Groundwater Potential Mapping using AHP and Machine Learning in the upper Subarnarekha River Basin, India

Farhin Tabassum [1], Akhouri Pramod Krishna [1]

[1] Department of Remote Sensing, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India
Email – phdrs10008.20@bitmesra.ac.in, apkrishna@bitmesra.ac.in

Keywords: Groundwater Potential (GWP), Analytic Hierarchy Process (AHP), Machine Learning, Subarnarekha River Basin (SRB).

Abstract

Water is one of the most vital and dynamic natural resources on Earth, essential for supporting various life forms and human activities. The rapid growth of the population has had a significant impact on the availability and quality of water. Consequently, groundwater has become the primary alternative source of water to meet human needs. Identifying regions with high groundwater potential is crucial for the effective utilization, management, and conservation of water resources. Finding groundwater with reasonable accuracy can be a challenging task due to its complex and nonlinear behavior. However, integrating geographic information systems (GIS) with machine learning techniques has proven to be an effective approach for mapping and predicting groundwater occurrence. This combination allows for a more reliable assessment of the various factors that influence groundwater distribution and availability. The upper Subarnarekha River Basin (SRB) region covers 17,037 square kilometers within the states of Jharkhand and West Bengal, situated between latitudes 22°01' and 23°35' N, and longitudes 85°05' and 86°55' E. The region receives an annual average rainfall of 1300 to 1800 mm, with the heaviest precipitation occurring from June to October. The annual average temperature ranges from 23°C to 28°C. Geologically, the basin comprises three main formations: Pre-Cambrian rocks in the upper and middle areas, and Tertiary and Quaternary formations in the lower area.

In the present study, the Analytical Hierarchy Process (AHP), and two machine learning techniques, Random Forest (RF) and Support Vector Machine (SVM)), have been used to map GroundWater Potential Zones (GWPZ) in the upper SRB, India. The Analytical Hierarchical Process (AHP), is a hierarchical weighting method for multi-criteria decision problems. AHP evaluates parameters based on their relative importance and ranks them through pairwise comparisons to organize criteria in a hierarchical structure. Each parameter is assigned a set of weights and the data is normalized. Lastly, the consistency index (CI) and consistency ratio (CR) are estimated to verify the reliability of the assigned weights. Random Forest (RF) is a widely used and highly accurate machine learning algorithm known for its tree-based method. It combines multiple decision trees to assess the relationships between factors. Random Forest generates a "forest" of trees using bootstrapped data. The Support Vector Machine (SVM) model is a widely used supervised machine learning technique grounded in structural risk minimization and statistical learning theory. The method works by separating the dataset with a hyperplane, which is defined as the center of the maximum margin of separation. To achieve the study's goal, ten factors influencing groundwater occurrence were chosen for inter-thematic correlation analysis and overlaid with well locations. The key factors that influence groundwater occurrence include Geology, Geomorphology, Elevation, Drainage Density (DD), Lineament Density (LD), Land Use Land Cover (LULC), Rainfall, Soil, Slope, and Topographical Wetness Index (TWI). A total of 81 well locations were obtained from the Central Groundwater Board (CGWB) and split into two groups: training and validation, with a distribution ratio of 70:30. The AHP, RF, and SVM models have been used to delineate groundwater potential zones and are classified into three categories: High, moderate and Low. The study also concludes that geomorphology, slope, rainfall, and elevation play a more significant role in influencing groundwater potential zones (GWPZs). Results reveal that approximately 18% of the area has good Ground Water Potential (GWP), primarily located in the southeast region. Additionally, around 57% of the area demonstrated moderate GWP in the central part of the basin, whereas 25% of the area shows poor GWP in the northeastern regions. The model's performance has been evaluated using methods such as receiver operating characteristics (ROC), accuracy (ACC), and kappa coefficient. The performance of the machine learning (ML) models with the limited dataset (81 well data points across an area of 17,037 sq km) was not much satisfactory. The data scarcity negatively impacted the performance of the ML models, while the WOA approach showed more satisfactory results. The study shows the effectiveness and limitations of user knowledge-based AHP methods and machine learning techniques in effectively mapping groundwater potential areas within a river basin that naturally exhibits low primary porosity due to lithological constraints. The findings can support watershed management efforts and guide the selection of optimal well locations in future projects.

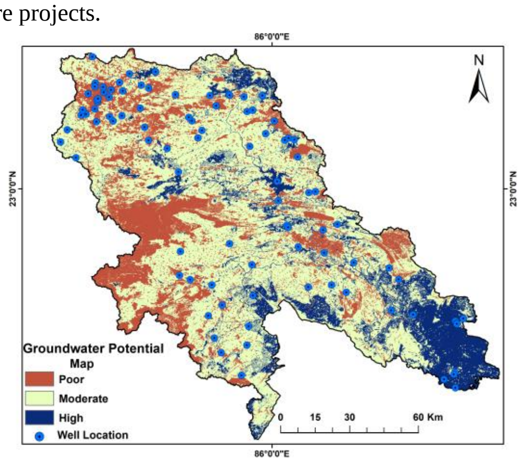

Figure 1: Groundwater Potential Map using AHP