation, fine soils (silt, clay) restrict infiltration, and loams are intermediate [46]. For this study, we used Bhutan's national soil map produced by the National Soil Services Centre (NSSC) via digital soil mapping [74] and clipped it to the basin. Within the mapped area, the soil units include Eutric Cambisols (8.15% area coverage; 59 spring points), Haplic Acrisols (9.67% area coverage; 6 spring points), and Dystric Cambisols (82.18% area coverage; 80 spring points). See Figure 3(h). Hydrologically, Cambisols are generally moderately permeable soils that allow infiltration and shallow subsurface flow, whereas Acrisols tend to be more weathered and clay-rich, often associated with lower permeability and reduced infiltration capacity [75].

3.3 Correlation and Multicollinearity Analysis

If strong correlations exist among conditioning factors in spatial mapping, they can reduce model accuracy and efficiency; therefore, highly correlated factors should be identified and excluded from the

analysis [76]. The two most common approaches for a correlation check are as follows:

a. Pearson correlation check

Pearson's correlation coefficient indicates both the strength and the direction of a linear association between two variables [77]. Pairwise correlations among the conditioning factors (r_{xy}) were assessed using Equation 3:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Equation 3

Where x_i and y_i are the observed values of variables x and y for the i^{th} individual, and \bar{x} , \bar{y} are their respective means. In statistical analyses, a Pearson correlation coefficient greater than 0.7 is generally considered to reflect substantial collinearity among variables [78] and [79].

b. Multicollinearity check

Multicollinearity arises when predictors in a dataset are highly correlated, reducing their independence and potentially biasing analytical outcomes [80]. In this study, multicollinearity was assessed using the Variance Inflation Factor (VIF) and Tolerance (TOL). A VIF value greater than 10 or a TOL value below 0.1 is considered indicative of significant multicollinearity [80] and [81]. The two measures can be expressed in Equations 4 and 5:

$$VIF_i = \frac{1}{TOL_i}$$

Equation 4

$$TOL_i = 1 - R_i^2$$

Equation 5

Where R_i^2 is the coefficient of determination obtained by regressing the i^{th} independent variable against all other independent variables. Highly collinear conditioning factors are dropped for the analysis.

3.4 Logistic Regression (LR)

Logistic regression is a fundamental technique widely used for classification problems. It examines the relationship between a set of input features and a categorical dependent variable, typically with two possible outcomes. The model estimates the probability of class membership by fitting the data to a logistic function. This function forms a sigmoid curve that transforms predicted values into

probabilities between 0 and 1. In this study, statistical analyses were initially performed in JASP (version 0.19.3.0) and subsequently verified in R to confirm the consistency of parameter estimates, significance levels, and overall model outputs. Since both JASP and R implement logistic regression using standard maximum-likelihood estimation, the analytical procedure is methodologically equivalent to widely used statistical workflows in environmental modeling. In practical applications, a decision threshold is applied to these probabilities, with values below the threshold classified as 0 and values above it as 1 [82]. In GWP studies, LR estimates the probability of groundwater presence or absence in the region, using input features (conditioning factors) and a categorical dependent variable (groundwater inventory presence and absence points) [45]. Mathematically, it can be expressed in Equation 6:

$$P = \frac{1}{1 + e^{-y}}$$

Equation 6

Here, P represents the probability of groundwater spring occurrence, ranging from 0 to 1, while y denotes the linear logistic function, defined in Equation (7), which extends from $-\infty$ to $+\infty$.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Equation 7

Where X_1 , X_2 , ..., and X_n are the explanatory variables, β_0 is the model intercept and β_1 , β_2 , ..., and β_n are the regression coefficients of the LR model.

3.5 Model Validation Method

Model validation is a critical step in the development of predictive models, and it is commonly carried out using a range of statistical evaluation measures [26]. In this study, model performance was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) [83], along with additional indicators including Accuracy, Sensitivity, Specificity, and Root Mean Square Error (RMSE) [84]. Detailed descriptions and theoretical backgrounds of these evaluation metrics are available in previous research [85][86] and [87]. The mathematical formulations of the applied metrics are presented below. Mathematically, these metrics are expressed in Equations 8 to 11:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 8

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

Equation 9

$$\text{Specification} = \frac{TN}{TN + FP}$$

Equation 10

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_p - X_a)^2}$$

Equation 11

Where *TP* (True Positive) refers to cases where actual groundwater springs are correctly predicted as springs, while *FP* (False Positive) indicates non-spring areas incorrectly predicted as springs. *FN* (False Negative) represents actual springs that are predicted as non-spring, and *TN* (True Negative) refers to non-spring areas correctly identified as non-spring. Here, X_p denotes the predicted value, and X_a represents the actual value, and n is the total number of observations used in the analysis.

3.6 Spring Density Index

Besides the validation techniques described earlier, density-index methods are also widely applied in engineering and geotechnical research to assess model performance. One commonly used approach is the Spring Density Index (SDI). The SDI measures how effectively a potential map groups actual spring locations within high-potential zones. It is calculated as the ratio of the percentage of validation springs located in a specific potential class to the percentage of the map area occupied by that class [88]. In simple terms, a higher SDI value indicates that the model is better at concentrating springs in the most favorable areas. The mathematical expression of the index is given in Equation 12:

$$\text{SDI}_i = \frac{P_{s,i}}{P_{a,i}}$$

Equation 12

Where SDI_i is the Spring Density Index for class i , $P_{s,i}$ is the percentage of validation springs in class i , and $P_{a,i}$ is the percentage of area occupied by class i .

When *SDI* values gradually increase from low to very high susceptibility classes, it indicates that the map performs consistently and can be regarded as reliable and well validated [89]. For the current study, we will also evaluate the map produced by LR model using

the *SDI* based on both the training and validation datasets of groundwater spring inventory points.

4. Results

4.1 Correlation and Multicollinearity Check

Pearson's correlation analysis (Figure 4) and multicollinearity diagnostics (Table 1) indicated generally low to moderate interrelationships among the conditioning factors ($r < 0.7$). However, a moderate correlation between altitude and precipitation was observed ($r = 0.698$), reflecting orographic controls on rainfall distribution in mountainous terrain.

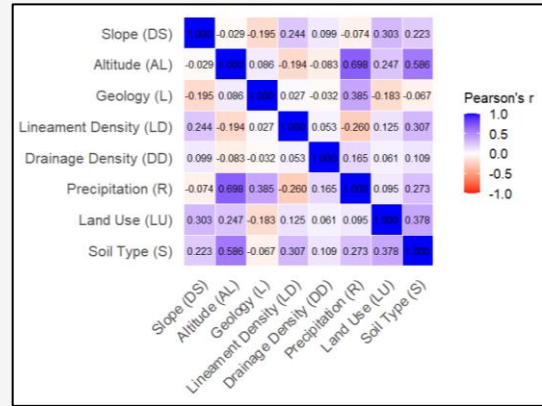


Figure 4: Pearson correlation matrices of conditioning factors used in the study

This indicates a degree of shared information between the two variables, as precipitation patterns are partly elevation-dependent. Nevertheless, variance inflation factor (VIF) values remained below the threshold of 10, confirming the absence of severe multicollinearity. Although some redundancy may exist, altitude and precipitation represent distinct but complementary hydrological processes. Altitude governs topographic controls on groundwater flow and discharge zones, while precipitation represents the primary input for recharge. Their combined inclusion allows the model to capture both elevation-driven climatic variability and direct recharge inputs, without introducing problematic multicollinearity.

4.2 Logistic Regression (LR) Model Results

The logistic regression (LR) model was developed using the original continuous dataset without prior transformation, thereby reducing the potential loss of information associated with discretizing predictor variables. All statistical analyses were performed using JASP (version 0.19.3.0) and reverified in R (see Table 2).

Table 1: Multicollinearity diagnostics (Tolerance and VIF) of conditioning factors

Conditioning factors	Tolerance	VIF
Slope (DS)	0.742	1.348
Altitude (AL)	0.185	5.391
Geology (L)	0.848	1.179
Lineament Density (LD)	0.608	1.643
Drainage Density (DD)	0.598	1.672
Precipitation (R)	0.248	4.037
Land Use (LU)	0.892	1.121
Soil Type(S)	0.448	2.232

Table 2: Logistic regression coefficients (Estimate, Standard Error, z, and Wald test statistics) for groundwater spring occurrence using conditioning factors

Model	Variable	Estimate (β)	Standard Error	z	Wald Test			95% Confidence interval	
					Wald Statistic	df	p	Lower bound	Upper bound
M₀	(Intercept)	0.01	0.14	0.07	0.005	1	0.944	-0.265	0.285
M₁	(Intercept)	-12.916	6.195	-2.085	4.347	1	0.037	-25.057	-0.774
	Slope (DS)	-0.092	0.025	-3.667	13.446	1	<0.001**	-0.141	-0.043
	Altitude (AL)	-0.003	0.001	-5.323	28.33	1	<0.001**	-0.005	-0.002
	Geology (L)	2.604	0.796	3.273	10.712	1	0.001**	1.045	4.163
	Lineament Density (LD)	-0.514	0.887	-0.58	0.336	1	0.562	-2.253	1.225
	Drainage Density (DD)	-1.181	0.486	-2.43	5.903	1	0.015*	-2.134	-0.228
	Precipitation (R)	0.013	0.005	2.439	5.949	1	0.015*	0.003	0.023
	Land Use (LU)	-0.246	0.334	-0.736	0.542	1	0.462	-0.901	0.409
	Soil Type (S)	0.771	0.347	2.219	4.922	1	0.027*	0.09	1.451

Note: M₀ denotes the null model (intercept-only), while M₁ denotes the full logistic regression model incorporating all selected conditioning variables, *Significant ($p < 0.05$), and **Highly Significant ($p < 0.01$)

Unlike stepwise selection approaches, which have been criticized for increasing the risk of overfitting, inflating Type I error rates, and producing unstable coefficient estimates, this study employed a full model approach in which all conditioning factors were included simultaneously. This approach allows for a comprehensive evaluation of the relative influence of each variable while avoiding biases associated with automated variable selection procedures [90]. The significance of model coefficients was evaluated using Wald statistics, p-values, and 95% confidence intervals. Predictors with p-values less than 0.05 and confidence intervals that did not include zero were considered statistically significant [91] and [92]. The results indicate that altitude ($p < 0.001$), slope ($p < 0.001$), geology ($p = 0.001$), drainage density ($p = 0.015$), precipitation ($p = 0.015$), and soil type ($p = 0.027$) show statistically significant associations with groundwater spring occurrence. In contrast, lineament density ($p = 0.562$) and land use ($p = 0.462$) do not show statistically

significant effects within the model. The 95% confidence intervals further support these findings. Significant predictors such as slope (-0.141 to -0.043), altitude (-0.005 to -0.002), geology (1.045 to 4.163), drainage density (-2.134 to -0.228), precipitation (0.003 to 0.023), and soil type (0.090 to 1.451) do not cross zero, indicating stable effects. In contrast, lineament density (-2.253 to 1.225) and land use (-0.901 to 0.409) show higher uncertainty. The regression coefficients indicate that geology ($\beta = 2.604$), precipitation ($\beta = 0.013$), and soil type ($\beta = 0.771$) have positive associations with groundwater spring occurrence, whereas slope ($\beta = -0.092$), altitude ($\beta = -0.003$), and drainage density ($\beta = -1.181$) have negative associations within the fitted logistic regression model.

The coefficients of the statistically significant conditioning factors were applied to generate the groundwater spring potential map (GSPM) using the logistic sigmoid function defined in Equation (6). The linear predictor y, formulated as a weighted

combination of the retained explanatory variables following Equation (7), was estimated from the fitted logistic regression model and is expressed in Equation 13:

$$Y = -12.916 - 0.092X_{DS} - 0.003X_{AL} + 2.604X_L - 1.181X_{DD} + 0.013X_R + 0.771X_S$$

Equation 13

Finally, the obtained GSPM index values were classified into three categories (low, moderate, and high) using the Natural Breaks (Jenks) method. This approach was selected because it identifies class boundaries based on inherent groupings within the data by minimizing within-class variance and maximizing between-class differences, making it suitable for representing non-uniform probability distributions [93]. The Natural Breaks method has been widely used in groundwater and environmental susceptibility mapping studies to represent spatial variability and preserve meaningful patterns in model outputs [94] and [95]. In contrast to quantile classification, which forces equal proportions of data into each class regardless of the underlying distribution, the Natural Breaks method allows the classification to reflect genuine statistical variations in groundwater spring potential [96]. Based on the final Jenks classification, low-potential zones corresponded to probability values of ≤ 0.16 , moderate-potential zones to 0.16–0.54, and high-potential zones to ≥ 0.54 . The resulting classification produced low (81.34%), moderate (11.80%), and high (6.87%) potential zones (Figure 5).

4.3 Model Validation

The performance of the LR model is presented in Table 3 and Figure 6. The SRC (Success Rate Curve) reflects model fit to the training dataset, whereas the PRC (Prediction Rate Curve) represents predictive performance on the validation dataset [97] and [98]. The areas under the ROC curve (AUC) for the training dataset (SRC) and validation dataset (PRC) were 0.86 and 0.85, respectively. These values indicate “very good” discriminatory performance, showing that the model can reliably distinguish spring locations from non-spring locations in pairwise comparisons for both the training and validation datasets [85] [99] and [100]. The predictive performance of the LR model was further evaluated using accuracy, sensitivity, specificity, and RMSE for both training and validation datasets (Table 3). In the training dataset, the model achieved an accuracy of 0.80, with sensitivity and specificity values of 0.75 and 0.85, respectively, indicating slightly better identification of non-spring locations than spring locations. The RMSE value of 0.50 reflects a moderate level of prediction error. In the validation dataset, the model achieved an accuracy of 0.85, with sensitivity increasing to 0.88, indicating improved identification of spring locations. However, specificity decreased slightly to 0.81, suggesting a minor reduction in correctly classifying non-spring locations. The RMSE remained stable at 0.51, indicating consistent prediction error between training and validation datasets. Overall, these results demonstrate that the LR model exhibits stable generalization performance, with a tendency toward higher sensitivity in the validation phase.

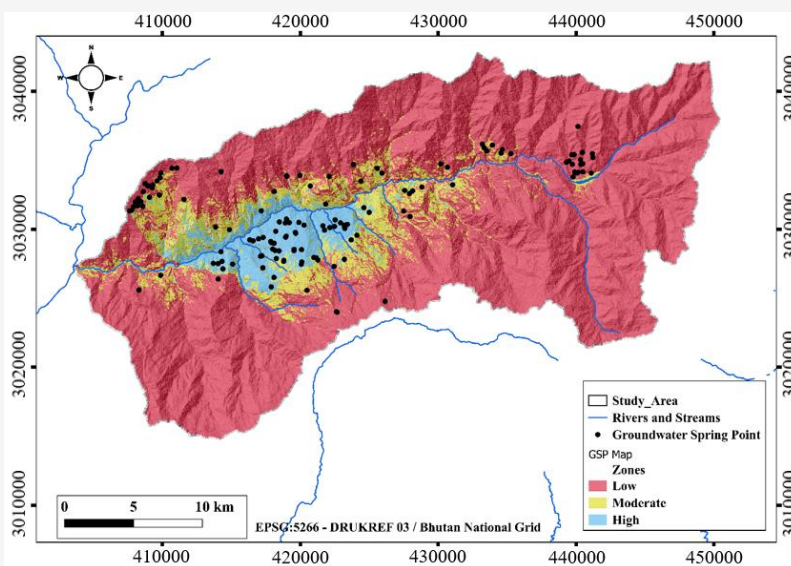


Figure 5: Groundwater spring potential maps based on LR model

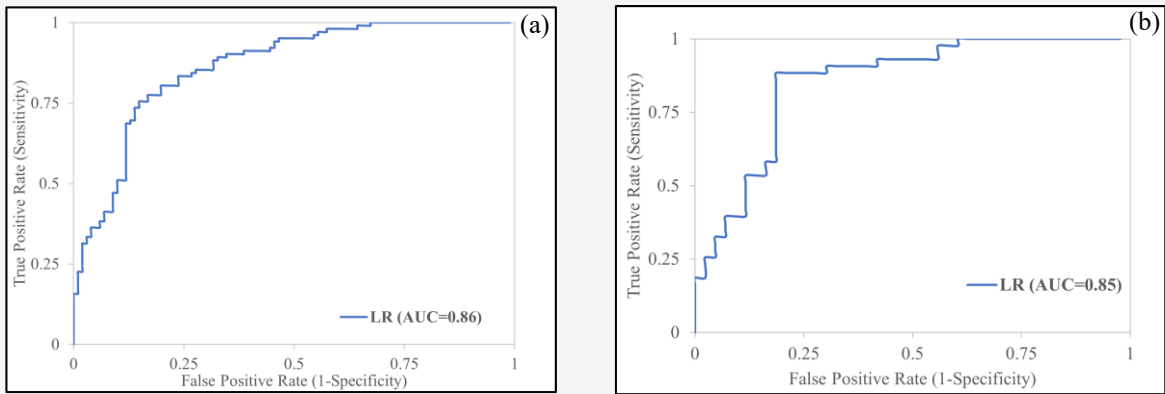


Figure 6: Success Rate Curve (SRC) (a) and Prediction Rate Curve (PRC) (b) for LR model

Table 3: Performance of the LR model

Model	AUC	Accuracy	Sensitivity	Specificity	RMSE
LR (training)	0.86 (SRC)	0.8	0.75	0.85	0.50
LR (validation)	0.85 (PRC)	0.85	0.88	0.81	0.51

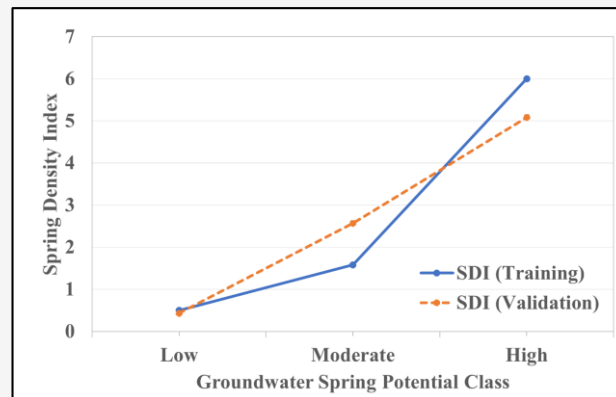


Figure 7: Spring Density Index (SDI) based on training and validation spring points for LR model

Table 4: Spring Density Index (SDI) and spring distribution across different potential classes

Model	Class	Area (km ²)	Area (%)	Springs (Train)	Springs % (Train)	Spring (Validation)	Spring % (Validation)	SDI (Training)	SDI (Validation)
LR	Low	597	81.34	41	40.20	15	34.89	0.50	0.43
	Moderate	86.6	11.80	19	18.63	13	30.24	1.58	2.57
	High	50.4	6.87	42	41.18	15	34.89	6.00	5.08
	Total	734	100.0	102		43	100.0		

4.4 Spring Density Index (SDI) Results

The Spring Density Index (SDI) results for the LR model are presented in Table 4 and Figure 7. In both the training and validation datasets, SDI values increased progressively from low- to high-potential classes. This trend satisfies the reliability criterion suggested by Mersha and Meten [89], which states

that a valid susceptibility map should display a monotonic increase in density index toward higher potential classes. The LR model achieved the highest SDI values in the high-potential class, with 6.00 for the training dataset and 5.08 for the validation dataset. Moreover, the progressive increase in SDI from low to high classes indicates good spatial

agreement between the classified groundwater spring potential zones and the observed spring distribution. This pattern is consistent with findings reported in previous groundwater and landslide susceptibility studies [88] and [101]. Overall, the SDI results support the spatial reliability of the LR-based spring potential map.

5. Discussion

The present study developed a GIS-based LR model to assess groundwater spring potential in the Gamri Chhu Basin of eastern Bhutan. The model demonstrated strong predictive capability, achieving AUC values of 0.86 for the training dataset and 0.85 for the validation dataset. According to established classification criteria, AUC values above 0.80 indicate very good model performance [99] and [100]. While directly comparing AUC values across regions has limitations due to differences in hydrogeological conditions, data quality, and sampling methods, the model's performance aligns with those reported in other mountainous areas [28][29][16][27] and [102]. These findings suggest that logistic regression is a practical and interpretable approach for preliminary mapping of groundwater spring potential in complex Himalayan terrain, although its effectiveness depends on the study area's specific features.

Among the eight conditioning factors analyzed, altitude, drainage density, slope, precipitation, soil type, and geology were found to be statistically significant predictors of spring occurrence, while lineament density and land use showed no significant influence. A negative elevation–spring relationship suggests that spring occurrence is more likely at lower topographic positions, where groundwater can converge and discharge after being recharged upslope through lateral subsurface flow pathways [103]. Similar patterns have been observed in mountain catchments, where groundwater contributions to spring flow increase toward lower elevations due to downslope accumulation and discharge behavior [36].

Slope exhibited a significant negative association with spring occurrence, indicating that gentler terrain favors infiltration and percolation, while steeper slopes promote rapid overland flow and limit recharge. This finding is consistent with established runoff–infiltration theory and experimental evidence showing that infiltration decreases and runoff increases with increasing slope gradients [104]. Drainage density showed a strong negative association with spring occurrence. Areas with low drainage density typically indicate more permeable terrain and greater infiltration capacity, whereas high drainage density is commonly associated with rapid

runoff and reduced groundwater recharge. Similar interpretations have been reported in Himalayan basins in studies mapping groundwater potential and recharge zones [55] and [105]. Precipitation was positively associated with spring occurrence, reflecting its fundamental role in groundwater recharge and the maintenance of subsurface flow paths that sustain springs [106]. Isotope-based studies from Himalayan catchments further demonstrate that spring discharge often reflects precipitation signatures and can be traced to recharge occurring at higher elevations, supporting the concept of upslope recharge feeding downslope spring emergence through subsurface pathways [107].

Geology emerged as a significant predictor, highlighting the strong lithological control on groundwater storage and movement. The Chekha Formation (Pzc) exhibited much higher probabilities of spring occurrence than the Orthogneiss unit (GHlo), reflecting important structural and lithological contrasts. The Orthogneiss is a mechanically strong, erosion-resistant rock that weathers into massive blocks with low primary porosity and limited infiltration capacity [53]. In contrast, the Chekha Formation comprises layered quartzite and schist, where interbedded schist layers create planes of weakness and enhanced secondary permeability, favoring groundwater storage and spring discharge.

Regarding soil characteristics, the distribution of springs reflects the hydro-physical properties of the mapped units. Dystric Cambisols, which dominate the study area, are moderately developed and allow infiltration but tend to promote lateral subsurface flow under humid conditions, supporting spring emergence along slopes. Eutric Cambisols, with comparatively balanced drainage and moisture-retention properties, are also favorable for localized groundwater recharge and temporary storage, contributing to spring occurrence [75]. Haplic Acrisols are more weathered and clay-enriched in subsurface horizons, which can restrict vertical infiltration and enhance runoff, thereby limiting groundwater recharge and reducing groundwater occurrence [108]. This behavior aligns with the statistical results, where soil type was identified as a significant predictor of spring distribution, with a high concentration of springs in Dystric Cambisols, followed by Eutric Cambisols, and fewer in Haplic Acrisols.

The non-statistically significant results for land use and land cover ($p > 0.05$) are likely due to the basin being mostly covered by forested and vegetated landscapes (about 90% of the area), which reduces class contrast in the model. This explanation aligns

with studies showing that unbalanced data distributions can decrease sensitivity to underrepresented categories, bias models toward the majority class, and limit the model's ability to distinguish within-variable effects. [109] and [110]. In addition, some spatial confounding may exist between land use and altitude because forest and vegetation cover in the basin occur predominantly in mid-elevation zones, which overlap the elevation range where spring occurrence is relatively common. This overlap may partly reduce the independent explanatory effect of land use within the fitted model. Nevertheless, the highest number of groundwater spring points was recorded in forest and vegetated areas, followed by agricultural land and grassland, whereas no springs occurred in built-up areas, bare or sparsely vegetated zones, ice-covered areas, or water bodies. Although this pattern is consistent with hydrological understanding that vegetation enhances infiltration and groundwater recharge [69], similar trends have also been reported in the Himalayan region, where spring distribution and persistence are closely associated with vegetation cover and recharge potential, with forested areas supporting higher spring densities [111] and [112].

The non-significant contribution of lineament density (LD) in the logistic regression model is likely due to the coarse resolution of the available geological datasets (1:500,000), which capture only major tectonic features and fail to represent the fine-scale fracture networks that primarily control groundwater flow in mountainous terrains. Despite these few non-significant factors, the combination of significant predictors in our model effectively differentiates spring-potential areas. The Spring Density Index (SDI) analysis further supported the reliability of the spring potential map: SDI values increased from low to high potential classes for both training and validation datasets, consistent with density-index validation logic commonly used in susceptibility mapping, where an increasing density index toward higher classes indicates a valid zonation [88] and [89].

Compared to previous groundwater mapping studies in Bhutan, which relied primarily on multi-criteria decision-making approaches such as AHP [30] [31] and [32], the present study offers a statistically objective alternative. Unlike expert-based weighting methods, logistic regression derives factor influence directly from observed data, reducing subjectivity and improving reproducibility. This represents a useful methodological contribution to the assessment of groundwater and spring potential in Bhutan.

Despite the model's satisfactory performance, several sources of uncertainty need to be

acknowledged. Spatial uncertainty may arise from the input data's resolution (30 m), which can affect the representation of terrain and environmental variables. Additionally, the use of randomly generated pseudo-absence points introduces sampling-related uncertainty, as their locations might not fully reflect true absence conditions. Uncertainty is also linked to the logistic regression model itself, especially in estimating coefficients and their confidence intervals, which can impact the stability and interpretation of predictor effects. These factors together may influence the spatial accuracy of the predicted spring potential zones and should be considered when interpreting the results.

6. Conclusion

This study applied a GIS-based logistic regression model to map groundwater spring potential in the Gamri Chhu Basin, integrating topographic, geological, hydrological, climatic, and land-use factors to identify the main controls on spring occurrence in a mountainous environment. The resulting map functions as a regional-scale screening tool to support the prioritization of sites for field investigation, resource allocation, and further hydrogeological assessment. From a remote sensing and geospatial analysis perspective, the study demonstrates the value of integrating multisource spatial datasets to delineate potential spring occurrence zones in complex terrain where direct field data are limited. Compared with knowledge-driven approaches such as the Analytic Hierarchy Process (AHP), logistic regression provides a more objective and statistically reproducible framework for evaluating factor-response relationships. However, AHP remains a useful alternative in data-scarce settings where expert knowledge can compensate for limited empirical observations. The practical value of the model lies in its ability to support early-stage planning by identifying areas that warrant further investigation. When used alongside established frameworks such as the ICIMOD six-step spring revival protocol, particularly for spring mapping, recharge-zone identification, and monitoring, the outputs can support intervention-site selection and improve resource allocation. In this way, the approach can reduce the uncertainty and cost associated with traditional field-based exploration while helping local authorities and communities connect spatial evidence with actionable spring management strategies.

Despite these strengths, the study remains exploratory. The lack of direct field validation and the dependence on pseudo-absence data introduce uncertainty; therefore, the model outputs should be interpreted with caution. In addition, although the

methodological framework is adaptable, the selection of conditioning factors and associated regression coefficients is basin-specific and requires local recalibration before application in other regions. The classified area proportions are also influenced by the choice of classification method. Although the Natural Breaks (Jenks) method was adopted to better represent the inherent distribution of the model output, alternative classification schemes may yield different class boundaries and area percentages. Therefore, the mapped potential classes should be interpreted as regionally useful screening categories rather than fixed or absolute thresholds.

Future research should prioritize uncertainty quantification and field-based validation to enhance confidence in model outputs. This includes evaluating the sensitivity of results to pseudo-absence sampling strategies, quantifying spatial uncertainty in predicted classes, and validating high-potential zones through targeted hydrogeological investigations. In addition, comparative assessment with alternative modeling approaches such as Weight-of-Evidence–Logistic Regression (WoE–LR), Random Forest–Logistic Regression (RF–LR), and Support Vector Machine (SVM)-based models is recommended to improve predictive performance while maintaining interpretability. Furthermore, incorporating higher-resolution datasets may enhance predictive accuracy by capturing fine-scale terrain and hydrogeological variability, while preserving model interpretability for practical water resource management in mountainous regions. Integrating these models into web-based and cloud-enabled geospatial platforms could also improve data handling, scalability, and real-time decision support.

Acknowledgment

This research was conducted as part of a Master's thesis in the Water Resources Engineering Program at Naresuan University, Thailand. The author gratefully acknowledges the academic support and facilities provided by Naresuan University. The author also acknowledges Tzu Chi University and the Buddhist Tzu Chi Medical Foundation for providing financial support for the article processing charge (APC) under the scheme 發表於英文期刊 : TCMF-SP 115-04. The author thanks colleagues and faculty members who provided constructive feedback during the development of this research.

References

- [1] Giacomo, B., Karishma, K., Alice, A., Bobby, M. and Sanjeev, B., (2024). Spring Revival in the Mid-Hills of the Himalaya: A Socioeconomic Assessment Using Benefit–Cost Analysis. *Mountain Research and Development*, Vol. 44(4); 151-159. <https://doi.org/10.1659/mrd.2024.00020>.
- [2] Rawat, S., Jose, P., Rai, S. and Hakhoo, N., (2018). Spring Sanctuary Development Sustaining Water Security in the Himalayan Region in Changing Climate. *Proceedings of the International Conference on Water Environment and Climate Change Knowledge Sharing and Partnership*, Kathmandu, Nepal; 151-159.
- [3] Shrestha, R., Desai, J., Mukherji, A., Dhakal, M., Kulkarni, H. and Acharya, S., (2017). Application of Eight-Step Methodology for Reviving Springs and Improving Springshed Management in the Mid-Hills of Nepal. *International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal*. <https://doi.org/10.53055/ICIMOD.652>.
- [4] Chapagain, P. S., Ghimire, M. and Shrestha, S., (2019). Status of Natural Springs in the Melamchi Region of the Nepal Himalayas in the Context of Climate Change. *Environment, Development and Sustainability*, Vol. 21(1); 263-280. <https://doi.org/10.1007/s10668-017-0036-4>.
- [5] Tiwari, S., Kumar, G., Arrawatia, M. L., Kumar, H., Mishra, K. and Gupta, A. K., (2012). Reviving Dying Springs: Climate Change Adaptation Experiments from the Sikkim Himalaya. *Mountain Research and Development*, Vol. 32(1); 62-72. <https://doi.org/10.1659/MRD-JOURNAL-D-11-00079.1>.
- [6] Verma, R. and Jamwal, P., (2022). Sustenance of Himalayan Springs in an Emerging Water Crisis. *Environmental Monitoring and Assessment*, Vol. 194(2). <https://doi.org/10.1007/s10661-021-09731-6>.
- [7] NITI Aayog., (2018). Inventory and Revival of Springs in the Himalayas for Water Security. Department of Science and Technology, Government of India, New Delhi, India. [Online]. Available: https://dst.gov.in/sites/default/files/Final_NITI%20Report_Himalayan_Springs_23Aug2018.pdf. [Accessed: Sep. 19, 2025].

- [8] Wester, P., Mishra, A., Mukherji, A. and Shrestha, A., (2019). The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People. *Springer Nature*. <https://doi.org/10.1007/978-3-319-92288-1>.
- [9] Watershed Management Division., (2021). Assessment and Mapping of Water Sources in Bhutan: A Comprehensive Inventory and Status of Water Sources Used by the Bhutanese Communities. Department of Forests and Park Services, Ministry of Agriculture and Forests, Thimphu, Bhutan, *Technical Report*. [Online]. Available: <https://www.dofps.gov.bt/reports/>. [Accessed: Oct.01, 2025].
- [10] Ghimire, M., Chapagain, P. and Shrestha, S., (2019). Mapping of Groundwater Spring Potential Zone Using Geospatial Techniques in the Central Nepal Himalayas: A Case Example of Melamchi–Larke Area. *Journal of Earth System Science*, Vol. 128. <https://doi.org/10.1007/s12040-018-1048-7>.
- [11] Ranjan, P., Pandey, P. and Pandey, V., (2023). Delineation of Spring Potential Zones Using Geospatial Techniques: Arunachal Pradesh, India. *Current World Environment*, Vol. 18(1); 87-101. <https://doi.org/10.12944/CWE.18.1.8>.
- [12] ICIMOD., (2021). *Reviving and Protecting Our Springs*. International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal.
- [13] Israil, M., Alhadithi, M. and Singhal, D., (2006). Application of a Resistivity Survey and Geographical Information System (GIS) Analysis for Hydrogeological Zoning of a Piedmont Area, Himalayan Foothill Region, India. *Hydrogeology Journal*, Vol. 14; 753-759. <https://doi.org/10.1007/s10040-005-0483-0>.
- [14] Jha, M., Chowdary, V. 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; 1713-1728. <https://doi.org/10.1007/s10040-010-0631-z>.
- [15] Zeinivand, H. and Ghorbani, S., (2018). Application of GIS-Based Data-Driven Models for Groundwater Potential Mapping in Kuhdasht Region of Iran. *Geocarto International*, Vol. 33; 651-666. <https://doi.org/10.1080/10106049.2017.1289560>.
- [16] Rahmati, O., Naghibi, S. A., Shahabi, H., Bui, D. T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A. N. and Melesse, A. M., (2018). Groundwater Spring Potential Modelling: Comprising the Capability and Robustness of Three Different Modeling Approaches. *Journal of Hydrology*, Vol. 565; 248-262. <https://doi.org/10.1016/j.jhydrol.2018.08.027>
- [17] Chen, M., Zhang, S., Liu, S., Li, M., Zhang, T., Wu, T. and Bu, X., (2025). Mapping the Groundwater Potential Zones in Mountainous Areas of Southern China Using GIS, AHP and Fuzzy AHP. *Scientific Reports*, Vol. 15(1). <https://doi.org/10.1038/s41598-025-01837-y>.
- [18] Nugroho, J. T., Lestari, A. I., Gustiandi, B., Sofan, P., Prasasti, I., Rahmi, K. I. N., Noviar, H., Sari, N. M., Manalu, R. J., Arifin, S. and Taufiq, A., (2024). Groundwater Potential Mapping Using Machine Learning Approach in West Java, Indonesia. *Groundwater for Sustainable Development*, Vol. 27. <https://doi.org/10.1016/j.gsd.2024.101382>.
- [19] Al-Shabeeb, A. R., Hamdan, I., Al-Fugara, A. K., Al-Adamat, R. and Alrawashdeh, M., (2023). Spatial Mapping of Water Spring Potential Using Four Data Mining Models. *Water Supply*, Vol. 23(5); 1743-1759. <https://doi.org/10.2166/ws.2023.087>.
- [20] Chen, W., Li, Y., Tsangaratos, P., Shahabi, H., Ilija, I., Xue, W. and Bian, H., (2020). Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest and Alternating Decision Tree Models. *Applied Sciences*, Vol. 10(2). <https://doi.org/10.3390/ap10020425>.
- [21] Rahmati, O. and Melesse, A. M., (2016). Application of Dempster–Shafer Theory, Spatial Analysis and Remote Sensing for Groundwater Potentiality and Nitrate Pollution Analysis in the Semi-Arid Region of Khuzestan, Iran. *Science of the Total Environment*, Vol. 568; 1110-1123. <https://doi.org/10.1016/j.scitotenv.2016.06.176>.
- [22] Naghibi, S. A. and Moradi Dashtpajgerdi, M., (2017). Evaluation of Four Supervised Learning Methods for Groundwater Spring Potential Mapping in Khalkhal Region (Iran) Using GIS-Based Features. *Hydrogeology Journal*, Vol. 25(1); 169-189. <https://doi.org/10.1007/s10040-016-1466-z>.

- [23] Golkarian, A., Naghibi, S. A., Kalantar, B. and Pradhan, B., (2018). Groundwater Potential Mapping Using C5.0, Random Forest and Multivariate Adaptive Regression Spline Models in GIS. *Environmental Monitoring and Assessment*, Vol. 190(3). <https://doi.org/10.1007/s10661-018-6507-8>.
- [24] Lee, S., Hong, S. M. and 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*, Vol. 33(8); 847-861. <https://doi.org/10.1080/10106049.2017.1303091>.
- [25] Rudin, C., (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, Vol. 1(5); 206-215. <https://doi.org/10.1038/s42256-019-0048-x>.
- [26] Nguyen, P. T., Bui, D. T., Pradhan, B., Lofman, O. and Revhaug, I., (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(7). <https://doi.org/10.3390/ijerph17072473>.
- [27] 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(1-2); 123-136. <https://doi.org/10.1016/j.jhydrol.2011.05.015>.
- [28] Sheikh, H. A., Bhat, M. S., Alam, A., Ahsan, S. and Shah, B., (2022). Assessing the Groundwater Spring Potential of Sindh Basin in the Kashmir Himalaya. *Arabian Journal of Geosciences*, Vol. 15(23); <https://doi.org/10.1007/s12517-022-10965-y>.
- [29] Jadoud, M., El Achheb, A., Laftouhi, N., Khouz, A., Eloudi, H., El Behari, F., Rachidi, S., Bougadi, B. and Saouab, T., (2023). Assessment of Groundwater Potential in a Mountainous Area Using Machine Learning and GIS, Rherhaya Basin, High Atlas, Morocco. *Iraqi Geological Journal*, Vol. 56(2D); 288-306. <https://doi.org/10.46717/igj.56.2D.22ms-2023-10-28>.
- [30] Gyeltshen, S., Kannaujiya, S., Chhetri, I. K. and Chauhan, P., (2023). Delineating Groundwater Potential Zones Using an Integrated Geospatial and Geophysical Approach in Phuentsholing, Bhutan. *Acta Geophysica*, Vol. 71. <https://doi.org/10.1007/s11600-022-00856-x>.
- [31] Thapa, M., Yuden, K. and Acharya, S., (2022). Exploration of Groundwater Potential Zones: A Case Study in Phuentsholing. *Bhutan Journal of Research and Development*. <https://doi.org/10.17102/bjrd.rub.10.2.012>.
- [32] Thongley, T., Choki, T. and Choden, K., (2023). Groundwater Potential Mapping at Highly Populated Sub-Districts of Thimphu District Using Fuzzy Analytic Hierarchy Process. *Applied Environmental Research*, Vol. 45(2). <https://doi.org/10.35762/AER.2023012>.
- [33] United Nations Development Programme (UNDP), (2016). A Community-Based Approach to Resilient and Sustainable Landscapes: Lessons from Phase II of the COMDEKS Programme. United Nations Development Programme (UNDP), New York, USA. [Online]. Available: <https://comdeksproject.com/knowledge-management-products/publications/>. [Accessed: Oct. 20, 2025].
- [34] Hoy, A., Katel, O., Thapa, P., Dendup, N. and Matschullat, J., (2016). Climatic Changes and their Impact on Socio-Economic Sectors in the Bhutan Himalayas: An Implementation Strategy. *Regional Environmental Change*, Vol. 16; 1-15. <https://doi.org/10.1007/s10113-015-0868-0>.
- [35] Jambay, J. and Uden, K., (2023). *Springsheds: Hydrogeological Mapping Recharge Areas and Assessing Drying Spring Water Quality in Southwestern Bhutan (Volume II)*. Report, Bhutan.
- [36] Dendup, T., Tshering, D., Tobgay, S. and Liu, F., (2024). Sources and Pathways of Spring Flow and Climate Change Effects in the Dungju Ri and Yude Ri Catchments, Bhutan Himalaya. *Heliyon*, Vol. 10(16). <https://doi.org/10.1016/j.heliyon.2024.e36211>.
- [37] Dahal, K., Sharma, S., Shakya, A., Talchabhadel, R., Adhikari, S., Pokharel, A., Sheng, Z., Pradhan, A. M. S. and Kumar, S., (2023). Identification of Groundwater Potential Zones in Data-Scarce Mountainous Region Using Explainable Machine Learning. *Journal of Hydrology*, Vol. 627. <https://doi.org/10.1016/j.jhydrol.2023.130417>.
- [38] Senay, S. D., Worner, S. P. and Ikeda, T., (2013). Novel Three-Step Pseudo-Absence Selection Technique for Improved Species Distribution Modelling. *PLoS ONE*, Vol. 8. <https://doi.org/10.1371/journal.pone.0071218>.
- [39] Pradhan, A. M. S., Kim, Y. T., Shrestha, S., Huynh, T. C. and Nguyen, B. P., (2021). Application of Deep Neural Network to Capture Groundwater Potential Zone in Mountainous Terrain, Nepal Himalaya.

- Environmental Science and Pollution Research*, Vol. 28(15); 18501–18517. <https://doi.org/10.1007/s11356-020-10646-x>.
- [40] Barbet-Massin, M., Jiguet, F., Albert, C. H. and Thuiller, W., (2012). Selecting Pseudo-Absences for Species Distribution Models: How, Where and How Many. *Methods in Ecology and Evolution*, Vol. 3(2); 327-338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- [41] Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S., (2009). Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data. *Ecological Applications*, Vol. 19(1); 181–197. <https://doi.org/10.1890/07-2153.1>.
- [42] Sivakumar, M., Parthasarathy, S. and Padmapriya, T., (2024). Trade-Off Between Training and Testing Ratio in Machine Learning for Medical Image Processing. *PeerJ Computer Science*, Vol. 10. <https://doi.org/10.7717/peerj-cs.2245>.
- [43] Gholamy, A., Kreinovich, V. and Kosheleva, O., (2018). *Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation*. Departmental Technical Reports (CS), University of Texas at El Paso, USA. [Online]. Available: https://scholarworks.utep.edu/cs_techrep/1209. [Accessed: Jun. 25, 2025].
- [44] Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I. and Pham, B. T., (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, Vol. 2021. <https://doi.org/10.1155/2021/4832864>.
- [45] Nguyen, T., Thunyawatcharakul, P., Nguyen, H. N. and Chotpantarat, S., (2022). Global Review of Groundwater Potential Models in the Last Decade: Parameters, Model Techniques and Validation. *Journal of Hydrology*, Vol. 614. <https://doi.org/10.1016/j.jhydrol.2022.128501>.
- [46] Diaz-Alcaide, S. and Martinez-Santos, P., (2019). Advances in Groundwater Potential Mapping: A Review. *Hydrogeology Journal*, Vol. 27(7); 2307-2324. <https://doi.org/10.1007/s10040-019-02001-3>.
- [47] Moore, I. D., Grayson, R. B. and Ladson, A. R., (1991). Digital Terrain Modelling: A Review of Hydrological, Geomorphological and Biological Applications. *Hydrological Processes*, Vol. 5(1); 3-30. <https://doi.org/10.1002/hyp.3360050103>.
- [48] Ali, H. Y., Priju, C. P. and Prasad, N. B. N., (2015). Delineation of Groundwater Potential Zones in Deep Midland Aquifers Along Bharathapuzha River Basin, Kerala Using Geophysical Methods. *Aquatic Procedia*, Vol. 4; 1039-1046. <https://doi.org/10.1016/j.aqpro.2015.02.131>.
- [49] Magesh, N. S., Chandrasekar, N. and Soundranayagam, J. P., (2012). Delineation of Groundwater Potential Zones in Theni District, Tamil Nadu Using Remote Sensing, GIS and MIF Techniques. *Geoscience Frontiers*, Vol. 3(2); 189-196. <https://doi.org/10.1016/j.gsf.2011.10.007>.
- [50] Prasad, P., Loveson, V. J., Kotha, M. and Yadav, R., (2020). Application of Machine Learning Techniques in Groundwater Potential Mapping Along the West Coast of India. *GIScience and Remote Sensing*, Vol. 57(6); 735-752. <https://doi.org/10.1080/15481603.2020.1794104>.
- [51] Tahmassebpour, N., Rahmati, O., Noormohamadi, F. and Lee, S., (2016). Spatial Analysis of Groundwater Potential Using Weights-of-Evidence and Evidential Belief Function Models and Remote Sensing. *Arabian Journal of Geosciences*, Vol. 9(1). <https://doi.org/10.1007/s12517-015-2166-z>.
- [52] Xie, X., Wang, Y., Ellis, A., Liu, C., Duan, M. and Li, J., (2014). Impact of Sedimentary Provenance and Weathering on Arsenic Distribution in Aquifers of the Datong Basin, China: Constraints from Elemental Geochemistry. *Journal of Hydrology*, Vol. 519; 3541-3549. <https://doi.org/10.1016/j.jhydrol.2014.10.044>.
- [53] Long, S., McQuarrie, N., Tobgay, T., Grujic, D. and Hollister, L., (2011). Geologic Map of Bhutan. *Journal of Maps*. <https://doi.org/10.4113/jom.2011.1159>.
- [54] Nhu, V. H., Rahmati, O., Falah, F., Shojaei, S., Al-Ansari, N., Shahabi, H., Shirzadi, A., Górski, K., Nguyen, H. and Ahmad, B. B., (2020). Mapping of Groundwater Spring Potential in Karst Aquifer System Using Novel Ensemble Bivariate and Multivariate Models. *Water*, Vol. 12(4). <https://doi.org/10.3390/w12040985>
- [55] Shah, R. A. and Lone, S. A., (2019). Hydrogeomorphological Mapping Using Geospatial Techniques for Assessing the Groundwater Potential of Rambiar River Basin, Western Himalayas. *Applied Water Science*, Vol. 9(3). <https://doi.org/10.1007/s13201-019-0941-9>.

- [56] Haridas, V., Aravindan, S. and Girish, G., (1998). Remote Sensing and Its Applications for Groundwater Favourable Area Identification. *Quarterly Journal of GARC*, Vol. 6; 18-22. <https://doi.org/10.1007/BF03007340>
- [57] Gansser, A., (1983). *Geology of the Bhutan Himalaya*. Birkhäuser Verlag, Boston, USA.
- [58] RECS International Inc., Research Institute for Urban and Environmental Development (Japan), Nippon Koei Co. Ltd., and Kokusai Kogyo Co. Ltd., (2019). The project for formulation of comprehensive development plan for Bhutan 2030: final report, Volume II: existing conditions and development issues. Ministry of Works and Human Settlement (MoWHS) and Japan International Cooperation Agency (JICA), Thimphu, Bhutan, Technical Report. [Online]. Available: https://openjicareport.jica.go.jp/360/360/360_102_12340576.html. [Accessed: Oct. 20, 2025].
- [59] Strahler, A. N., (1964). Quantitative Geomorphology of Drainage Basins and Channel Networks. In: Handbook of Applied Hydrology. Chow, V.T. (Ed.), McGraw-Hill, New York, 4-39.
- [60] Tucker, G. E., Catani, F., Rinaldo, A. and Bras, R. L., (2001). Statistical Analysis of Drainage Density from Digital Terrain Data. *Geomorphology*, Vol. 36(3); 187-202. [https://doi.org/10.1016/S0169-555X\(00\)00056-8](https://doi.org/10.1016/S0169-555X(00)00056-8).
- [61] Bagyaraj, M., Ramkumar, T., Venkatramanan, S. and Gurugnanam, B., (2013). Application of Remote Sensing and GIS Analysis for Identifying Groundwater Potential Zone in Parts of Kodaikanal Taluk, South India. *Frontiers of Earth Science*, Vol. 7(1); 65-75. <https://doi.org/10.1007/s11707-012-0347-6>.
- [62] Ghosh, P. K., Bandyopadhyay, S. and Jana, N. C., (2015). Mapping of Groundwater Potential Zones in Hard Rock Terrain Using Geoinformatics: A Case of Kumari Watershed in Western Part of West Bengal. *Modeling Earth Systems and Environment*, Vol. 2(1). <https://doi.org/10.1007/s40808-015-0044-z>.
- [63] Shekhar, S. and Pandey, A. C., (2015). Delineation of Groundwater Potential Zone in Hard Rock Terrain of India Using Remote Sensing, Geographical Information System (GIS) and Analytic Hierarchy Process (AHP) Techniques. *Geocarto International*, Vol. 30(4); 402-421. <https://doi.org/10.1080/10106049.2014.894584>.
- [64] Agarwal, R. and Garg, P., (2016). Remote Sensing and GIS-Based Groundwater Potential and Recharge Zones Mapping Using Multi-Criteria Decision Making Technique. *Water Resources Management*. <https://doi.org/10.1007/s11269-015-1159-8>.
- [65] Chotpantarat, S., Wongsasuluk, P., Siriwong, W., Borjan, M. and Robson, M., (2014). Non-Carcinogenic Hazard Maps of Heavy Metal Contamination in Shallow Groundwater for Adult and Aging Populations at an Agricultural Area in Northeastern Thailand. *Human and Ecological Risk Assessment: An International Journal*, Vol. 20(3); 689-703. <https://doi.org/10.1080/10807039.2013.832998>.
- [66] Tarafdar, S., Bruijnzeel, L. A. and Kumar, B., (2019). Improved Understanding of Spring and Stream Water Responses in Headwaters of the Indian Lesser Himalaya Using Stable Isotopes, Conductivity and Temperature as Tracers. *Hydrological Sciences Journal*, Vol. 64(7); 757-770. <https://doi.org/10.1080/02626667.2019.1600698>.
- [67] Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K. and Helkowski, J. H., (2005). Global Consequences of Land Use. *Science*, Vol. 309; 570-574. <https://doi.org/10.1126/science.1111772>.
- [68] Jacobson, C. R., (2011). Identification and Quantification of the Hydrological Impacts of Imperviousness in Urban Catchments: A Review. *Journal of Environmental Management*, Vol. 92(6); 1438-1448. <https://doi.org/10.1016/j.jenvman.2011.01.018>
- [69] Owuor, S. O., Butterbach-Bahl, K., Guzha, A. C., Rufino, M. C., Pelster, D. E., Díaz-Pinés, E. and Breuer, L., (2016). Groundwater Recharge Rates and Surface Runoff Response to Land Use and Land Cover Changes in Semi-Arid Environments. *Ecological Processes*, Vol. 5(1). <https://doi.org/10.1186/s13717-016-0060-6>.
- [70] Caschetto, M., Sacchi, E., Pinti, D. L., Riparbelli, C., Bruno, S., Zanotti, C., Bonomi, T. and Rotiroti, M., (2025). Surface-Water-Irrigation Return Flow Dominates Groundwater Recharge, Groundwater Age and Nitrate Dynamics in an Alluvial Basin Aquifer. *Water Research*, Vol. 285. <https://doi.org/10.1016/j.watres.2025.124040>.
- [71] Al-Ghazal, A., (2000). Effect of Tillage Practices and Compaction on Infiltration Rate of Sandy Soils. *Pakistan Journal of Biological Sciences*, Vol. 3(9). <https://doi.org/10.3923/pjbs.2000.1443.1446>.

- [72] Sun, J., He, F., Zhang, Z., Shao, H., Pan, Y., Yang, R., Li, W., Li, P. and Zheng, M., (2018). Analysis of Saline Groundwater Infiltration into Two Loam Soils. *Land Degradation and Development*, Vol. 29(10); 3795-3802. <https://doi.org/10.1002/ldr.3089>.
- [73] Oh, H. J., Kim, Y., Choi, J. K., Park, E. and Lee, S., (2011). GIS Mapping of Regional Probabilistic Groundwater Potential in the Area of Pohang City, Korea. *Journal of Hydrology*, Vol. 399; 158-172. <https://doi.org/10.1016/j.jhydrol.2010.12.027>.
- [74] National Soil Services Centre (NSSC), (2023). Soil classes of Bhutan. National Soil Services Centre, Bhutan. [Online]. Available: <https://drive.google.com/file/d/1YIFE60M100vEJjnG7e4Uv7RtFiabQBnT/view?usp=sharing>. [Accessed: Feb. 21, 2025]
- [75] Food and Agriculture Organization., (2015). World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps – Update 2015. *World Soil Resources Reports*, Vol. 106; FAO, Rome, Italy. [Online]. Available: www.fao.org/publications. [Accessed: Jan. 12, 2025]
- [76] Deng, H., Wu, X., Zhang, W., Liu, Y., Li, W., Li, X., Zhou, P. and Zhuo, W., (2022). Slope-Unit Scale Landslide Susceptibility Mapping Based on the Random Forest Model in Deep Valley Areas. *Remote Sensing*, Vol. 14(17). <https://doi.org/10.3390/rs14174245>.
- [77] Papageorgiou, S., (2022). On Correlation Coefficients and their Interpretation. *Journal of Orthodontics*, Vol. 49; 359-361. <https://doi.org/10.1177/14653125221076142>.
- [78] Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J. and Münkemüller, T., (2013). Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography*, Vol. 36; 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- [79] Long, G., Tantane, S., Nusit, K. and Sooraksa, P., (2025). Flood Susceptibility Mapping in the Yom River Basin, Thailand: Stacking Ensemble Learning Using Multi-Year Flood Inventory Data. *Geocarto International*. Vol. 40(1). <https://doi.org/10.1080/10106049.2025.2461531>.
- [80] Bui, D. T., Tuan, T., Klempe, H., Pradhan, B. and Revhaug, I., (2015). Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression and Logistic Model Tree. *Landslides*, Vol. 13; 1-18. <https://doi.org/10.1007/s10346-015-0557-6>.
- [81] Kim, J. H., (2019). Multicollinearity and Misleading Statistical Results. *Korean Journal of Anesthesiology*, Vol. 72(6); 558-569. <https://doi.org/10.4097/kja.19087>.
- [82] Mangrulkar, R. S., Michalas, A., Shekokar, N. M., Narvekar, M. and Chavan, P. V., (Eds.), (2021). *Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques*. CRC Press, Boca Raton.
- [83] Zhao, D., Miao, F., Chen, Y., Wu, Q., Feng, G., Chang, B., Su, H., Ren, P., Hao, C., Li, Z. and Li, X., (2025). Enhanced Groundwater Potential Mapping Using a GIS-Based Chaotic Sparrow Search Algorithm Optimized Weighted Broad Learning System: A Case Study of the Guozhuang Spring Region, Northern China. *Journal of Hydrology: Regional Studies*, Vol. 61. <https://doi.org/10.1016/j.ejrh.2025.102708>.
- [84] Wang, Z., Le, T., Tian, K., Phong, T. V., Bien, T. X. and Pham, B. T., (2024). Novel Ensemble Models Based on the Split-Point Sampling and Node Attribute Subsampling Classifier for Groundwater Potential Mapping. *Earth and Space Science*, Vol. 11(6). <https://doi.org/10.1029/2023EA003338>.
- [85] Fawcett, T., (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*, Vol. 27(8); 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- [86] Pham, B. T., Nguyen-Thoi, T., Ly, H. B., Nguyen, M. D., Al-Ansari, N., Tran, V. Q. and Le, T. T., (2020). Extreme Learning Machine Based Prediction of Soil Shear Strength: A Sensitivity Analysis Using Monte Carlo Simulations and Feature Backward Elimination. *Sustainability*, Vol. 12(6). <https://doi.org/10.3390/su12062339>.
- [87] Rainio, O., Teuhio, J. and Klén, R., (2024). Evaluation Metrics and Statistical Tests for Machine Learning. *Scientific Reports*, Vol. 14(1). <https://doi.org/10.1038/s41598-024-56706-x>.

- [88] Hamamin, D. F., Othman, A. A. and Al-Manmi, D. A. M. A., (2023). Comparison of Different Statistical Approaches to Assess Spring Potential Mapping in a Multi-Aquifer System: A Case Study of Kurdistan Region, Iraq. *Environmental Earth Sciences*, Vol. 82(24); 606. <https://doi.org/10.1007/s12665-023-11250-1>.
- [89] Mersha, T. and Meten, M., (2020). GIS-Based Landslide Susceptibility Mapping and Assessment Using Bivariate Statistical Methods in Simada Area, Northwestern Ethiopia. *Geoenvironmental Disasters*, Vol. 7(1). <https://doi.org/10.1186/s40677-020-00155-x>.
- [90] Harrell, F. E. Jr., (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, 2nd ed., Springer, New York.
- [91] Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X., (2013). *Applied Logistic Regression*, 3rd ed., John Wiley & Sons, New York.
- [92] Menard, S., (2010). *Logistic Regression: From Introductory to Advanced Concepts and Applications*. SAGE Publications, Thousand Oaks, California, USA. [Online]. Available: <https://methods.sagepub.com/book/mono/logistic-regression-from-introductory-to-advanced-concepts-and-applications/toc>. [Accessed: March. 28, 2026].
- [93] Jenks, G. F., (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, Vol. 7; 186-190.
- [94] Arabameri, A., Roy, J., Saha, S., Blaschke, T., Ghorbanzadeh, O. and Bui, D. T., (2019). Application of Probabilistic and Machine Learning Models for Groundwater Potentiality Mapping in Damghan Sedimentary Plain, Iran. *Remote Sensing*, Vol. 11(24). <https://doi.org/10.3390/rs11243015>.
- [95] Singha, C., Swain, K. C., Pradhan, B., Rusia, D. K., Moghimi, A. and Ranjgar, B., (2024). Mapping Groundwater Potential Zone in the Subarnarekha Basin, India, Using a Novel Hybrid Multi-Criteria Approach in Google Earth Engine. *Heliyon*, Vol. 10(2). <https://doi.org/10.1016/j.heliyon.2024.e24308>.
- [96] Brewer, C. A. and Pickle, L., (2002). Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. *Annals of the Association of American Geographers*, Vol. 92(4); 662-681. <https://doi.org/10.1111/1467-8306.00310>.
- [97] Gupta, V., Kumar, S., Kaur, R. and Tandon, R., (2022). Regional-Scale Landslide Susceptibility Assessment for the Hilly State of Uttarakhand, NW Himalaya, India. *Journal of Earth System Science*. <https://doi.org/10.1007/s12040-021-01746-4>.
- [98] Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C. and Bălteanu, D., (2016). Different Landslide Sampling Strategies in a Grid-Based Bivariate Statistical Susceptibility Model. *Geomorphology*, Vol. 253; 508-523. <https://doi.org/10.1016/j.geomorph.2015.10.030>.
- [99] Guru, B., Seshan, K. and Bera, S., (2017). Frequency Ratio Model for Groundwater Potential Mapping and Its Sustainable Management in Cold Desert, India. *Journal of King Saud University - Science*, Vol. 29(3); 333-347. <https://doi.org/10.1016/j.jksus.2016.08.003>.
- [100] Naghibi, A. and Pourghasemi, H., (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*, Vol. 29; 5217-5236. <https://doi.org/10.1007/s11269-015-1114-8>.
- [101] Bai, S. B., Wang, J., Lü, G. N., Zhou, P. G., Hou, S. S. and Xu, S. N., (2010). GIS-Based Logistic Regression for Landslide Susceptibility Mapping of the Zhongxian Segment in the Three Gorges Area, China. *Geomorphology*, Vol. 115(1); 23-31. <https://doi.org/10.1016/j.geomorph.2009.09.025>.
- [102] Zandi, J., Ghazvinei, P. T., Hashim, R., Yusof, K. B. W., Ariffin, J. and Motamedi, S., (2016). Mapping of Regional Potential Groundwater Springs Using Logistic Regression Statistical Method. *Water Resources*, Vol. 43(1); 48-57. <https://doi.org/10.1134/S0097807816010097>.
- [103] Fan, Y., Clark, M., Lawrence, D. M., Swenson, S. C., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G. and Kirchner, J. W., (2019). Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resources Research*, Vol. 55(2); 1737-1772. <https://doi.org/10.1029/2018WR023903>.
- [104] Morbidelli, R., Saltalippi, C., Flammini, A. and Govindaraju, R. S., (2018). Role of Slope on Infiltration: A Review. *Journal of Hydrology*, Vol. 557; 878-886. <https://doi.org/10.1016/j.jhydrol.2018.01.019>.

- [105] Dar, T., Rai, N. and Bhat, A., (2021). Delineation of Potential Groundwater Recharge Zones Using Analytical Hierarchy Process (AHP). *Geology, Ecology, and Landscapes*, Vol. 5(4); 292-307. <https://doi.org/10.1080/24749508.2020.1726562>.
- [105] Negi, G. C. S. and Joshi, V., (2004). Rainfall and Spring Discharge Patterns in Two Small Drainage Catchments in the Western Himalayan Mountains, India. *Environmentalist*, Vol. 24(1); 19-28. <https://doi.org/10.1023/B:ENVR.0000046343.45118.78>.
- [107] Matheswaran, K., Khadka, A., Dhaubanjari, S., Bharati, L., Kumar, S. and Shrestha, S., (2019). Delineation of Spring Recharge Zones Using Environmental Isotopes to Support Climate-Resilient Interventions in Two Mountainous Catchments in Far-Western Nepal. *Hydrogeology Journal*, Vol. 27(6); 2181-2197. <https://doi.org/10.1007/s10040-019-01973-6>.
- [108] Wangmo, C., Wangmo, P., Rinchen, N., Choerzer, N. and Pasang, S., (2025). Predictive Analytics for Identifying Surface Water Sources for Domestic Water Supply in Phuentsholing, Bhutan. *Zorig Melong: A Technical Journal of Science, Engineering and Technology*, Vol. 8(2); 231-238.
- [109] Chen, W., Yang, K., Yu, Z., Shi, Y. and Chen, C. L. P., (2024). A Survey on Imbalanced Learning: Latest Research, Applications and Future Directions. *Artificial Intelligence Review*, Vol. 57(6). <https://doi.org/10.1007/s10462-024-10759-6>.
- [110] He, H. and Garcia, E. A., (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21(9); 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>
- [111] Kumar, M., Sen, S., Kulkarni, H., Badiger, S., Varma, G. R. and Krishnaswamy, J., (2024). Ecohydrological and Hydrogeological Dynamics of Groundwater Springs in Eastern Himalaya, India. *Groundwater for Sustainable Development*, Vol. 27. <https://doi.org/10.1016/j.gsd.2024.101311>.
- [112] Pant, N., Hagare, D., Maheshwari, B., Rai, S. P., Patel, A., Puthiyottil, N., Jain, S. K., Sahu, L. N., Sen, S. and Upadhyay, M., (2026). Understanding the Influence of Landcover on Spring Dynamics and Evaporation in the Himalayan Region Using Stable Isotope and Discharge. *Science of the Total Environment*, Vol. 1012. <https://doi.org/10.1016/j.scitotenv.2025.181215>.