

Machine learning–based groundwater potential mapping in fractured terrain: A case study of the Azrou–Khenifra basin (Central Morocco)

Anouar Taibou^{1*}, Halima Jounaid¹, Nachchach Badr¹, Said Ou Moua¹

¹Laboratory of Geosciences, Geomatics, Planetology and Environment (GEOPEN), Hassan II University, Faculty of Sciences Ain Chok, km 8 Route d'El Jadida, 20150, Casablanca, Morocco

*anouar.taibou-etu@etu.univh2c.ma

ABSTRACT

Identifying groundwater resources in fractured and semi-arid environments remains a major challenge due to geological complexity and limited hydrogeological data. In this context, this study explores the potential of machine learning and ensemble learning approaches for groundwater potential mapping in the Azrou–Khenifra basin (Central Morocco). The methodology integrates twenty groundwater conditioning factors derived from remote sensing data, digital elevation models, geological datasets, and climatic information. These include morphometric parameters, hydrological indices, structural features, lithological characteristics, and environmental indicators such as NDVI, LST, rainfall, and land use/land cover.

Groundwater potential maps were produced using three machine learning algorithms (Random Forest, CatBoost, and K-Nearest Neighbors) and their ensemble combinations (RF–CatBoost, RF–KNN, CatBoost–KNN, and RF–CatBoost–KNN) to evaluate the predictive capability of individual and hybrid models.

The predictive capability of the models was assessed using overall accuracy, precision, recall, F1-score, and ROC–AUC metrics. Results show that CatBoost achieved the best performance among the individual models, with a testing accuracy of **0.76** and an AUC value of **0.8304**, demonstrating strong generalization ability. Ensemble models also showed competitive performance, particularly the RF–CatBoost model (accuracy = **0.74**, AUC = **0.8144**) and CatBoost–KNN (accuracy = **0.72**, AUC = **0.8176**). The combined RF–CatBoost–KNN voting model produced stable and consistent predictions, reaching an accuracy of **0.70** and an AUC of **0.8086**.

Analysis of variable importance indicates that rainfall, fault density, and proximity to faults play a dominant role in groundwater occurrence, highlighting the structural control of groundwater storage in fractured environments. Model validation using the spatial distribution of groundwater productivity further confirmed the reliability of the results, as high-yield water points were mainly located within high and very high groundwater potential zones.

Overall, this study demonstrates that hybrid machine learning approaches constitute effective tools for groundwater exploration and decision support, particularly in heterogeneous and data-limited hydrogeological settings. The proposed framework provides valuable insights for sustainable groundwater management and future hydrogeological investigations in similar fractured basins.

Keywords: Groundwater potential mapping, Machine learning, Ensemble learning, CatBoost, Fractured terrain