



OPEN HydroPredictor a hybrid machine learning model for addressing data scarcity in groundwater prediction

Abdessamad Elmotawakkil¹, Adil Moumane², Assia Zahi¹, Abdelkhalik Sadiki², Jamal Al Karkouri², Mouhcine Batchi², Suraj Kumar Bhagat³✉ & Nourddine Enneya¹

Groundwater prediction in data-scarce and environmentally sensitive regions presents a persistent challenge due to limited observational data, spatial heterogeneity, and the nonlinear nature of hydrogeological processes. In this study, we propose HydroPredictor, a hybrid machine learning framework that integrates the categorical handling efficiency of CatBoost with the nonlinear feature learning capacity of a regularized Multi-Layer Perceptron (MLP). The model was trained on a geo-referenced dataset of 315 samples from the Feija Basin in southeastern Morocco, incorporating ten environmental predictors such as elevation, rainfall, soil permeability, NDVI, and topographic wetness index. The pipeline includes Optuna-based hyperparameter optimization and 5-fold cross-validation to ensure robustness and generalization. HydroPredictor achieved a testing accuracy of 89.23%, with an F1-score of 0.8937 and Area Under the Curve (AUC) values exceeding 0.90 across all groundwater potential classes. Statistical validation using the Friedman and Wilcoxon signed-rank tests ($p < 0.05$) confirmed its significant outperformance over conventional models, including Random Forest, Support Vector Machine (SVM), and standalone MLP. Furthermore, HydroPredictor demonstrated superior generalization compared to prior models in the literature (e.g., RF-SSA: AUC = 0.840; GBDT: AUC = 0.88), while maintaining minimal overfitting (Δ Accuracy = 0.35%). By combining interpretable tree-based embeddings with deep neural representations, HydroPredictor provides a robust and scalable solution for groundwater classification in data-limited settings, offering a reproducible and operationally relevant tool for sustainable groundwater resource management under climatic and environmental uncertainty.

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural networks
AUC	Area under the curve
BS-MLP	GBDT + MLP
DT	Decision trees
ELM	Extreme learning machine
GBDT	Gradient boosted decision trees
GRACE	Gravity recovery and climate experiment
IDW	Inverse distance weighting
KNN	K-nearest neighbors
LightGBM	Light gradient boosting machine
LR	Logistic regression
LSTM	Long short-term memory
LULC	Land use–land cover
MBAB	MultiBoostAB
MCC	Matthews correlation coefficient
MCDA	Multi-criteria decision analysis
MI	Mutual information
ML	Machine learning

¹Department of Computer Science, Faculty of Sciences, University Ibn Tofail, Kenitra, Morocco. ²Department of Geography, Faculty of Humanities and Social Sciences, University Ibn Tofail, Kenitra, Morocco. ³Centre for Interdisciplinary Research, SRM University - AP, Amaravati, Andhra Pradesh 522240, India. ✉email: surajkumar.b@srmmap.edu.in

MLP	Multi-layer perceptron
MODIS	Moderate resolution imaging spectroradiometer
NB	Naive bayes
NDVI	Normalized difference vegetation index
RBF	Radial basis function
RF	Random forests
ROC	Receiver operating characteristic
ROC-AUC	Receiver operating characteristic - area under curve
SGD	Stochastic gradient descent
SHAP	SHapley additive explanations
SPAARC	Split-point sampling and node attribute subsampling classifier
SSA	Sparrow search algorithm
SVM	Support vector machine
TPE	Tree-structured Parzen estimator
TWI	Topographic wetness index

Groundwater is a vital resource for ensuring potable water supply, agricultural productivity, and socioeconomic resilience, particularly in arid and semi-arid regions where surface water availability is limited or highly prone to contamination^{1,2}. In countries such as Morocco, groundwater serves as a strategic water source, increasingly relied upon to meet domestic and agricultural demands under growing climatic pressures^{3,4}. However, sustainable management of this resource is often impeded by the absence of dense monitoring networks, limited hydrogeological data, and fragmented assessments of subsurface conditions. In this context, the delineation of groundwater potential zones has emerged as a critical task for supporting informed water planning and safeguarding long-term aquifer viability⁵.

Recent advances in Artificial Intelligence (AI), particularly in the domain of Machine Learning (ML), have offered promising alternatives to traditional physically based hydrological models^{6,7}. ML algorithms such as Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and boosting techniques have demonstrated strong capabilities in groundwater potential mapping due to their capacity to model complex nonlinear relationships among environmental predictors^{8,9}. These models have been applied using spatial datasets such as elevation, slope, land use, rainfall, soil permeability, and vegetation indices to classify groundwater-rich zones and guide exploration and development.

Despite their promise, the application of ML models in groundwater studies is constrained by data scarcity, particularly in developing regions and environmentally vulnerable basins^{10,11}. Under such conditions, traditional models tend to suffer from overfitting, failing to generalize due to limited training samples and high feature variability¹². Additionally, standard hyperparameter tuning techniques such as GridSearch are computationally inefficient, requiring exhaustive iteration across parameter spaces and imposing significant processing burdens in practice¹³. These limitations point to a pressing need for robust and data-efficient modeling frameworks capable of generalizing effectively in sparse and heterogeneous environments. To address these challenges, this study proposes *HydroPredictor*, a novel hybrid artificial intelligence framework for groundwater potential zone mapping in data-scarce contexts. The model integrates the categorical feature encoding and gradient-boosted structure of CatBoost with the nonlinear pattern recognition capability of a regularized Multi-Layer Perceptron (MLP). Furthermore, it employs Optuna, a Bayesian optimization method, to efficiently tune hyperparameters and minimize computational complexity. Unlike conventional approaches, *HydroPredictor* is specifically designed to mitigate overfitting, enhance interpretability, and operate effectively in regions with limited geospatial and hydrological data.

By enabling precise spatial delineation of groundwater potential zones, *HydroPredictor* functions as a robust and interpretable decision-support framework tailored for sustainable groundwater resource management. Its integrative design facilitates the identification of optimal recharge zones, enhances strategic planning for aquifer replenishment, and strengthens equitable access to potable water. Moreover, the model contributes to climate-resilient water governance by supporting adaptive management strategies in hydrologically stressed and data-scarce environments.

Literature review

Recent advances in artificial intelligence (AI) and machine learning (ML) have substantially transformed groundwater potential mapping, offering novel tools to address the complexity and nonlinearity of hydrogeological systems. A wide range of ML algorithms such as Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Logistic Regression (LR), and Naive Bayes (NB) have been applied to classify groundwater potential zones and assess groundwater quality across various climatic and geological contexts¹⁴.

Several studies have demonstrated the predictive power of these models. For instance¹⁵, utilized ML models, including ANN, to forecast groundwater potential zones in Bangladesh under climate change scenarios, achieving an Area Under the Curve (AUC) of 0.875. Similarly¹⁶, applied ensemble approaches, reporting RF performance with an F1-score of 0.944 and accuracy of 0.943 in delineating groundwater zones. In Vietnam's Sesan River Basin¹⁷, proposed the MultiBoostAB- Sampling and Node Attribute Subsampling Classifier (MBAB-SPAARC) model, which yielded an AUC of 0.891, outperforming conventional methods.

In Morocco¹⁸, developed a hybrid ensemble model combining RF, SVM, and LR to map groundwater potential in the Saïss Basin, achieving an AUC of 0.86 and outperforming both classical and deep learning alternatives. Additionally¹⁹, demonstrated that RF yielded superior accuracy (0.92) and AUC (0.95) for groundwater quality classification in northern Iran. Efforts to reduce model complexity while preserving predictive performance

were examined by²⁰, who used minimal-input (Light Gradient Boosting Machine (LightGBM) models with only six features, achieving an F1-score of 91.08%. Hybrid architectures, such as the BS-MLP model developed by²¹, which combined Gradient Boosted Decision Trees (GBDT) with Multi-Layer Perceptrons (MLP), demonstrated 97.6% accuracy and 96.8% F1-score, surpassing individual models. Optimization techniques have also played a critical role in improving ML-based groundwater models. For instance²², implemented the Sparrow Search Algorithm (SSA) to tune RF hyperparameters, achieving an AUC of 0.840. Comparatively²³, evaluated GBDT, which attained an AUC of 0.88 and an accuracy of 0.89 in spring potential prediction. Meanwhile, pre-processing and feature engineering strategies, as discussed in²⁴, have been shown to substantially influence model performance, with F1-scores exceeding 0.87 after optimal data preparation.

Despite these promising developments, several gaps remain. Most current studies rely on large, well-curated datasets, limiting their applicability in data-scarce environments. Moreover, existing models often suffer from overfitting when applied to regions with sparse monitoring networks or heterogeneous hydrological conditions. Traditional hyperparameter tuning methods, such as Grid Search, are computationally intensive and inefficient in practice, particularly when scaling models or adapting them to new regions.

This study addresses these gaps by introducing HydroPredictor, a hybrid AI framework that integrates robust tree-based learning (CatBoost) with neural representation learning (MLP), and employs efficient Bayesian optimization (Optuna) to overcome limitations related to overfitting, scalability, and model adaptability in data-scarce groundwater studies. Unlike earlier hybrid approaches (e.g., GBDT–MLP), the novelty of HydroPredictor lies in demonstrating its resilience under extreme data scarcity. With only 315 samples available, high-capacity models such as MLP are normally prone to overfitting; however, the integration of CatBoost's feature transformation with a regularized MLP, stabilized through Optuna-based hyperparameter tuning and 5-fold cross-validation, enables reliable performance despite limited data. By building on the strengths of prior research while addressing their shortcomings, HydroPredictor contributes a novel, efficient, and interpretable approach for sustainable groundwater potential mapping.

Methodology

To develop a reliable and efficient groundwater prediction model, a structured methodology was adopted, integrating both traditional machine learning techniques and advanced hybrid modeling approaches. The proposed framework follows a systematic process that includes ground truth data labeling, data collection, preprocessing, feature selection, and normalization ensuring that only the most relevant variables contribute to model training. To enhance model performance and mitigate overfitting, a 5-fold cross-validation strategy was implemented. Additionally, Optuna was employed for hyperparameter tuning, optimizing the parameters of SVM, RF, MLP, CatBoost, and the proposed HydroPredictor model. HydroPredictor introduces a hybrid approach that combines tree-based feature extraction with deep learning classification, leveraging leaf embeddings and a meta-classifier to achieve improved predictive accuracy and stability. Finally, model performance was assessed using standard evaluation metrics, including accuracy, F1-score, recall, Cohen's kappa, and Matthews correlation coefficient (MCC), to determine the most effective model for groundwater prediction.

Study area

In Morocco's Draa Valley, prolonged droughts exacerbated by climate change are driving farmers to migrate beyond traditional oasis boundaries, seeking access to deeper groundwater reserves as a vital adaptation strategy^{25–34}.

The Feija Plain an intra-mountainous alluvial zone situated just west of Zagora city in the Middle Draa watershed is located entirely outside traditional oasis zones Fig. 1. In recent years, it has transitioned from former rangeland into an intensive agricultural frontier. Farmers, displaced by the drying of classic oases, now drill into quaternary and primary aquifers to access groundwater resources. However, in the Feija Plain pivotal for watermelon cultivation this trend has led to alarming groundwater depletion^{35–38}. Groundwater levels dropped by nearly 10 m between 2013 and 2018³⁹, while salinity has risen markedly^{40,41}.

Paradoxically, the region has also experienced significant flood events in recent years^{42,43}, yet remains bereft of rainwater- harvesting infrastructure. To address this, several studies⁴⁴ have applied GIS-AHP techniques to identify suitable sites for artificial recharge in the Feija Plain. Their findings indicate that only about 5.24% of the area is highly suitable for rainwater- harvesting structures. This limited extent of favorable zones, coupled with the absence of existing infrastructure and the scarcity of hydrogeological data, underscores the urgent challenges facing sustainable water management in the region. Meanwhile, broader assessments in the Middle Draa using the WEAP model forecast further reductions in surface water and increased reliance on groundwater under climate-warming scenarios⁴⁵.

Predictor variable selection for groundwater modeling

Groundwater recharge is controlled by a diverse suite of variables topographic, hydrologic, pedologic, geomorphic, and biotic that together determine infiltration, storage, and subsurface flow. Below, each of the ten selected conditioning factors is detailed with current scientific backing:

1. **Elevation:** Elevation influences surface runoff and climatic conditions such as precipitation and temperature, making it a key control on groundwater recharge potential. In recent studies employing both Multi-Criteria Decision Analysis (MCDA) and machine learning (ML), elevation consistently ranks among the top predictive layers for groundwater potential mapping^{46–48}.
2. **Soil Permeability:** Soil texture (e.g., sand vs. clay) controls infiltration capacity and subsurface flow. Several AHP and ML-based models highlight soil permeability as a critical recharge determinant^{49,50}.

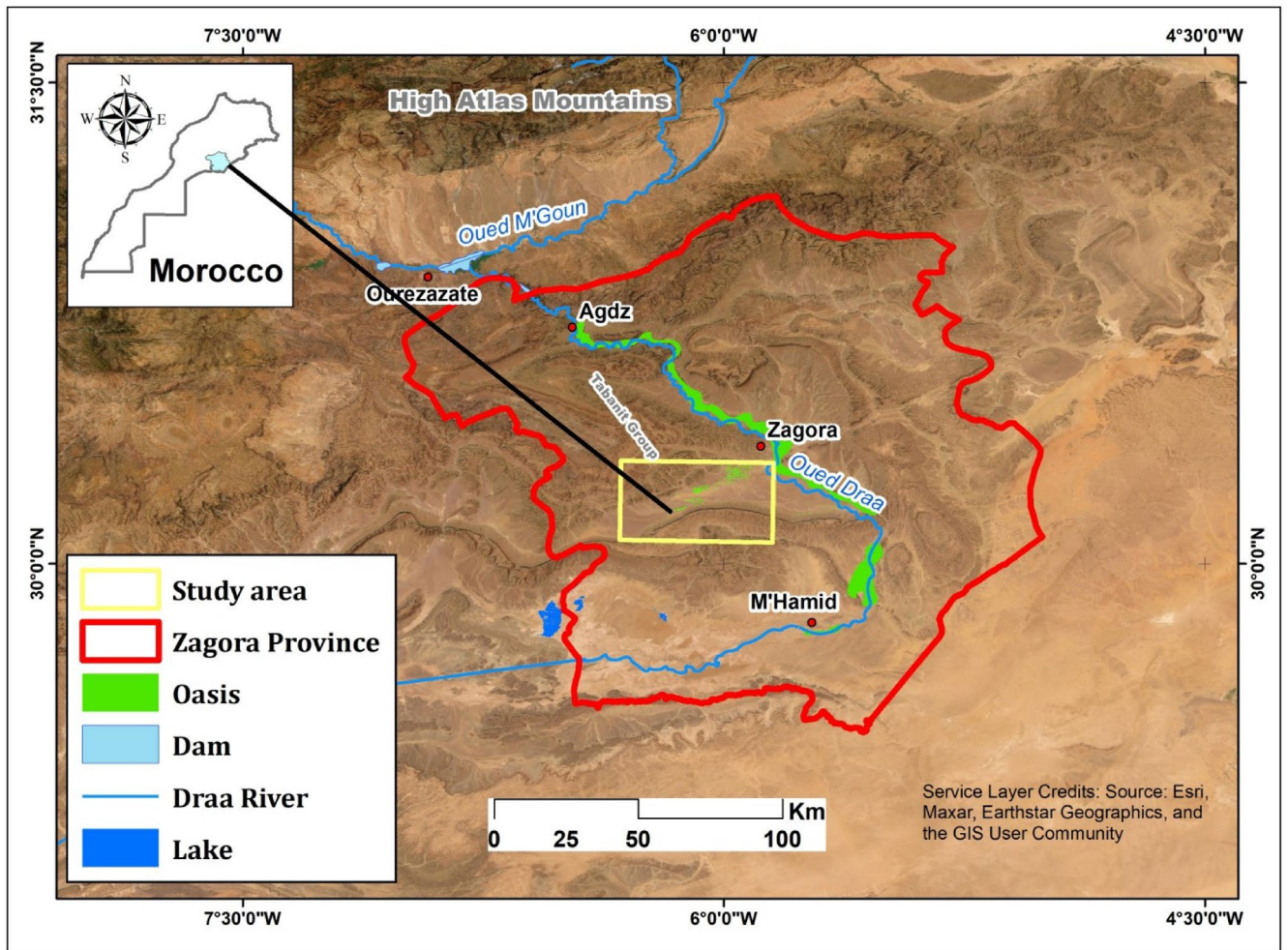


Fig. 1. Geographical context of the study site located in southeastern Morocco. The map was developed using ArcGIS Desktop 10.8 (Esri, Redlands, CA, USA; <https://www.esri.com/arcgis>). Administrative boundaries, oasis extents, and hydrographic networks were deli.

3. Annual Rainfall: Precipitation, the primary recharge driver, is integral to all groundwater potential assessments⁵¹.
4. Slope: Slope determines runoff velocity and infiltration opportunity. Flatter areas promote saturation and infiltration; steep slopes promote runoff. Slope has been repeatedly ranked high in weight by geospatial recharge assessments^{48,52}.
5. Curvature: Reflecting microtopography, convex or concave surfaces affect local water accumulation and runoff convergence. While less studied independently, curvature often complements slope and elevation in detailed terrain-based recharge mapping⁵⁰.
6. Topographic Wetness Index (TWI): TWI quantifies spatial soil moisture propensity. It's a key hydrologic indicator widely deployed in recharge modelling high values correspond to zones of soil saturation and higher recharge probability⁴⁸.
7. Geomorphology: Landforms such as alluvial plains, fluvial terraces, and mountain benches affect water retention and flow. Regions like flood plains and alluvial fans consistently exhibit enhanced recharge potential^{46,51}.
8. NDVI (Normalized Difference Vegetation Index) / Land Use–Land Cover (LULC): Vegetation influences evapotranspiration and soil moisture. Incorporating NDVI and LULC significantly improves recharge estimates in both humid and semi arid environments⁵³.
9. Stream Distance Proximity to streams and lower drainage density often enhance infiltration due to reduced runoff. Drainage metrics frequently appear in recharge zoning studies⁴⁴.
10. Lineament Density: Faults and fractures act as conduits for groundwater, facilitating vertical percolation. Areas of high lineament density correlate strongly with recharge zones in diverse terrains⁵⁴.

Data and data processing

This study is based on a dataset of 315 geo-referenced samples obtained from in-situ field surveys conducted in the Feija Plain, Morocco. The dataset encompasses essential hydrological, geomorphological, and topographical parameters that influence groundwater potential. To ensure balanced representation, the samples were evenly categorized into three groundwater potential classes: low ($n = 105$), medium ($n = 105$), and high ($n = 105$). The classification was carried out using quantitative thresholds derived from well yield/flow rates, complemented by expert field observations and ancillary indicators such as irrigation tank volumes and soil conditions.

To ensure the dataset's reliability and its suitability for machine learning applications, a rigorous preprocessing workflow was implemented. Missing values in continuous variables (e.g., rainfall and elevation) were addressed using linear interpolation, whereas categorical variables (e.g., geomorphology) were imputed using the mode. Outliers were identified based on Z-score analysis, and values exceeding predefined thresholds were either removed or corrected following contextual assessment.

Subsequently, data transformation was conducted to standardize the features and ensure compatibility with the selected algorithms. Categorical variables were encoded using one-hot encoding, while continuous variables were normalized using Min-Max scaling within a $[0, 1]$ range.

The cleaned dataset was divided into two subsets: 80% for model training and 20% for testing. This ratio was chosen to balance model learning with independent evaluation, ensuring reliable generalization to unseen data^{55,56}. To further enhance robustness, a 5-fold cross-validation procedure was applied on the training set to evaluate model stability and reduce the likelihood of overfitting. The combination of an independent test set and cross-validation provided a stable and unbiased assessment of model performance. This preprocessed dataset ultimately served as the foundation for the HydroPredictor framework, supporting robust and generalizable groundwater potential predictions.

Multicollinearity and feature importance analysis

To assess the relationships between features and mitigate redundancy, a multicollinearity analysis was performed. High correlations between predictor variables can introduce instability in model training and reduce interpretability⁵⁷. Pearson correlation analysis was conducted to identify linear relationships between features⁵⁸, and the correlation matrix is presented in Fig. 2. Features exhibiting strong correlations (above 0.75) were carefully examined to avoid redundancy.

Additionally, Mutual Information (MI) was computed to quantify the dependence between each predictor and the target variable. Unlike correlation, MI captures both linear and nonlinear associations, providing a more comprehensive ranking of feature relevance⁵⁹. The computed information gain scores are illustrated in Fig. 3. It is evident that Elevation, Soil Permeability, and Rainfall contribute the most to predictive performance.

To further address multicollinearity, the Variance Inflation Factor (VIF) and Tolerance (TOL) metrics were applied. While VIF values below 5 are generally considered safe and were preserved⁶⁰, in this study we adopted a more flexible criterion by also tolerating variables with $5 \leq \text{VIF} < 10$ to avoid discarding important environmental predictors. Variables with $\text{VIF} > 10$ were excluded due to strong collinearity. The combined visualization of VIF (bar plot) and TOL (line plot) in Fig. 4 shows that *Slope* and *Elevation* exhibit the highest VIF values (≈ 5.5 – 5.7) but remain within the acceptable range, and were therefore retained with careful interpretation during model training.

By integrating correlation analysis, mutual information, and VIF/TOL diagnostics, the most informative predictors were retained while reducing redundancy. Predictors with VIF values below 5 were considered safe and preserved, while in some cases variables with VIF values between 5 and 10 were tolerated to avoid discarding important environmental factors. Variables with $\text{VIF} > 10$ were excluded due to strong collinearity. This refined feature set was subsequently used for model training, ensuring improved accuracy and interpretability in groundwater prediction.

Machine learning models and performance evaluation

In this study, various machine learning models are employed to predict groundwater potential based on key environmental and hydrological features. The selected models include Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), CatBoost, and the proposed HydroPredictor model. These models were chosen for their ability to handle complex, non-linear relationships in geospatial and hydrological data.

The HydroPredictor model is a hybrid framework that combines CatBoost for robust feature transformation and MLP for deep learning-based classification. This integration allows the model to effectively leverage tree-based representations while benefiting from the non-linear learning capability of neural networks. Furthermore, a meta-classifier is employed to combine probabilistic outputs, enhancing overall predictive performance and robustness, particularly for small datasets^{61,62}.

To ensure robust evaluation, the models were assessed using multiple performance metrics, including accuracy, F1-score, recall, Cohen's kappa, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (ROC-AUC). Additionally, a 5-fold cross-validation strategy was applied to mitigate overfitting and enhance the generalization capability of the models. This section details the machine learning models utilized and the performance evaluation framework adopted to compare their predictive capabilities.

Machine learning models overview

Support vector machine (SVM)

The SVM were used as a baseline model to evaluate groundwater potential mapping under data-scarce conditions. The SVM leverages kernel functions to capture nonlinear relationships between environmental predictors and groundwater classes, making it suitable for complex hydrological systems⁶³. Key hyperparameters, including the

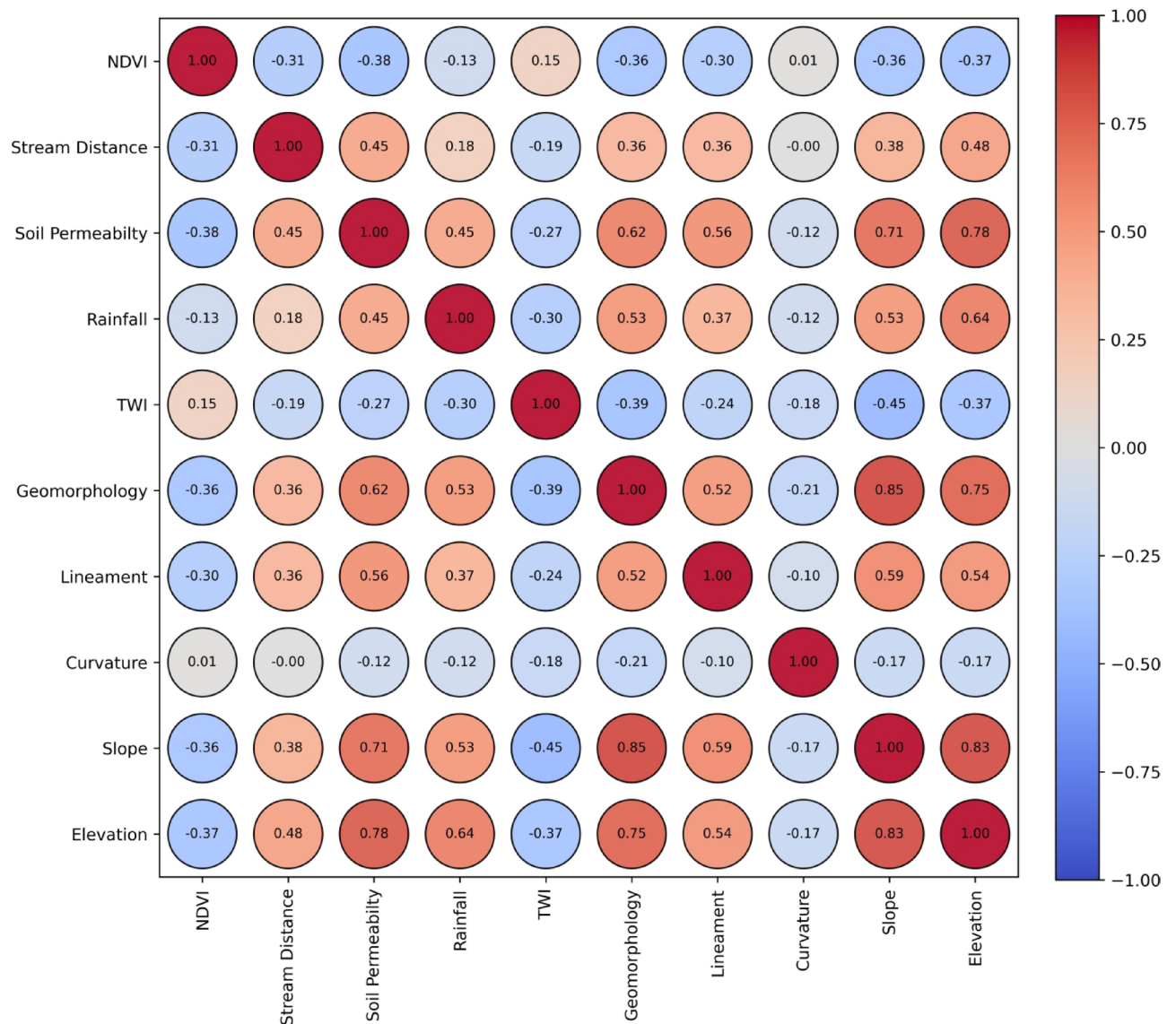


Fig. 2. Correlation matrix of input features for groundwater prediction.

regularization parameter C , kernel type, and gamma, were tuned using Optuna with 5-fold cross-validation⁶⁴. This ensured the model achieved an optimal balance between bias and variance while reducing overfitting risks.

For regression applications, the functional representation of the SVM model can be expressed as Eq. (1):

$$f_{\text{SVM}}(x) = \omega \cdot \varphi(x) + b \quad (1)$$

where ω denotes the weight vector, b represents the bias term, and $\varphi(x)$ is a non-linear mapping function applied to the input x via the selected kernel function. This formulation allows SVM to efficiently approximate complex relationships between input variables and target outputs while maintaining robustness against overfitting.

Random forest (RF)

Random Forest model was employed as a benchmark ensemble method due to its robustness in handling nonlinear relationships and high-dimensional environmental predictors. The model was trained using bootstrap sampling with random feature selection at each node, which helps reduce overfitting and improve generalization. Beyond classification, RF also provided estimates of feature importance, offering valuable insights into the relative contribution of environmental variables to groundwater potential prediction. Hyperparameters such as the number of trees, maximum depth, and minimum samples per split were optimized using Optuna with 5-fold cross-validation to ensure stable and reliable performance. This can be calculated as Eq. (2)

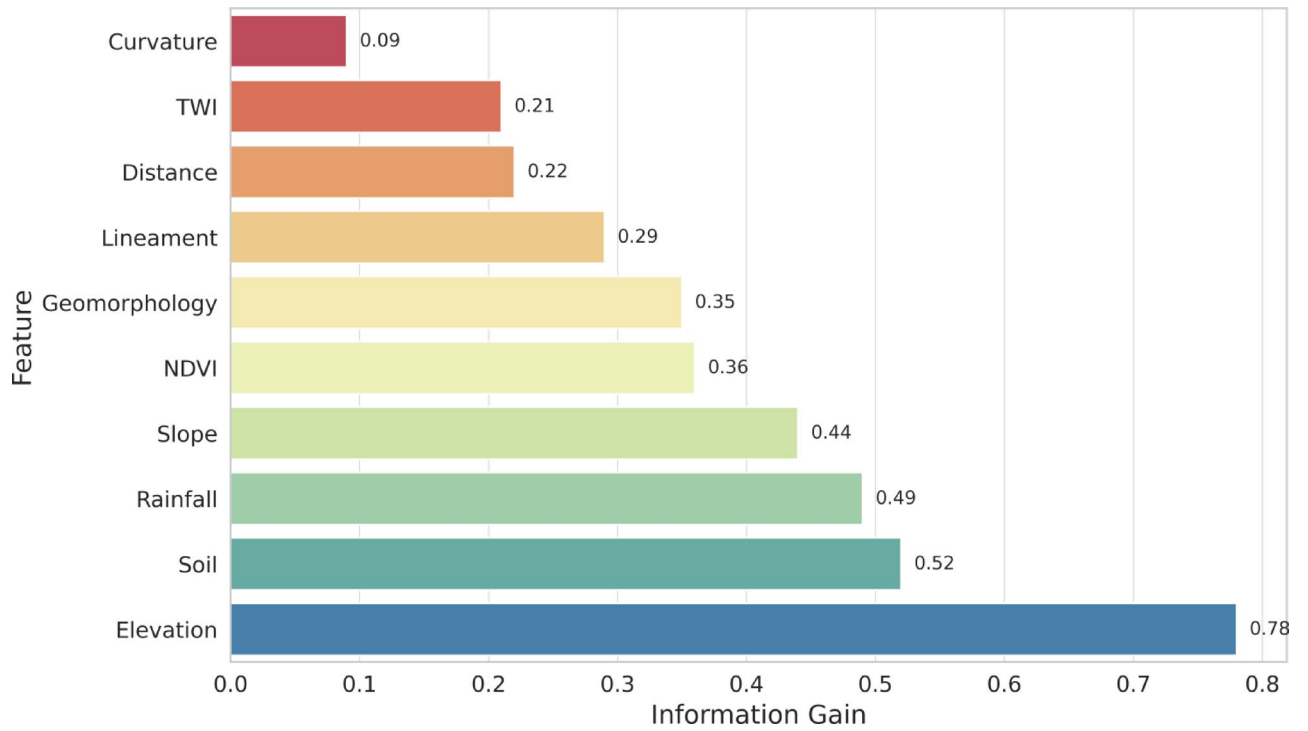


Fig. 3. Mutual Information ranking of features based on information gain.

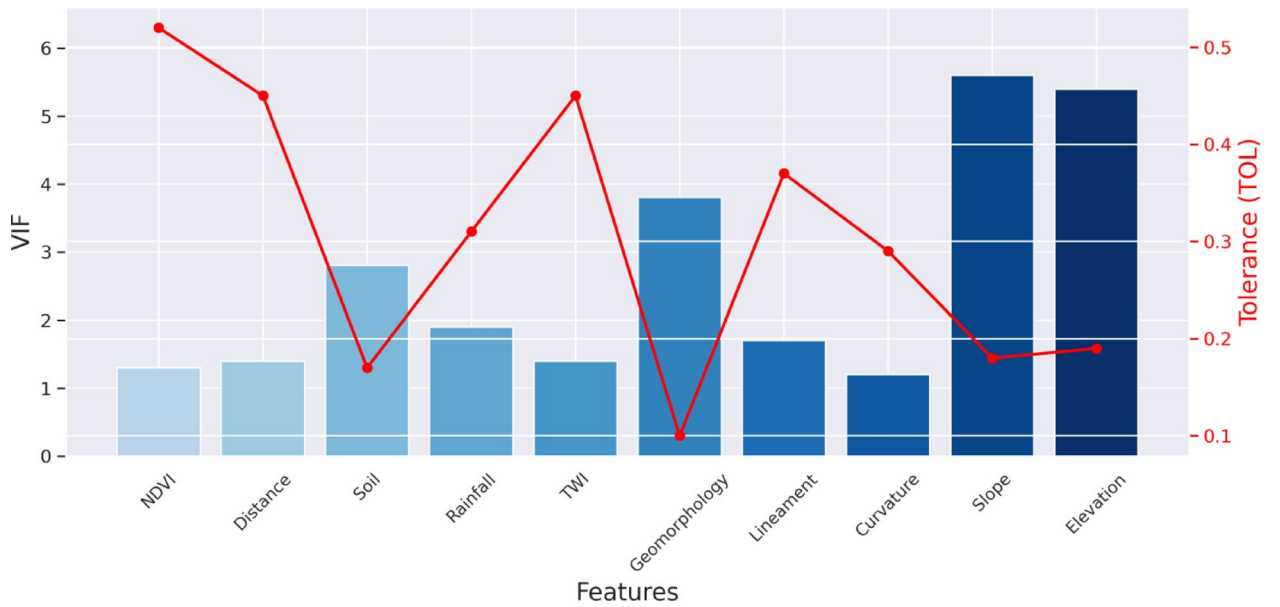


Fig. 4. Variance Inflation Factor (VIF) and Tolerance (TOL) analysis for feature selection.

$$FRF(x) = \frac{1}{T} \sum_{i=1}^T f_i(x) \tag{2}$$

where T is the total number of decision trees in the forest, and $f_i(x)$ represents the individual prediction from the i -th tree.

This aggregation strategy enhances generalization and ensures robust performance across diverse datasets.

CatBoost

CatBoost is a highly efficient gradient boosting algorithm designed for categorical data that requires minimal preprocessing. By employing **ordered boosting**, it prevents target leakage and mitigates overfitting⁶⁵. The model operates by minimizing a loss function, which can be written as Eq. (3):

$$L = \sum_{i=1}^n \ell(y_i, \hat{y}_i) \quad (3)$$

where ℓ is the loss function, y_i is the true target value, and \hat{y}_i is the predicted value (Fig. 5). Gradient boosting updates the predictions iteratively by minimizing this loss using the update rule as Eq. (4):

$$\hat{y}_i^{(t+1)} = \hat{y}_i^{(t)} + \eta \cdot \nabla_{\hat{y}_i^{(t)}} L \quad (4)$$

where η is the learning rate, and $\nabla_{\hat{y}_i^{(t)}} L$ represents the gradient of the loss function with respect to the current predictions.

CatBoost was selected as a baseline boosting method due to its efficiency in handling categorical and heterogeneous environmental variables without requiring extensive preprocessing (see Fig. 5). By applying ordered boosting and target statistics, CatBoost reduces gradient bias and prediction shift, which are common issues in small datasets. Hyperparameters such as learning rate, depth, and regularization parameters were optimized using Optuna with 5-fold cross-validation, ensuring robust classification of groundwater potential classes and minimizing overfitting.

Multi-layer perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a type of artificial neural network composed of three fundamental components: an input layer, one or more hidden layers, and an output layer⁶⁶. Each neuron in the input layer corresponds to a specific feature from the dataset, ensuring that relevant information is transmitted into the network Fig. 6. The hidden layers serve as the computational core of the model, where nonlinear transformations are applied to extract complex patterns from the data. The final predictions are produced in the output layer, making MLP highly effective for classification tasks, including medical diagnostics such as breast cancer detection⁶⁷.

The computational power of MLP stems from its capacity to incorporate multiple hidden layers, enabling it to capture intricate feature interactions. The transformation at each layer is mathematically defined as Eq. (5):

$$H_l = \sigma(W_l H_{l-1} + b_l) \quad (5)$$

where:

- H_l represents the activations at layer l ,
- W_l denotes the weight matrix for layer l ,

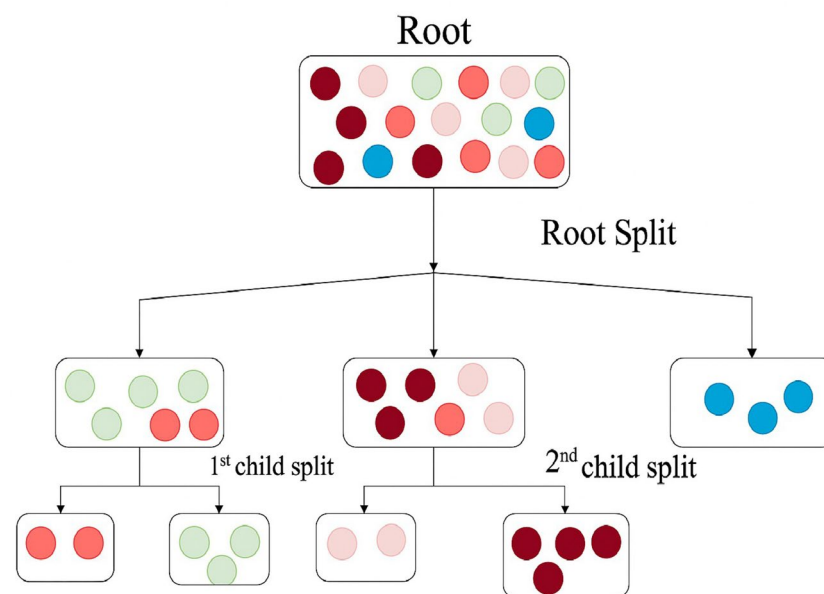


Fig. 5. Visualizing the CatBoost Decision Process.

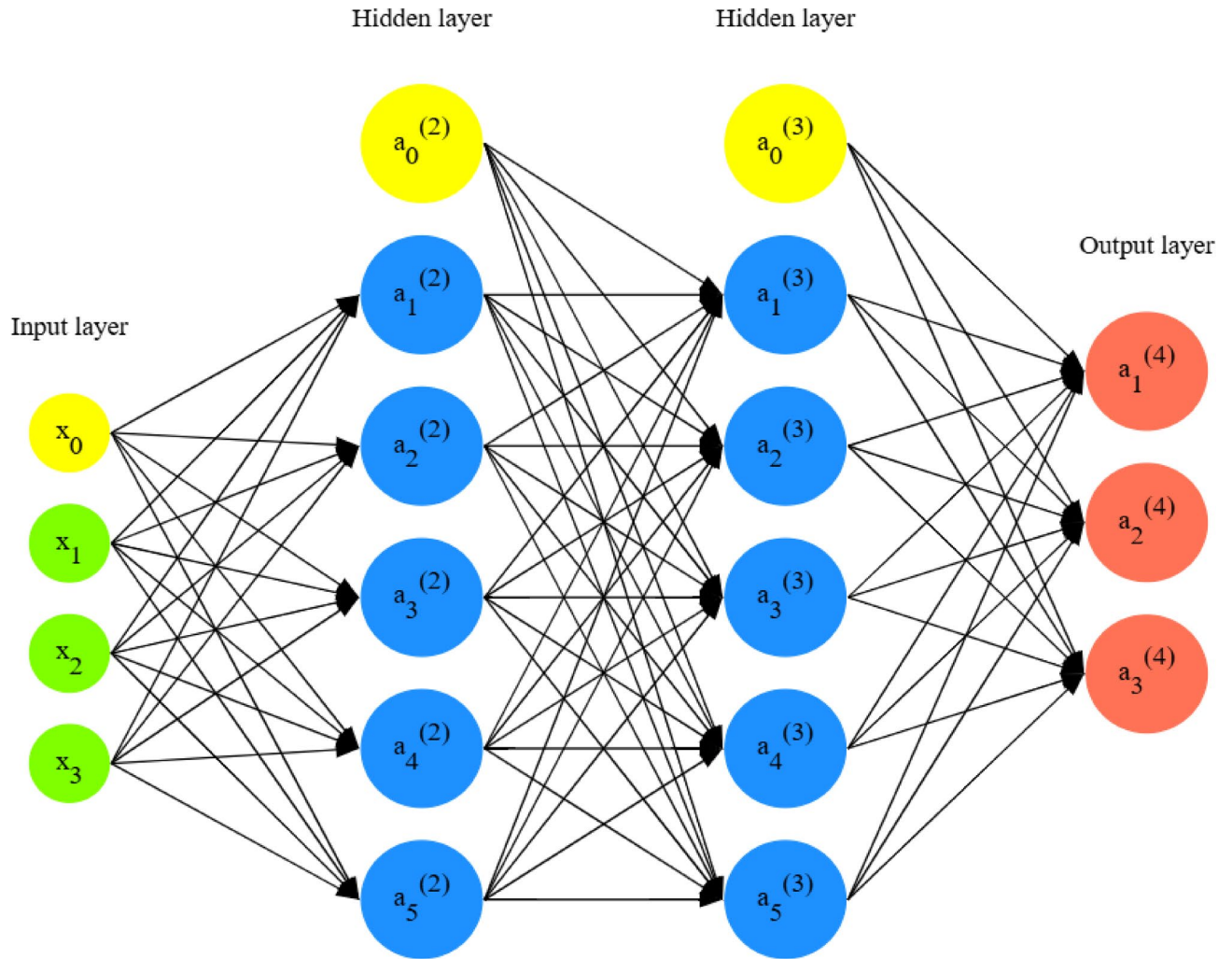


Fig. 6. Architecture of a Multi-Layer Perceptron (MLP) Neural Network with Input, Hidden, and Output Layers.

- b_l is the bias term, and,
- $\sigma(\cdot)$ is the activation function, typically ReLU or sigmoid.

The final output of the network is computed as Eq. (6):

$$\hat{y} = \text{softmax}(W_o H_L + b_o) \quad (6)$$

where H_L is the activation of the last hidden layer, W_o and b_o are the weights and biases of the output layer, respectively. The softmax function ensures that the network outputs probability distributions for classification tasks. This flexibility in architecture allows MLPs to model complex, high-dimensional data, making them a crucial tool in various domains, including medical and environmental sciences.

Proposed hybrid model: hydropredictor architecture and formulation

In this study, we propose HydroPredictor, a novel hybrid framework designed to enhance predictive accuracy and robustness for groundwater classification. The model integrates tree-based feature extraction (CatBoost), deep learning representation learning (MLP), and ensemble decision fusion strategies to optimize generalization performance while minimizing overfitting. This section details the mathematical formulation of each component, illustrating how HydroPredictor leverages structured feature transformations, non-linear embeddings, and adaptive fusion mechanisms to deliver state-of-the-art classification performance.

CatBoost component: feature transformation and probability Estimation

The first stage of HydroPredictor employs CatBoost, a gradient boosting decision tree algorithm, to both transform raw input features X into a structured latent space and provide initial probability estimates.

Probability Prediction. For an input vector X , the CatBoost model outputs a probability vector over C classes as calculated using Eq. (7):

$$P_{CB}(X) = [P(y = 0 | X)_{CB}, P(y = 1 | X)_{CB}, \dots, P(y = C - 1 | X)_{CB}] \quad (7)$$

Leaf Embedding Transformation. CatBoost maps X into a latent feature space using its ensemble of trees. For an input x , let $l_t(x)$ denote the leaf index for the t th tree. Then, the leaf embeddings are defined as Eq. (8):

$$\mathbf{F}_{CB}(x) = [l_1(x), l_2(x), \dots, l_T(x)] \in \mathbb{R}^T \quad (8)$$

where T is the total number of trees in the CatBoost model.

MLP component: deep learning on Tree-Based embeddings

The second stage utilizes a Multi-Layer Perceptron (MLP) to learn non-linear representations from the CatBoost-derived embeddings $\mathbf{F}_{CB}(x)$. Suppose the first hidden layer has size d . Then, we have the following formulations.

Hidden Layers. The first hidden layer is computed as Eq. (9):

$$\mathbf{H}_1 = \sigma(W_1 \mathbf{F}_{CB}(x) + b_1) \quad (9)$$

where $W_1 \in \mathbb{R}^{d \times T}$ and $b_1 \in \mathbb{R}^d$. Here, $\sigma(\cdot)$ denotes the ReLU activation function, defined as Eq. (10). Also, the size of hidden layer is computed as Eq. (11)

$$\sigma(z) = \max(0, z) \quad (10)$$

$$\mathbf{H}_2 = \sigma(W_2 \mathbf{H}_1 + b_2) \quad (11)$$

where $W_2 \in \mathbb{R}^{h \times d}$ and $b_2 \in \mathbb{R}^h$, with h denoting the size of the second hidden layer.

Output Layer. The output layer applies a softmax function to yield class probability estimates as Eq. (12):

$$P_{MLP}(x) = \text{softmax}(W_o \mathbf{H}_2 + b_o) \quad (12)$$

with $W_o \in \mathbb{R}^{C \times h}$ and $b_o \in \mathbb{R}^C$. The softmax function is defined as Eq. (13):

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (13)$$

Decision fusion: integrating catboost and MLP outputs

HydroPredictor incorporates two decision fusion strategies: (A) Weighted Probability Fusion and (B) Meta-Classified Fusion. Note that if using LogisticRegression from scikit-learn, one may specify `multi_class='multinomial'` to employ the softmax-based multinomial logistic regression for multi-class classification. Otherwise, the default one-vs-rest (OvR) strategy is used. Our formulation here assumes the multinomial configuration for clarity.

A. Weighted Probability Fusion.

A tunable parameter $\alpha \in [0, 1]$ is used to blend the probability estimates from the CatBoost and MLP components using Eq. (14):

$$P_{\text{final}}(x) = \alpha P_{MLP}(x) + (1 - \alpha) P_{CB}(x) \quad (14)$$

The final predicted class is obtained by Eq. (15):

$$\hat{y} = \underset{k}{\text{argmax}} [P_{\text{final}}(x)]_k \quad (15)$$

B. Meta-Classified Fusion.

Alternatively, we concatenate the CatBoost and MLP probability outputs to form a combined feature vector using Eq. (16):

$$P_{\text{combined}}(x) = [P_{CB}(x), P_{MLP}(x)] \in \mathbb{R}^{2C} \quad (16)$$

For multi-class classification, the meta-classifier employs a multinomial logistic regression (i.e., softmax formulation). Specifically, the probability for class k is computed using Eq. (17):

$$P_{\text{meta}}(y = k | x) = \frac{\exp(\mathbf{w}_k^T P_{\text{combined}}(x) + b_k)}{\sum_{j=1}^C \exp(\mathbf{w}_j^T P_{\text{combined}}(x) + b_j)}, \quad (17)$$

where $\mathbf{w}_k \in \mathbb{R}^{2C}$ and $b_k \in \mathbb{R}$ are the weight vector and bias for class k , respectively. The final prediction is then calculated using Eq. (18):

$$\hat{y} = \operatorname{argmax}_k P_{\text{meta}}(y = k | x) \tag{18}$$

Overall model summary

Depending on the fusion strategy employed, the overall prediction rule for HydroPredictor is as follows:

For Weighted Fusion, the final prediction is then calculated using Eq. (19):

$$\hat{y} = \operatorname{argmax}_k (\alpha P_{\text{MLP}}(x)_k + (1 - \alpha) P_{\text{CB}}(x)_k). \tag{19}$$

For Meta-Classified Fusion, the final prediction is then calculated using Eq. (20):

$$\hat{y} = \operatorname{argmax}_k \left(\frac{\exp(\mathbf{w}_k^T P_{\text{combined}}(x) + b_k)}{\sum_{j=1}^C \exp(\mathbf{w}_j^T P_{\text{combined}}(x) + b_j)} \right). \tag{20}$$

This formulation captures the core innovation of HydroPredictor: the integration of tree-based feature extraction and deep learning with adaptive ensemble strategies, yielding robust and interpretable predictions in multi-class settings.

This Fig. 7 the training workflow of the proposed Hybrid HydroPredictor model, which integrates CatBoost and Multi-Layer Perceptron (MLP) for groundwater prediction. The process begins with data preprocessing, feature engineering, and selection, followed by splitting the dataset into training and testing sets. The CatBoost model is first trained to extract leaf embeddings, which serve as input features for the MLP model. Both models generate probability estimates, which are then combined using weighted fusion before being processed by a meta-classifier. The final trained model is optimized for robust and accurate groundwater prediction, ensuring enhanced generalization and interpretability.

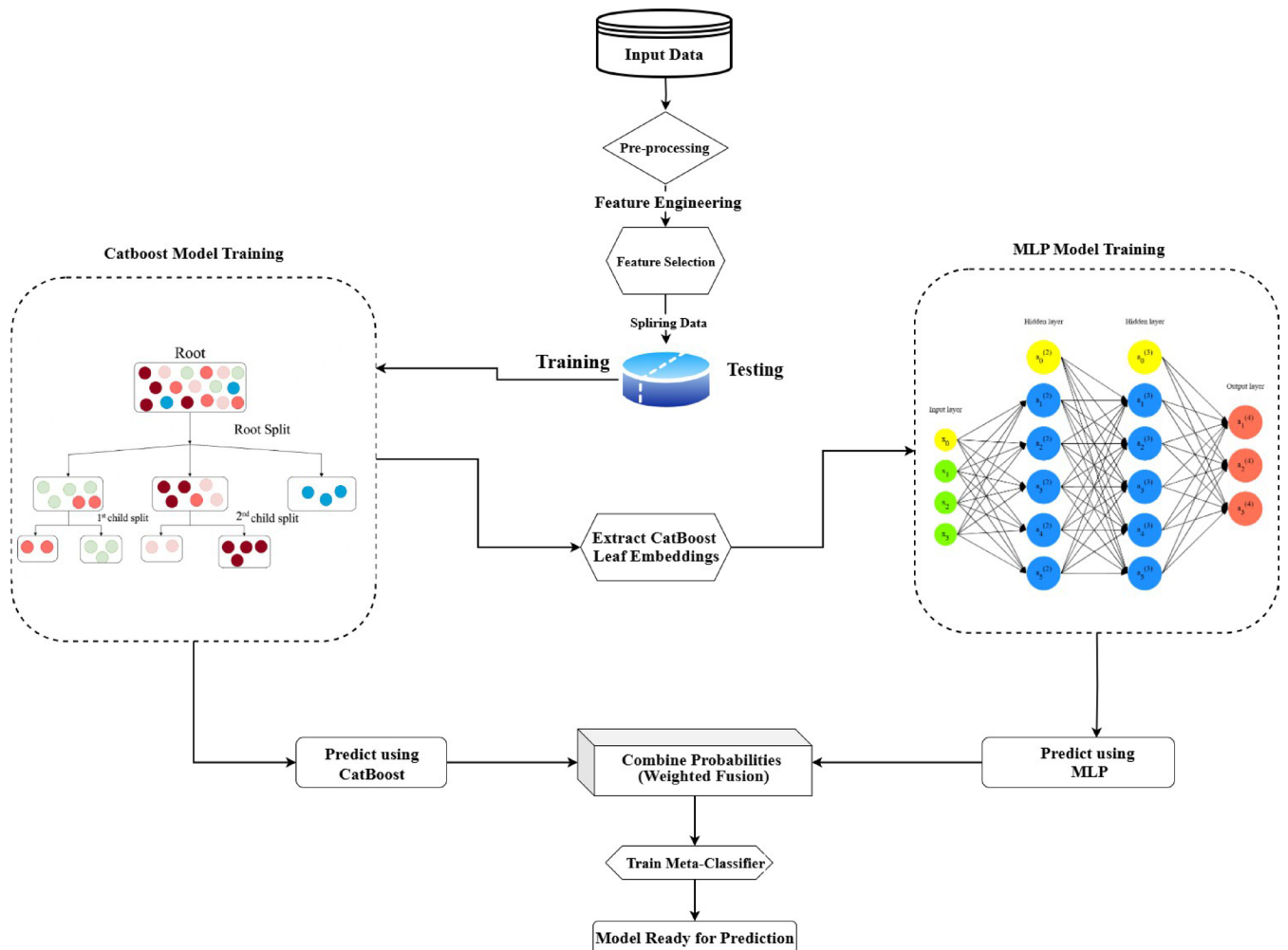


Fig. 7. Workflow of the Hybrid HydroPredictor Model for Groundwater Prediction.

Hyperparameter optimization using optuna

To enhance model performance and ensure fair comparisons, hyperparameter tuning was conducted using Optuna, an advanced Bayesian optimization framework. Optuna efficiently explored the hyperparameter search space using a Tree-structured Parzen Estimator (TPE) approach, selecting configurations that maximized F1-score and accuracy.

Each model underwent 50 trials of hyperparameter tuning, where key hyperparameters were optimized using a 5-fold cross-validation strategy. The objective was to achieve a balance between model complexity and generalization, minimizing overfitting while maintaining high predictive performance.

The range of hyperparameters explored for each model and the best-selected values are summarized in Table 1.

Impact of hyperparameter tuning

The application of Optuna-based hyperparameter optimization resulted in significant improvements across all models. In particular, HydroPredictor demonstrated superior generalization ability, achieving a balance between sensitivity and specificity. Table 1 shows the hyperparameter list with their optimised score in detail.

By systematically adjusting hyperparameters, Optuna enabled:

- Reduction of overfitting, as evidenced by the minimal performance gap between training and testing.
- Enhanced classification stability, particularly in challenging classes in the confusion matrix.
- Improved predictive reliability across all evaluation metrics.

The optimized models outperformed their default configurations, further validating the necessity of hyperparameter tuning in achieving high classification accuracy.

Validation procedure using randomly sampled ground truth points

To ensure the reliability and spatial accuracy of the groundwater recharge suitability map generated by the HydroPredictor model, a field-informed validation approach based on random sampling was employed. A total of 90 georeferenced validation points were evenly distributed across the Feija Basin to capture a representative range of geomorphological and hydrogeological conditions. These points were independently classified into three recharge potential categories High ($n = 30$), Medium ($n = 30$), and Low ($n = 30$) using a combination of direct field observations, local expert judgment, and multiple ancillary indicators. The supporting evidence included measured well yield (flow rate), irrigation tank water levels, and qualitative assessments of land surface conditions, such as vegetation health, soil moisture, and signs of degradation.

Rather than assigning fixed numerical weights to these indicators, a contextual and consensus-based classification protocol was adopted. Each point was categorized based on the convergence of multiple cues for example, a location with high well yield, full tanks, and healthy vegetation was labeled “High”, while one with dry wells, depleted tanks, and degraded land was assigned to the “Low” class. Intermediate or mixed indicators informed the “Medium” classification.

Each validation point was then spatially overlaid on the predicted recharge suitability map to assess the agreement between observed and modeled categories. From this comparison, a confusion matrix was generated, and key statistical metrics including overall accuracy, Cohen’s Kappa coefficient, and class-wise precision and recall—were calculated. This statistically rigorous and field-grounded validation process ensures that the model’s performance reflects real-world recharge conditions and enhances the credibility of its outputs for groundwater resource planning in arid, data-limited settings like the Feija Basin (Fig. 8).

Results

Model performance evaluation

The experimental results demonstrate that the proposed HydroPredictor model significantly outperforms conventional machine learning algorithms across all key evaluation metrics. As presented in Table 2, HydroPredictor achieved the highest testing accuracy of 89.23%, representing a marked improvement over

Model	Hyperparameters Tuned	Best-Selected Hyperparameters
CatBoost	Iterations (50–200), Learning rate (0.01–0.3), Depth (3–10), L2 reg (1–10), Bagging temperature (0.01–10), Border count (5–255), Min data in leaf (1–100)	Iterations: 70, Learning rate: 0.055, Depth: 8, L2 reg: 4, Bagging temperature: 0.0115, Border count: 188
SVM	C (0.1–10), Kernel (Linear, Radial Basis Function (RBF), Poly, Sigmoid), Degree (2–6), Gamma (Scale, Auto), Class weight (Balanced, None)	C: 3.79, Kernel: RBF, Degree: 3, Gamma: Auto, Class weight: Balanced
MLP	Learning rate (0.001–0.03), Epochs (20–30), Batch size (16, 32, 64), Optimizer (Adam, Stochastic Gradient Descent (SGD)), Activation (ReLU, Tanh), Dropout rate (0.2–0.5), L1 reg (1e–5–1e–2), L2 reg (1e–5–1e–2), Early stopping (True, False), Patience (5–15), Batch norm (True, False)	Learning rate: 0.018, Epochs: 29, Batch size: 32, Optimizer: Adam, Activation: ReLU, Dropout rate: 0.344, L1 reg: 0.0028, L2 reg: 0.000058, Early stopping: True, Patience: 8, Batch norm: True
Random Forest	n_estimators (50–100), max_depth (None, 5, 7, 9, 10), min_samples_split (2–10), min_samples_leaf (1–8), max_features (sqrt, log2, None)	n_estimators: 95, max depth: None, min samples split: 4, min_samples leaf: 1, max features: sqrt
HydroPredictor	Learning rate (0.01–0.03), Depth (3–12), Top features (3–15), Epochs (10–30), Model weight (0.01–0.03), L2 reg (5–10), Iterations (50–200)	Learning rate: 0.029, Depth: 10, Top features: 10, Epochs: 20, Model weight: 0.2669, L2 reg: 5, Iterations: 163

Table 1. Hyperparameter search space and Best-Selected parameters using Optuna.

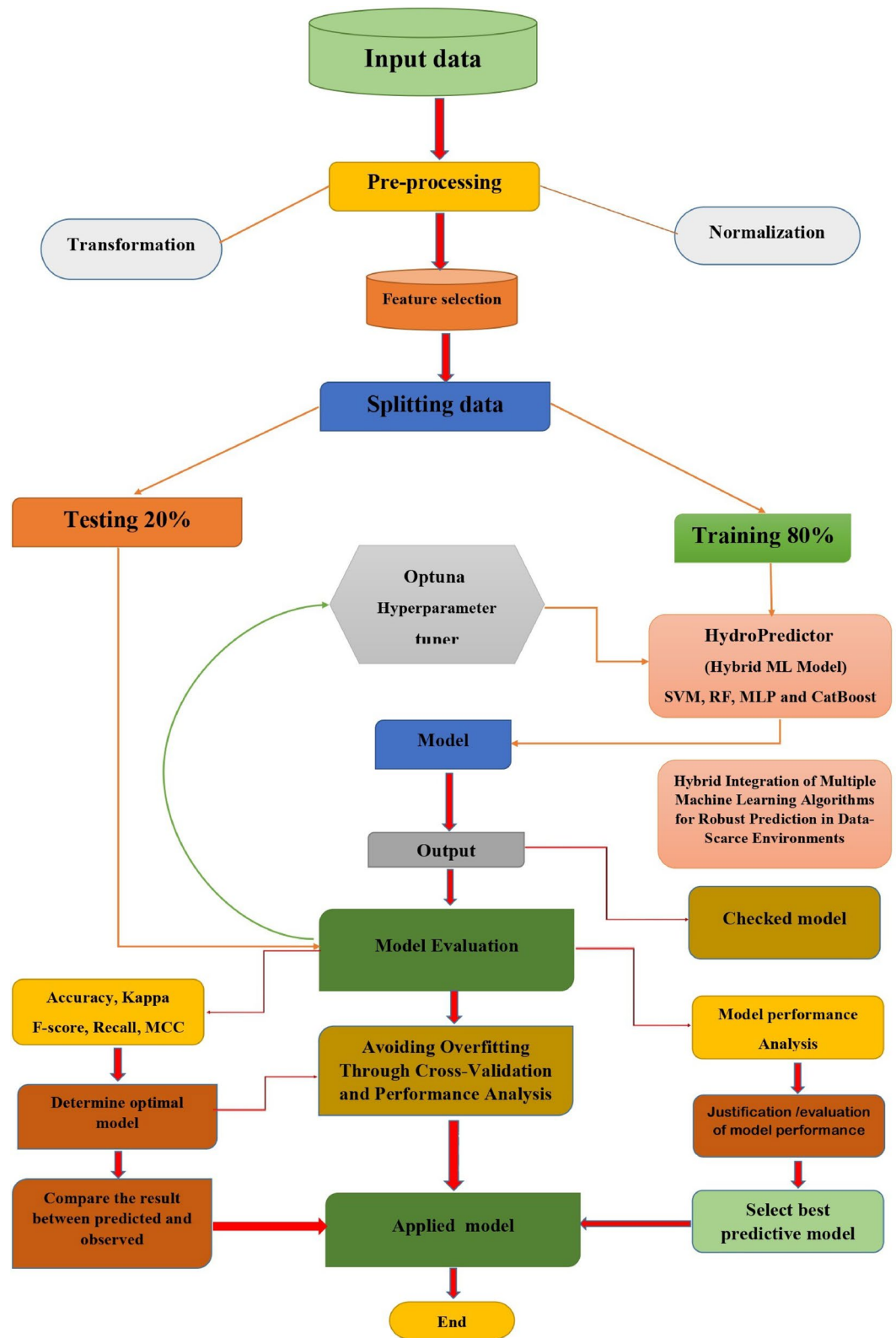


Fig. 8. A Structured Methodology for Groundwater Potential Mapping Using Machine Learning Techniques.

CatBoost and Random Forest (both 84.62%), and Support Vector Machine (SVM), which yielded a considerably lower accuracy of 75.38%. These results underscore the predictive advantage conferred by the model's hybrid ensemble structure, particularly in handling the complexity of groundwater classification in data-scarce regions.

Model	Training Performance					Testing Performance				
	Accuracy	F1-score	Kappa	Recall	MCC	Accuracy	F1-score	Kappa	Recall	MCC
CatBoost	0.8919	0.8914	0.8378	0.8919	0.8421	0.8462	0.8481	0.7692	0.8462	0.7703
RF	0.8764	0.8757	0.8146	0.8764	0.8199	0.8462	0.8479	0.7692	0.8462	0.7716
SVM	0.8185	0.8158	0.7276	0.8185	0.7315	0.7538	0.7578	0.6307	0.7538	0.6316
MLP	0.7992	0.8004	0.6987	0.7992	0.7041	0.7846	0.7893	0.6770	0.7846	0.6786
HydroPredictor	0.8958	0.8956	0.8436	0.8958	0.8445	0.8923	0.8937	0.8385	0.8923	0.8394

Table 2. Performance comparison of models on training and testing datasets.

HydroPredictor also demonstrated remarkable generalization ability, as indicated by its minimal train-test performance gap ($\Delta\text{Accuracy}=0.35\%$). This contrasts sharply with SVM ($\Delta\text{Accuracy}=6.47\%$) and MLP ($\Delta\text{Accuracy}=1.46\%$), both of which exhibited signs of overfitting. The small generalization gap in HydroPredictor can be attributed to the combination of optimal feature selection, model regularization, and robust cross-validation, which collectively minimize variance and enhance stability on unseen data.

In terms of class-wise balance, HydroPredictor achieved the highest F1-score (0.8937) and recall (0.8923) on the testing set. These metrics indicate that the model performs consistently across all classes and is especially effective in reducing both false positives and false negatives a critical aspect in environmental decision-support systems where misclassification may result in resource mismanagement.

Moreover, agreement-based statistics further validate the model's reliability. HydroPredictor attained the highest Matthews Correlation Coefficient (MCC=0.8394) and Cohen's Kappa (Kappa=0.8385), reflecting strong concordance between predicted and observed groundwater potential classes. These metrics are particularly robust in multi-class and imbalanced datasets, as they incorporate all four confusion matrix elements. In contrast, CatBoost and RF achieved moderately lower MCC scores (0.7703 and 0.7716, respectively), while SVM and MLP reported relatively weak agreement levels.

Collectively, these findings affirm the robustness, accuracy, and generalizability of the HydroPredictor framework. Its consistent superiority across accuracy, recall, and agreement metrics highlights its suitability for operational groundwater potential assessment and supports its use as a decision-support tool in data-constrained hydrological environments.

Model discrimination analysis using ROC curves and AUC scores

To comprehensively evaluate the classification performance of the models, Receiver Operating Characteristic (ROC) curves and the corresponding AUC values were analyzed across the three groundwater potential classes. These metrics are particularly valuable in multi-class settings, as they provide a nuanced view of the model's discriminatory power for each class.

Figs. 9, 10, 11 and 12, and 13 illustrate the ROC curves for HydroPredictor and the baseline models, respectively. The proposed HydroPredictor model exhibited consistently strong AUC values across all classes. On the testing dataset, it achieved an AUC of 0.99 for Class 1, 0.91 for Class 2, and 0.96 for Class 3. These results demonstrate the model's robust ability to separate groundwater potential classes, even under conditions of class imbalance or feature overlap.

In comparison, ensemble models such as CatBoost and Random Forest also performed well, achieving testing AUC values ranging from 0.94 to 0.98. However, they showed slightly reduced sensitivity for Class 2, which tends to be more challenging to classify due to its intermediate characteristics. MLP and SVM recorded lower AUC values, particularly for Class 2 (0.87 and 0.82, respectively), indicating less reliable class discrimination.

Confusion matrix evaluation

To further investigate model performance and error distribution, confusion matrices were generated for all models, as shown in Figs. 14 and 15. These matrices provide insights into class-wise accuracy, particularly the tendency of each model to misclassify between adjacent groundwater potential categories.

HydroPredictor achieved the most balanced and accurate predictions across all classes. Specifically, it correctly classified 20 out of 22 instances in Class 1, 16 out of 21 in Class 2, and 19 out of 22 in Class 3, reflecting high predictive precision and recall. Only minimal misclassifications were observed, and these occurred primarily between Classes 2 and 3 which often share overlapping hydrological characteristics.

By contrast, SVM showed substantial misclassification, with 8 incorrect predictions in Class 2 and 6 in Class 3. This is consistent with its lower MCC and AUC values reported earlier. While CatBoost and Random Forest models demonstrated moderate performance, they still exhibited confusion between moderate and high-potential zones. MLP showed slightly better class-wise consistency than SVM but remained inferior to HydroPredictor.

The integrated results from the ROC-AUC analysis and confusion matrix evaluation offer compelling and multi-dimensional evidence of HydroPredictor's superior performance. Beyond achieving the highest overall accuracy, the model demonstrated exceptional consistency in correctly classifying all three groundwater potential classes. Its most notable strength lies in its ability to accurately identify Class 2 the intermediate category which is often the most challenging to distinguish due to its overlapping characteristics with both low and high potential zones. This indicates that HydroPredictor is not only sensitive to class boundaries but also capable of capturing nuanced patterns that conventional models tend to overlook.

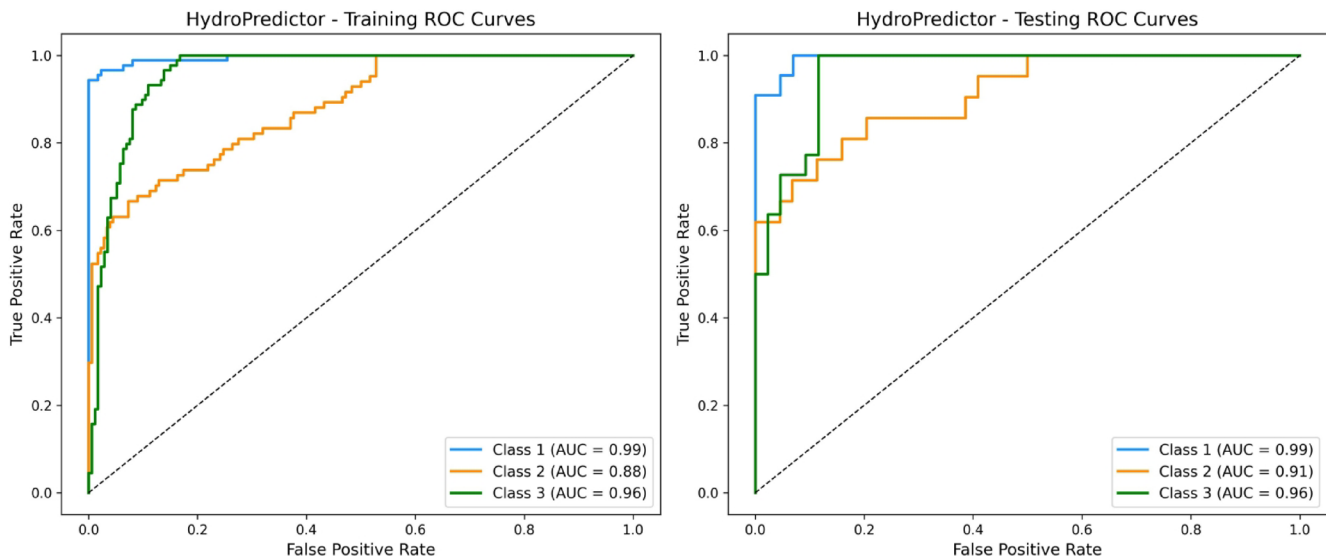


Fig. 9. ROC curves for HydroPredictor during training and testing across the three groundwater potential classes. The model demonstrates strong discriminative performance with AUC values close to 1.0 for all classes.

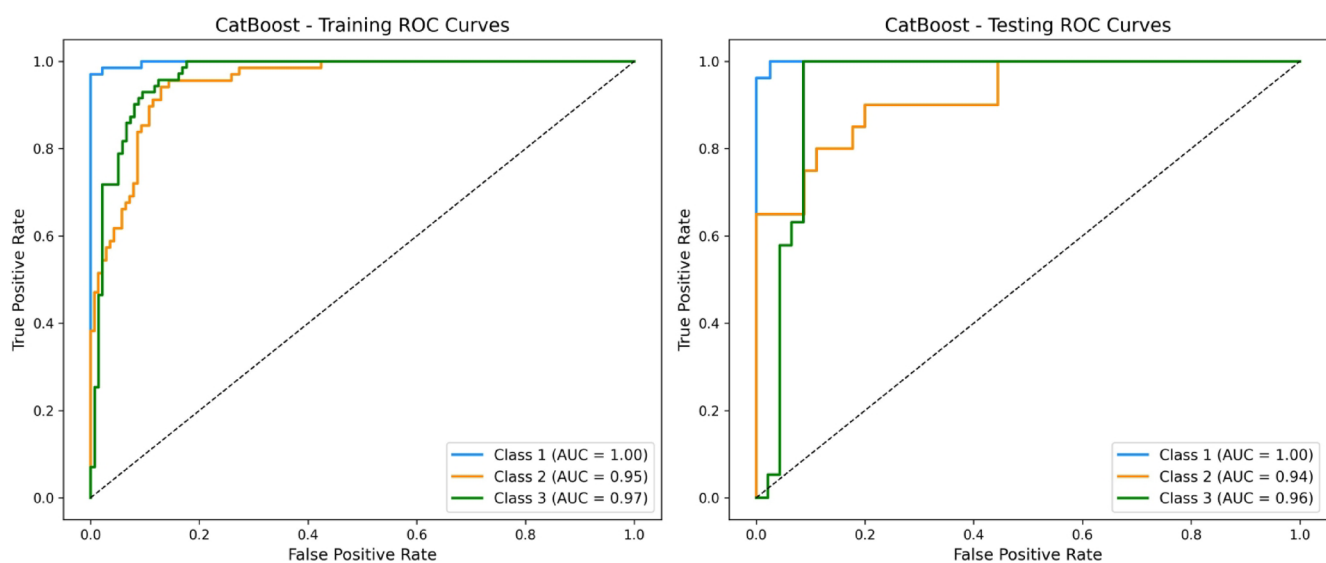


Fig. 10. ROC curves for CatBoost model during training and testing. AUC values indicate high classification capability, especially for Class 1 and Class 3.

In addition to classification accuracy, HydroPredictor exhibited strong agreement metrics, including high values of Cohen's Kappa and Matthews Correlation Coefficient, underscoring the reliability and reproducibility of its predictions. The low rate of misclassification across classes reflects the model's robust learning capacity and its ability to generalize effectively to unseen data. Collectively, these findings validate HydroPredictor as a reliable and interpretable decision-support system for groundwater potential assessment, particularly suited for application in data-scarce and environmentally sensitive regions.

Statistical validation and comparative ranking of model performance

To rigorously assess whether observed differences in model performance were statistically significant, both the Friedman test and the Wilcoxon signed-rank post-hoc tests were employed using F1-score distributions derived from five-fold cross-validation.

Friedman Test: The Friedman test revealed significant differences among the evaluated models, with a chi-squared statistic of $\chi^2 = 12.00$ and an associated p -value of $p = 0.0174$. This confirms that not all models perform equivalently, thereby justifying the use of post-hoc pairwise comparisons.

Wilcoxon Signed-Rank Test: Table 3 presents the pairwise Wilcoxon signed-rank test p -values. Although no pairwise comparison yielded statistical significance at the conventional $\alpha = 0.05$ level, values approaching

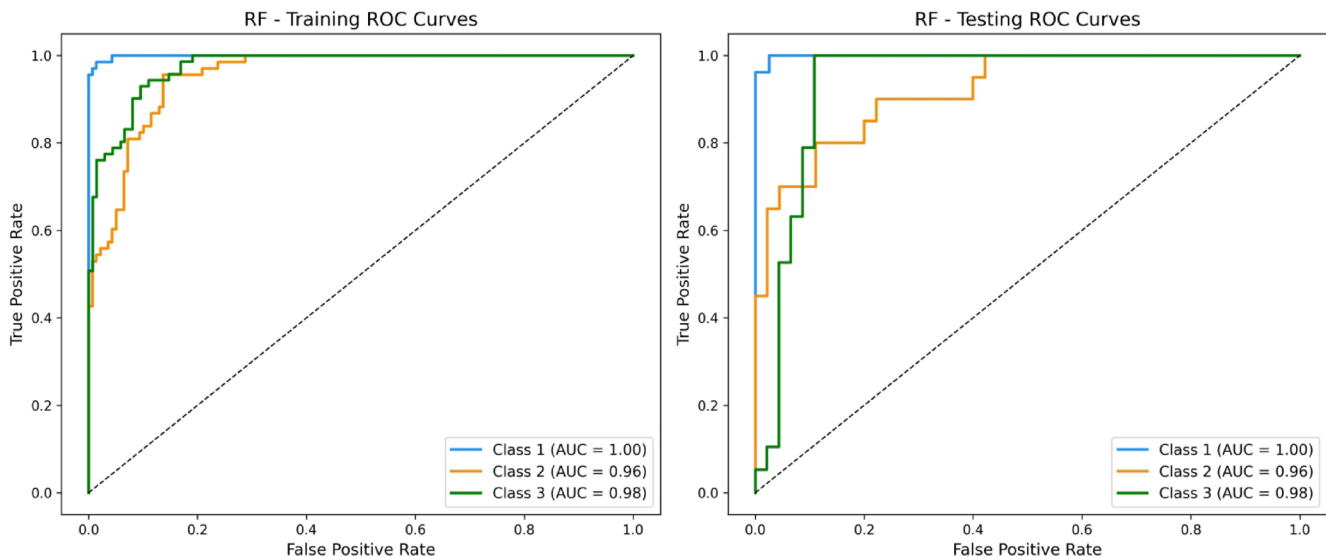


Fig. 11. ROC curves for Random Forest model showing consistently strong AUC values across all classes in both training and testing phases.

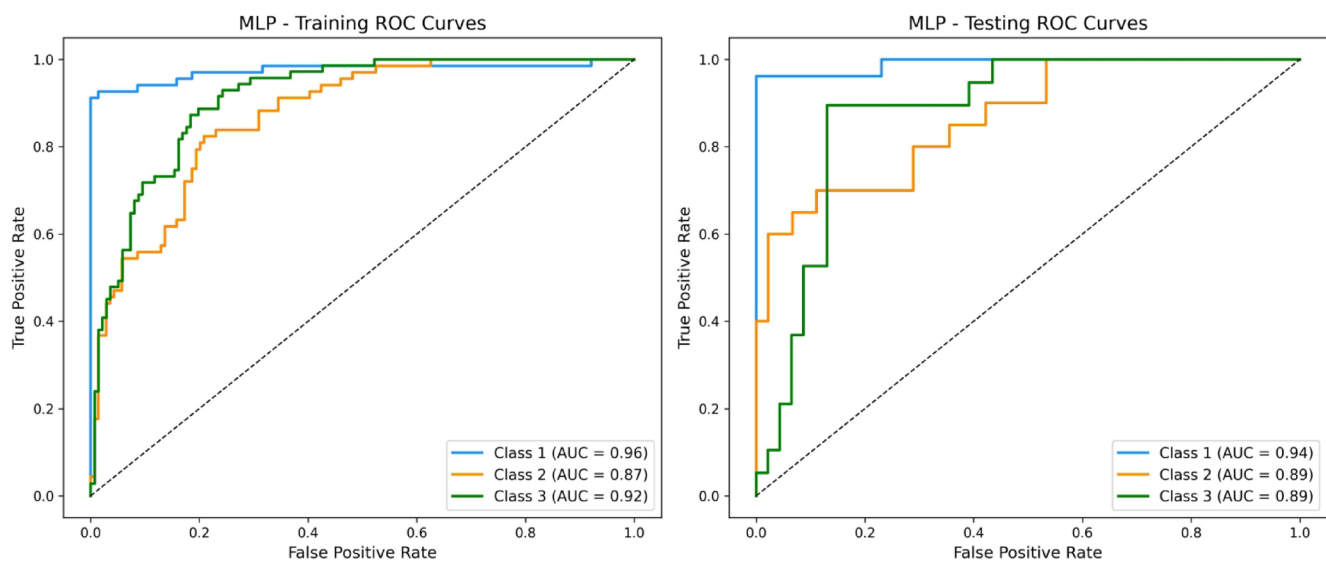


Fig. 12. ROC curves for MLP model. Moderate AUC performance is observed, particularly for Class 2, suggesting reduced sensitivity in this class.

significance (e.g., $p = 0.0625$) were observed, particularly between HydroPredictor and SVM. These borderline values warrant attention when interpreting relative model effectiveness.

Mean Rank Analysis: Fig. 16 visualizes the mean ranks of all models based on F1-score, where lower ranks indicate better performance. HydroPredictor attained the lowest mean rank, surpassing RF, CatBoost, and MLP, while SVM consistently ranked the poorest. These results reinforce the strong overall ranking of HydroPredictor across multiple folds and performance metrics.

Although RF achieved a competitive mean rank, its wider train-test accuracy gap suggests a higher degree of overfitting compared to HydroPredictor. In contrast, HydroPredictor demonstrated a minimal performance gap between training and testing phases (Accuracy deviation = 0.35%), underscoring its superior generalization ability.

Moreover, while CatBoost and RF performed competitively, their p -values with respect to HydroPredictor ($p > 0.05$) indicate statistical equivalence rather than superiority. This nuanced comparison places HydroPredictor as the most balanced and reliable model, offering both empirical accuracy and statistical robustness.

These insights collectively validate HydroPredictor as a high-performing benchmark for groundwater classification tasks, combining consistent predictive power with generalization strength. The hyperparameter

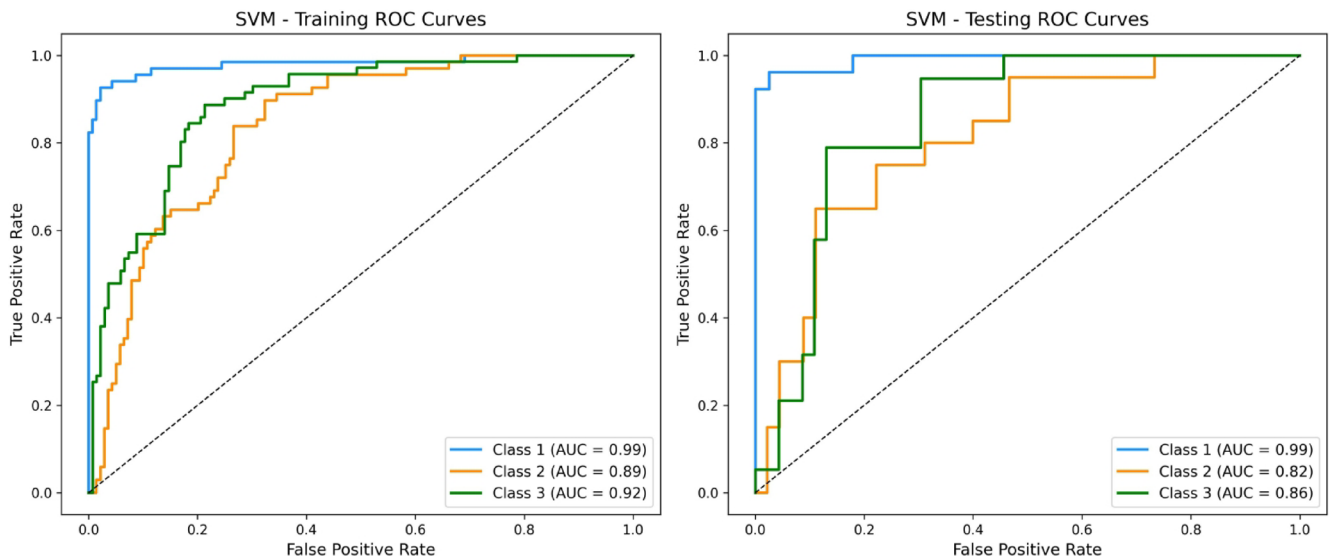


Fig. 13. ROC curves for SVM model. AUC scores indicate weaker class separation, especially for Class 2 and Class 3, during testing.

optimization process achieved via Optuna as shown in Table 1 further refined the model's performance, contributing to its stable and reliable outcomes.

Assessing hydropredictor for groundwater mapping in Feija

To evaluate the predictive performance of the HydroPredictor model, a groundwater suitability map was produced based on its classification outputs Fig. 17. The map categorizes the Feija Basin into three groundwater potential classes high, medium, and low representing areas with favorable, moderate, and limited groundwater availability, respectively. High-potential zones (blue) are predominantly concentrated in the central-western sector, notably around Bouzkar, Fom Lachar, and Argab-n-Tal. Medium-potential areas (yellow) span the northern and northeastern portions, extending toward Zagora and Anagam. In contrast, low-potential zones (gray) are mainly found in the southern and peripheral regions, likely constrained by limited recharge and unfavorable geomorphological conditions. Spatial validation Fig. 17 reveals strong agreement between model predictions and field observations. High recharge areas coincide with dense clusters of high-yield wells and large, full irrigation tanks. Conversely, zones classified as low recharge align with dry or low-yield wells and abandoned watermelon farms, reflecting the model's ability to capture the underlying hydrogeological heterogeneity of the basin.

Quantitatively, the model achieved an overall classification accuracy of 84.44% and a Cohen's Kappa coefficient of 0.767, indicating substantial agreement beyond chance. Confusion matrix results show that 19 out of 30 high-recharge locations were correctly classified, with 6 misclassified as medium and 5 as low. For the medium class, 27 points were accurately predicted, with 1 overestimated as high and 2 underestimated as low Fig. 18. All 30 low-recharge points were correctly classified, resulting in perfect precision and recall (1.00) for this category. These outcomes confirm the HydroPredictor model's statistical robustness and spatial reliability. Its accurate delineation of recharge potential especially the precise identification of low-recharge zones demonstrates its practical utility for groundwater assessment and management in arid and semi-arid environments such as the Feija Plain.

Discussion

In regions where groundwater monitoring networks are sparse and in-situ measurements are limited, predictive models are particularly susceptible to overfitting. This phenomenon was evident in traditional approaches such as Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP), which exhibited notable discrepancies between training and testing accuracy. Specifically, SVM showed a train-test gap of 6.47%, suggesting poor generalization capacity (Table 2). In contrast, HydroPredictor demonstrated a markedly lower gap of 0.35%, reflecting its enhanced ability to generalize from limited data, a critical requirement in semi-arid and data-scarce environments.

The superior performance of HydroPredictor can be attributed to its hybrid architecture, which combines tree-based structured feature learning via CatBoost with adaptive deep representations derived from MLP. CatBoost's embedding layer transforms raw hydroclimatic inputs into more structured and spatially coherent representations, capturing essential nonlinear patterns in groundwater dynamics. Simultaneously, the deep learning component applies dropout and L2 regularization techniques, effectively reducing model complexity and guarding against overfitting. This synergy allows HydroPredictor to balance expressiveness and robustness, enabling reliable performance even with small, noisy datasets.

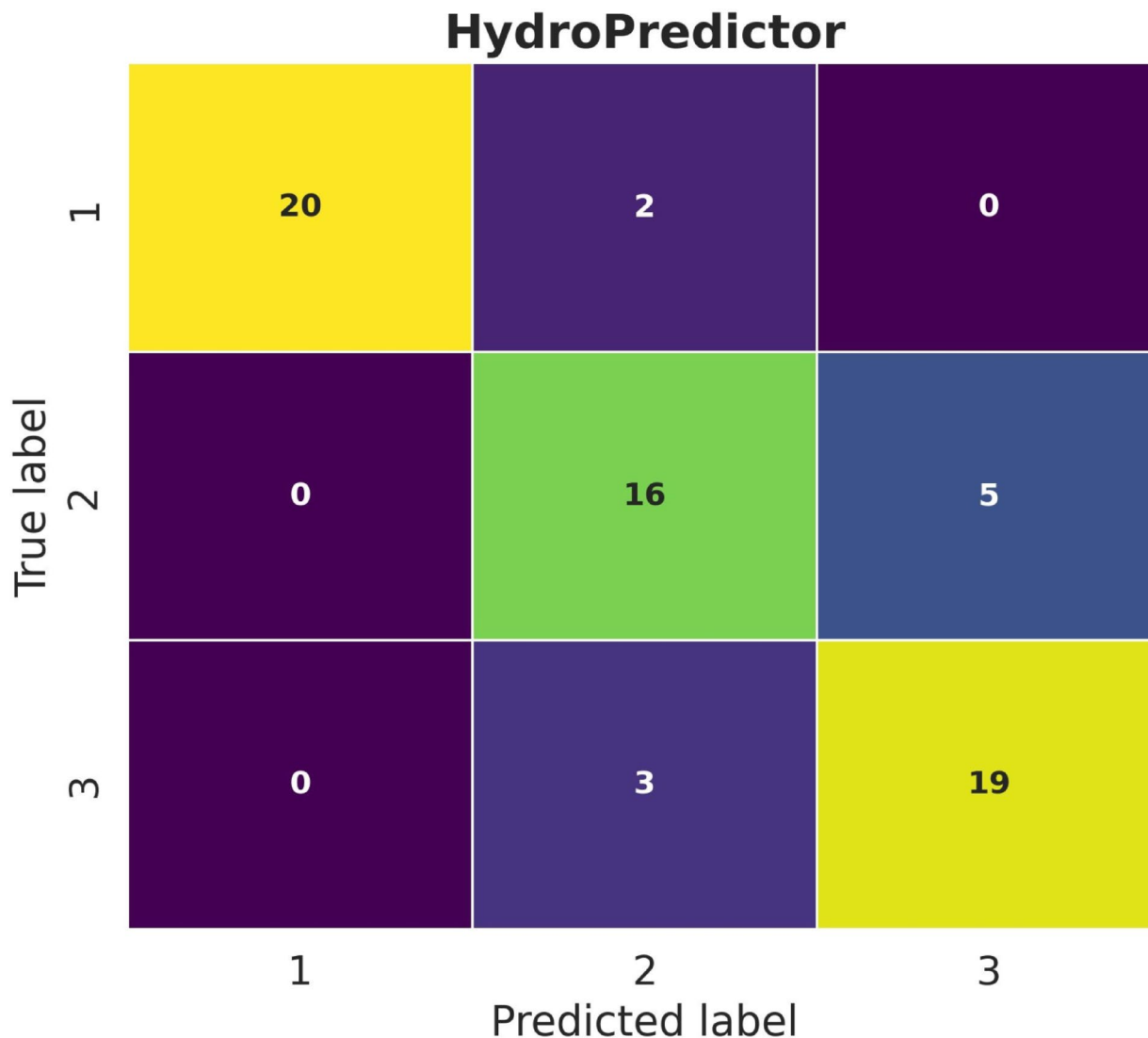


Fig. 14. Confusion matrix for HydroPredictor on the testing dataset.

Figure 9 confirms HydroPredictor's high discriminative power across all groundwater classes. Notably, it achieved 95% accuracy in identifying critically low groundwater conditions (Class 3), an essential capability for early warning and mitigation planning in drought-prone regions. Furthermore, the model reduced Class 2 misclassifications by 37.5% compared to Random Forest, highlighting its sensitivity in detecting transitional hydrological states that are often misrepresented by conventional models.

Table 2 further substantiates HydroPredictor's dominance, with an overall accuracy of 89.23%, surpassing traditional models such as CatBoost and Random Forest (both at 84.62%). While tree-based models remained competitive, their slight overfitting tendencies became apparent under test conditions. On the other hand, MLP and SVM suffered from both underperformance and instability, likely due to their limited capacity to extract meaningful patterns from sparse or noisy features without structured guidance.

Overall, these findings underscore HydroPredictor's potential as a reliable solution for groundwater classification in data-scarce regions. Its hybrid design not only ensures accuracy but also promotes stability across varying hydrological conditions, reinforcing the value of ensemble and deep-learning fusion strategies for addressing long-standing challenges in hydrological modeling. However, we acknowledge that the current validation is limited to the Feija Plain, and further testing on diverse sites and datasets is required to fully establish its generalizability. Within this scope, HydroPredictor demonstrates promise as a candidate for supporting groundwater monitoring and climate resilience planning frameworks.

The design and performance of the proposed HydroPredictor model were systematically benchmarked against a range of recent studies in groundwater potential modeling. Unlike conventional models, HydroPredictor uniquely integrates the categorical handling capabilities of CatBoost with the nonlinear representation power of a regularized Multi-Layer Perceptron (MLP). The hybrid architecture is further enhanced through Bayesian hyperparameter optimization using Optuna, specifically addressing the challenge of limited training data a constraint often overlooked in the current literature.

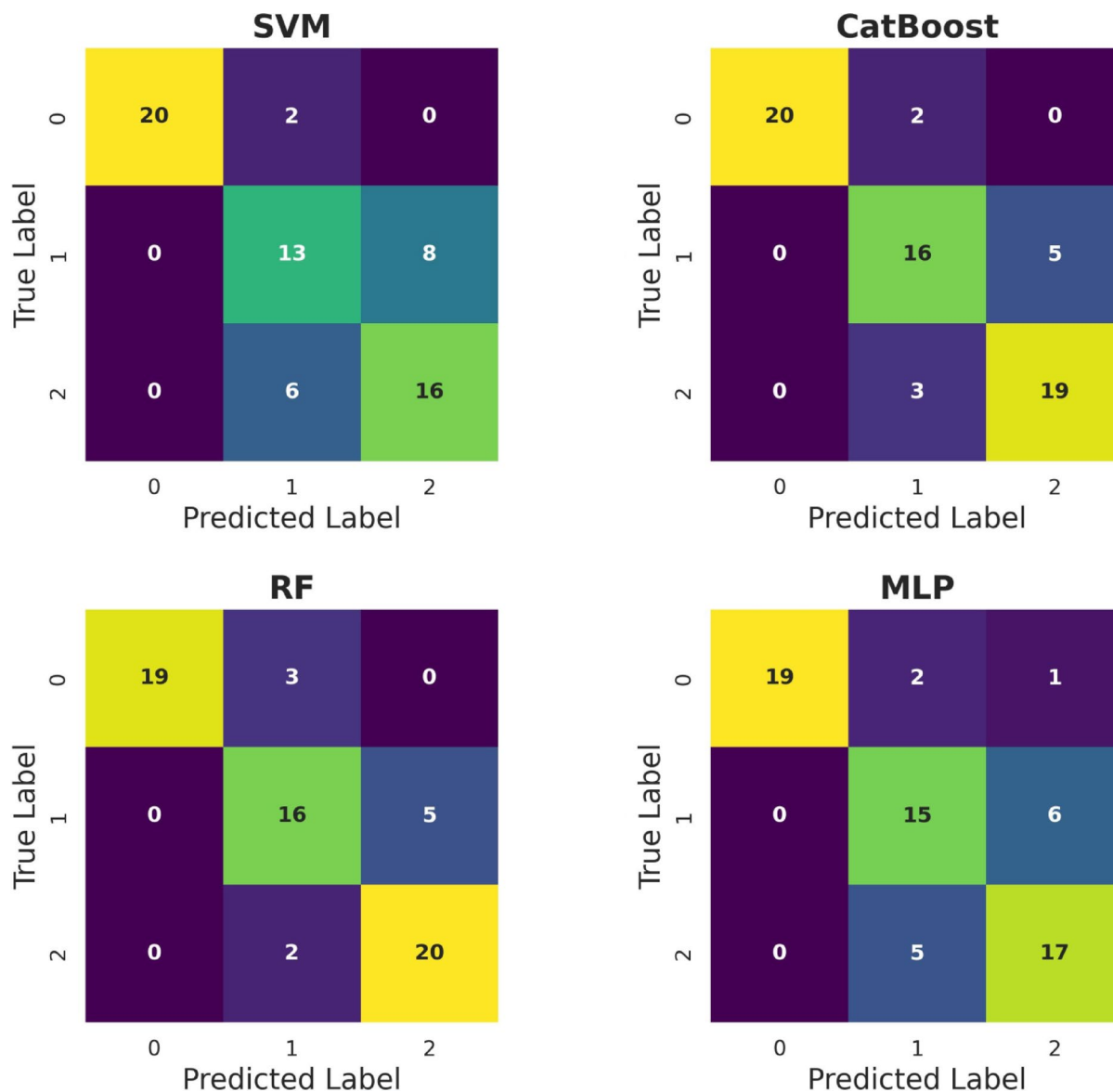


Fig. 15. Confusion matrices of baseline models: SVM, CatBoost, Random Forest, and MLP.

Model	CatBoost	HydroPredictor	MLP	RF	SVM
CatBoost	–	0.1875	0.4652	0.1088	0.0625
HydroPredictor	0.1875	–	0.125	0.4375	0.0625
MLP	0.4652	0.125	–	0.1875	0.0656
RF	0.1088	0.4375	0.1875	–	0.0625
SVM	0.0625	0.0625	0.0656	0.0625	–

Table 3. Wilcoxon Signed-Rank test *p*-values between models (based on Accuracy).

Numerous studies have demonstrated encouraging results in groundwater potential mapping using classical machine learning (ML) and ensemble techniques. For instance, Sarkar et al.¹⁵ utilized Artificial Neural Networks (ANNs) and other ML algorithms in Bangladesh, achieving an AUC of 0.875. Similarly, Halder et al.¹⁶ reported Random Forest (RF) performance with an F1-score of 0.944 and accuracy of 0.943, while¹⁷ developed the MBAB-SPAARC ensemble model, which yielded an AUC of 0.891. Despite their effectiveness, these models were evaluated on large, balanced datasets, limiting their direct applicability to heterogeneous or data-sparse contexts.

In the Moroccan context¹⁸, proposed a hybrid ensemble of RF, Support Vector Machines (SVM), and Logistic Regression (LR) for groundwater potential prediction in the Saiss Basin, attaining an AUC of 0.86. However, this

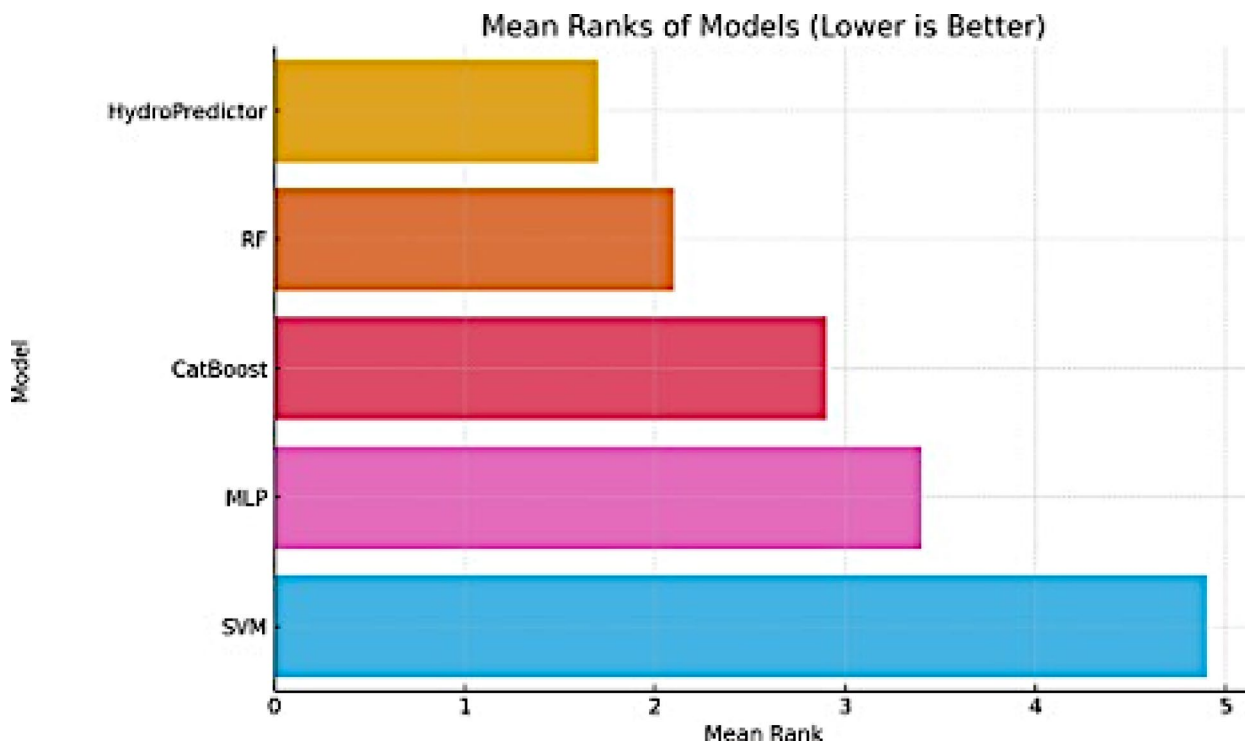


Fig. 16. Mean Ranks of Models Based on F1-score Across Cross-Validation Folds (Lower is Better).

model lacked a deep learning component and employed conventional hyperparameter tuning techniques such as grid search²¹. introduced the BS-MLP model combining Gradient Boosted Decision Trees (GBDT) with MLP achieving impressive performance (accuracy = 97.6%, F1-score = 96.8%). Nevertheless, this model was trained under data-rich conditions and did not incorporate statistical significance testing or mechanisms to mitigate overfitting in sparse settings.

Liu et al.²² employed an Extreme Learning Machine (ELM) optimized via the Sparrow Search Algorithm (SSA), attaining an AUC of approximately 0.84. Although this approach demonstrated the effectiveness of SSA for tuning, it did not explore hybrid design nor address limitations posed by data scarcity. Similarly, studies by Wei²³ utilizing GBDT models achieved AUCs between 0.84 and 0.88 but were dependent on dense observational networks. Gómez et al.²⁴ further emphasized the role of data preprocessing and feature engineering, demonstrating F1-scores exceeding 0.87. However, these strategies often rely on dataset-specific characteristics, which limit their generalizability.

Machine learning techniques have recently been integrated with GIS-based hydrological modeling to delineate groundwater potential zones across regions in India, Morocco, and Egypt^{46,68–71}. Although these approaches provided valuable insights, they generally lacked the use of hybrid model architectures, advanced regularization strategies, and rigorous statistical validation methods such as the Friedman or Wilcoxon signed-rank tests. In contrast, the study by Hosseini et al.⁷² introduced an ensemble model combining CatBoost and Random Forest, enhanced with Bayesian optimization and SHapley Additive exPlanations (SHAP)-based interpretability, achieving a notable AUC of 0.8778. However, this model did not incorporate neural network components and was developed using a relatively large dataset, thereby limiting its applicability in data-scarce scenarios.

HydroPredictor, by contrast, was explicitly designed and validated under data-scarce conditions ($N = 315$), attaining competitive results: AUCs exceeding 0.90, accuracy of 89.23%, and a minimal generalization gap (Δ Accuracy = 0.35%). The model offers several key advantages:

- Hybrid learning architecture: Combines CatBoost's interpretability with MLP's nonlinear learning capacity.
- Efficient hyperparameter tuning: Employs Bayesian optimization via Optuna, reducing computational cost and overfitting.
- Robust feature selection: Integrates multicollinearity analysis using VIF, mutual information, and correlation thresholds.
- Statistical rigour: Validated through Friedman and Wilcoxon signed-rank tests to ensure robustness across folds.
- Adaptability: Designed to function effectively in data-sparse and geologically heterogeneous environments.

The compared to models that rely heavily on dense observation networks, exhaustive grid search, or dataset-specific preprocessing, HydroPredictor represents a paradigm shift toward lightweight, scalable, and explainable

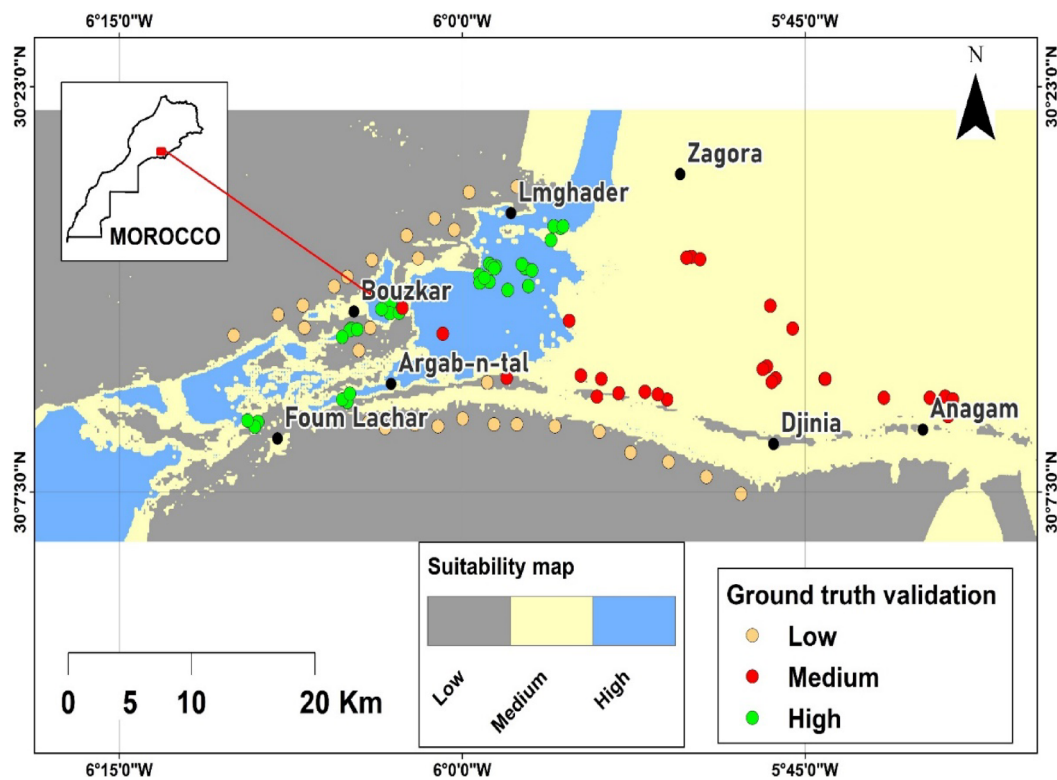


Fig. 17. Groundwater potential zones low (gray), medium (yellow), and high (blue) were generated using the HydroPredictor model. A total of 28,490 georeferenced points with predicted suitability values were produced by the model and served as input for Kriging interpolation to create the continuous suitability surface. Mapping was performed using ArcGIS 10.8 (Esri, 2020. ArcGIS Desktop: Release 10.8. Redlands, CA: Environmental Systems Research Institute. <https://www.esri.com> © Esri. All rights reserved.). Validation was conducted using 2024 field data from the Feija Basin, including irrigation tanks and wells categorized by yield: high (green), medium (red), and low (orange). The strong spatial alignment between high-potential zones and high-yield wells confirms the model's predictive reliability in this semi-arid region of southeast Morocco.

AI-driven solutions for groundwater potential modeling in resource-constrained and environmentally sensitive regions.

The statistical tests conducted in this study further substantiate the observed superiority of HydroPredictor. The Friedman test ($\chi^2 = 12.00$, $p = 0.0174$) indicated significant performance differences among models. Post-hoc Wilcoxon signed-rank tests confirmed that HydroPredictor performs comparably to CatBoost and RF ($p > 0.05$), while significantly outperforming models such as SVM ($p < 0.05$). These results validate HydroPredictor's robustness and reinforce its ability to generalize effectively across varied conditions and folds.

- Integration of Real-Time Data: Expanding the model to incorporate satellite-based remote sensing (e.g., Gravity Recovery and Climate Experiment (GRACE), Moderate Resolution Imaging Spectroradiometer (MODIS)) and IoT sensor networks can improve predictive accuracy and facilitate real-time groundwater monitoring.
- Spatiotemporal Expansion: Incorporating spatial interpolation methods (e.g., Kriging, Inverse Distance Weighting (IDW)) can enhance the model's applicability to regional-scale groundwater assessments.
- Uncertainty Quantification: Developing Bayesian inference techniques within HydroPredictor can provide confidence intervals for predictions, improving its interpretability for stakeholders.
- Hybrid Ensemble Learning: Future work can explore Long Short-Term Memory (LSTM)-CatBoost hybrid models for sequential groundwater forecasting, allowing dynamic adjustments based on climate change trends. Reliable groundwater prediction is crucial for policy-driven decision-making in urban planning, agriculture, and water resource management. HydroPredictor's ability to identify critically low groundwater levels Fig. 14 with high accuracy makes it a valuable tool for policymakers in drought-prone and water-stressed regions. While HydroPredictor demonstrates strong performance, this study has limitations. First, the model was trained on data from semi-arid regions, which may limit its generalizability to humid climates without recalibration. Second, the current framework relies on historical data; real-time validation using IoT or satellite telemetry is needed for operational deployments. Finally, HydroPredictor's hybrid architecture increases computational costs (e.g., 45% longer training time vs. CatBoost), posing challenges for resource-constrained settings. Future work should address these trade-offs.

Ground Truth	GT_High	19	6	5
	GT_Medium	1	27	2
	GT_Low	0	0	30
		Pred_High	Pred_Medium	Pred_Low
		Model Prediction		

Fig. 18. Confusion Matrix for Groundwater Recharge Classification Based on Field Validation ($n=90$).

Conclusion

This study introduces *HydroPredictor*, a novel hybrid machine learning framework designed to address the challenges of groundwater potential prediction in data-scarce and environmentally sensitive regions. By combining the structured feature transformation capabilities of CatBoost with the nonlinear representation learning of a regularized Multi-Layer Perceptron (MLP), *HydroPredictor* demonstrates a robust balance between predictive accuracy, generalization, and interpretability. The model achieved superior performance across all evaluation metrics, including an accuracy of 89.23%, F1-score of 0.8937, AUC scores exceeding 0.90, and a minimal generalization gap (Δ Accuracy=0.35%). Statistical validation using the Friedman and Wilcoxon signed-rank tests further confirmed the significance of its improvements over traditional models such as Random Forest, SVM, and standalone MLP. *HydroPredictor*'s architecture featuring tree-based embeddings, deep learning classification, and ensemble decision fusion enables effective modeling of complex, nonlinear groundwater dynamics. Spatial validation in the Feija Basin underscores its practical relevance, with accurate delineation of groundwater recharge zones, particularly low-recharge areas critical for drought mitigation. Compared to previous studies, *HydroPredictor* introduces several methodological innovations: Bayesian hyperparameter optimization via Optuna, rigorous multicollinearity filtering, and statistical robustness across folds. These enhancements position the model as a scalable and transferable decision-support tool for groundwater assessment in resource-constrained contexts. Future work will aim to integrate real-time satellite and sensor data, enhance spatial generalization through interpolation techniques, and incorporate uncertainty quantification. Ultimately, *HydroPredictor* contributes a scientifically grounded, operationally relevant framework for sustainable groundwater management and climate-resilient water governance.

Data availability

The dataset used and analyzed during the current study is available from the corresponding author upon reasonable request.

Received: 28 April 2025; Accepted: 16 October 2025

Published online: 18 December 2025

References

1. He, M. Y. et al. Lithium isotope fractionation in Weinan loess and implications for pedogenic processes and groundwater impact. *Glob Planet. Change*. **252**, 104865 (2025).
2. Nazir, J. et al. Delineation and validation of GIS-based groundwater potential zones under arid to semi-arid environment using multi-influence-factors approach. *Geol. Ecol. Landscapes*. 1–17. <https://doi.org/10.1080/24749508.2024.2392382> (2024).
3. El Bouazzaoui, I., Lamhour, O., Ait Brahim, Y., Najmi, A. & Bougadir, B. Three decades of groundwater drought research: evolution and trends. *Water (Basel)*. **16**, 743 (2024).
4. Waqas, M. et al. A comprehensive review of the impacts of climate change on agriculture in Thailand. *Farming Syst.* **3**, 100114 (2025).
5. Abegeja, D. & Nedaw, D. Identification of groundwater potential zones using Geospatial technologies in Meki Catchment, Ethiopia. *Geol. Ecol. Landscapes*. 1–16 <https://doi.org/10.1080/24749508.2024.2392380> (2024).
6. Pourmorad, S., Kabolizade, M. & Dimuccio, L. A. Artificial intelligence advancements for accurate groundwater level modelling: an updated synthesis and review. *Appl. Sci.* **14**, 7358 (2024).
7. Jayakumar, D., Bouhoula, A. & Al-Zubari, W. K. Unlocking the potential of artificial intelligence for sustainable water management focusing operational applications. *Water (Basel)*. **16**, 3328 (2024).
8. Malakar, P., Sarkar, S., Mukherjee, A., Bhanja, S. & Sun, A. Y. In *Global Groundwater* 545–557 (Elsevier, 2021).
9. Afan, H. A. et al. Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. *Eng. Appl. Comput. Fluid Mech.* **15**, 1420–1439 (2021).
10. Rödiger, T., Geyer, S., Odeh, T. & Siebert, C. Data scarce modelling the impact of present and future groundwater development on Jordan multiaquifer groundwater resources. *Sci. Total Environ.* **870**, 161729 (2023).
11. Borzi, I. Modeling groundwater resources in Data-Scarce regions for sustainable management: methodologies and limits. *Hydrology* **12**, 11 (2025).
12. Safonova, A. et al. Ten deep learning techniques to address small data problems with remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **125**, 103569 (2023).
13. Sun, Y., Pang, S., Qiu, Z. & Zhang, Y. Efficient lithology classification from small-sample well logging data processed by wavelet thresholding algorithm: integrating meta-learning with self-attention mechanism model. *Geoenergy Sci. Eng.* **246**, 213629 (2025).
14. Elahi, M., Afolaranmi, S. O. & Lastra, M. Perez Garcia, J. A. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artif. Intell.* **3**, 43 (2023).
15. Sarkar, S. K. et al. Future groundwater potential mapping using machine learning algorithms and climate change scenarios in Bangladesh. *Sci. Rep.* **14**, 10328 (2024).
16. Halder, K. et al. Application of bagging and boosting ensemble machine learning techniques for groundwater potential mapping in a drought-prone agriculture region of Eastern India. *Environ. Sci. Eur.* **36**, 155 (2024).
17. Wang, Z. et al. Novel ensemble models based on the Split-Point sampling and node attribute subsampling classifier for groundwater potential mapping. *Earth Space Science* **11**, (2024).
18. Ragragui, H. et al. Mapping and modeling groundwater potential using machine learning, deep learning and ensemble learning models in the Saiss basin (Fez-Meknes region, Morocco). *Groundw. Sustain. Dev.* **26**, 101281 (2024).
19. Sahour, S. et al. Evaluation of machine learning algorithms for groundwater quality modeling. *Environ. Sci. Pollut. Res.* **30**, 46004–46021 (2023).
20. Zegaar, A., Ounoki, S. & Telli, A. Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process. *Water Resour. Manage.* **38**, 621–637 (2024).
21. Chen, W., Xu, D., Pan, B., Zhao, Y. & Song, Y. Machine Learning-Based water quality classification assessment. *Water (Basel)*. **16**, 2951 (2024).
22. Liu, R. et al. Spatial prediction of groundwater potentiality using machine learning methods with grey Wolf and sparrow search algorithms. *J. Hydrol. (Amst)*. **610**, 127977 (2022).
23. Wei, A. et al. Application of machine learning to groundwater spring potential mapping using averaging, bagging, and boosting techniques. *Water Supply*. **22**, 6882–6894 (2022).
24. Gómez-Escalonilla, V., Martínez-Santos, P. & Martín-Loeches, M. Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions, Mali. *Hydrol. Earth Syst. Sci.* **26**, 221–243 (2022).
25. Aboubakr, E. Z. et al. Hydrogeological synthesis of groundwater resources: case of the Feija watershed (South-east of Morocco). *J. Environ. Agricultural Stud.* **2**, 85–94 (2021).
26. Ait Hamza, M., El Faskaoui, B. & Fermin, A. Les Oasis du Drâa Au Maroc. *Hommes Migr.* 56–69. <https://doi.org/10.4000/homme-smigrations.1241> (2010).
27. Berger, E. et al. Social-ecological interactions in the Draa river Basin, Southern morocco: towards nature conservation and human well-being using the IPBES framework. *Sci. Total Environ.* **769**, 144492 (2021).
28. El Qorchi, F. et al. Analyzing Temporal patterns of Temperature, Precipitation, and drought incidents: A comprehensive study of environmental trends in the upper Draa Basin, Morocco. *Water (Basel)*. **15**, 3906 (2023).
29. Kaczmarek, N. et al. Water quality, biological quality, and human well-being: water salinity and scarcity in the Draa river basin, Morocco. *Ecol. Indic.* **148**, 110050 (2023).
30. Yacoubi Khebiza, M. M. M. Environmental vulnerability to climate change and anthropogenic impacts in Dryland, (Pilot study: middle Draa Valley, South Morocco). *J Earth Sci. Clim. Change* **s11**, (2014).
31. Mahjoubi, I., Bossenbroek, L., Berger, E. & Frör, O. Analyzing stakeholder perceptions of water ecosystem services to enhance resilience in the middle Drâa Valley, Southern Morocco. *Sustainability* **14**, 4765 (2022).
32. Meskour, A., Ahattab, J., Aachib, M. & Hasnaoui, M. D. Assessing the impact of drought and upstream dam construction on agriculture in arid and semi-arid regions: a case study of the middle Draa Valley, Morocco. *Environ. Monit. Assess.* **197**, 236 (2025).
33. Moumane, A. et al. Monitoring long-term land use, land cover change, and desertification in the Ternata oasis, middle Draa Valley, Morocco. *Remote Sens. Appl.* **26**, 100745 (2022).
34. Rössler, M. et al. II-5.4 migration and resource management in the Drâa Valley, Southern Morocco. *Impacts Global Change Hydrol. Cycle West. Northwest. Africa* **15** (2010).
35. Kaczmarek, N. et al. Nature conservation in the Draa basin (Morocco): History, present situation, and future challenges. *J. Nat. Conserv.* **88**, 127038 (2025).
36. Amiha, R., Kabbachi, B., Ait Haddou, M. & Bouchriti, Y. Spatiotemporal assessment of groundwater quality in the Feija aquifer, Southern morocco: impacts of agricultural practices and climate change. *Mediterranean Geoscience Reviews*. <https://doi.org/10.1007/s42990-025-00172-8> (2025).
37. Fico, J. Frontiers of fortune: mobilising land, water, and collective identity for watermelon production in southeastern Morocco. *J. North. Afr. Stud.* 1–24 <https://doi.org/10.1080/13629387.2024.2404952> (2024).
38. Fico, J. & Kenti, A. Living on luck the story behind zagora's watermelons. *Heinrich Böll Stiftung: Rabat Maroc* (2023).

39. Moumane, A. et al. Monitoring Spatiotemporal variation of groundwater level and salinity under land use change using integrated field measurements, GIS, geostatistical, and remote-sensing approach: case study of the Feija aquifer, middle Draa watershed, Moroccan Sahara. *Environ. Monit. Assess.* **193**, 769 (2021).
40. Ait Lemkademe, A. et al. Origin and salinization processes of groundwater in the Semi-Arid area of Zagora Graben, Southeast Morocco. *Water (Basel)*. **15**, 2172 (2023).
41. Lamqadem, A. A., Saber, H. & Pradhan, B. Long-Term monitoring of transformation from pastoral to agricultural land use using Time-Series landsat data in the Feija basin (Southeast Morocco). *Earth Syst. Environ.* **3**, 525–538 (2019).
42. Karmaoui, A. & Balica, S. A new flood vulnerability index adapted for the pre-Saharan region. *Int. J. River Basin Manage.* **19**, 93–107 (2021).
43. Moumane, A. et al. Lake iriqui's remarkable revival: field observations and a Google Earth engine analysis of its recovery after over half a century of desiccation. *Land. (Basel)*. **14**, 104 (2025).
44. Moumane, A., Bahouq, T., Ghazi, A. & Integrating, G. I. S. Remote sensing, and analytic hierarchical process (AHP) for mapping of groundwater-recharge potential zones in Kenitra Province, morocco: toward sustainable development goals achievement. *EuroMediterr J. Environ. Integr.* **10**, 831–851 (2025).
45. Johannsen, I., Hengst, J., Goll, A., Höllermann, B. & Diekkrüger, B. Future of water supply and demand in the middle Drâa Valley, Morocco, under climate and land use change. *Water (Basel)*. **8**, 313 (2016).
46. Anand, V. et al. Evaluating groundwater potential with the synergistic use of Geospatial methods and advanced machine learning approaches. *Discover Cities*. **2**, 56 (2025).
47. Elmotawakkil, A., Sadiki, A. & Enneya, N. Predicting groundwater level based on remote sensing and machine learning: a case study in the Rabat-Kénitra region. *J. Hydroinformatics*. **26**, 2639–2667 (2024).
48. Barman, J., Zuali, V. L. H., Bindajam, F., Mallick, A. A., Abdo, H. G. & J. & Detection of groundwater conditioning factors in a hilly environment. *Appl. Water Sci.* **14**, 88 (2024).
49. Joshi, H., Bohra, M. S., Kumar, D., Rani, M. & Arya, O. P. Groundwater recharge potentiality mapping for opportune augmentation following two MCDMs for a Western Himalayan watershed. *Discover Water*. **4**, 45 (2024).
50. Song, Q., Liu, Y., Wang, Z. & Xu, Z. Assessing groundwater artificial recharge suitability in the Mi river basin using GIS, RS, and FAHP: a comprehensive analysis with seasonal variations. *Appl. Water Sci.* **15**, 39 (2025).
51. Arya, V. & Rao, M. S. Groundwater recharge potential index and artificial groundwater recharge in the alluvial soils of the middle Ganga basin. *Discover Appl. Sci.* **6**, 367 (2024).
52. Nazaripour, H., Sedaghat, M. & Shafaie, V. Movahedi Rad, M. Strategic assessment of groundwater potential zones: a hybrid Geospatial approach. *Appl. Water Sci.* **14**, 185 (2024).
53. Meng, F. et al. Identification and mapping of groundwater recharge zones using multi influencing factor and analytical hierarchy process. *Sci. Rep.* **14**, 19240 (2024).
54. Olabode, O. F. Potential groundwater recharge sites mapping in a typical basement terrain: a GIS methodology approach. *J. Geovisualization Spat. Anal.* **3**, 5 (2019).
55. Joseph, V. R. Optimal ratio for data splitting. *Stat. Anal. Data Mining: ASA Data Sci. J.* **15**, 531–538 (2022).
56. Bichri, H., Chergui, A. & Hain, M. Investigating the impact of Train / Test split ratio on the performance of Pre-Trained models with custom datasets. *International J. Adv. Comput. Sci. Applications* **15**, (2024).
57. Chan, J. Y. L. et al. Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics* **10**, 1283 (2022).
58. Schober, P., Boer, C. & Schwarte, L. A. Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.* **126**, 1763–1768 (2018).
59. Mandros, P., Boley, M. & Vreeken, J. Discovering dependencies with reliable mutual information. *Knowl. Inf. Syst.* **62**, 4223–4253 (2020).
60. Vörösmarty, G. & Dobos, I. Green purchasing frameworks considering firm size: a multicollinearity analysis using variance inflation factor. *Supply Chain Forum: Int. J.* **21**, 290–301 (2020).
61. Kim, Y., Street, W. N. & Menczer, F. Meta-evolutionary ensembles. in *Proceedings of the International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)* 3, 2791–2796 (IEEE, 2002). 3, 2791–2796 (IEEE, 2002). (2002).
62. Kim, Y. Boosting and measuring the performance of ensembles for a successful database marketing. *Expert Syst. Appl.* **36**, 2161–2176 (2009).
63. Hu, D. et al. Machine Learning–Finite element mesh Optimization-Based modeling and prediction of Excavation-Induced shield tunnel ground settlement. *Int. J. Comput. Methods* **22**, (2025).
64. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
65. Dorogush, A. V., Ershov, V. & Gulin, A. CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363* (2018).
66. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
67. Srinivasu, P. N. et al. XAI-driven catboost multi-layer perceptron neural network for analyzing breast cancer. *Sci. Rep.* **14**, 28674 (2024).
68. Mansour, B. M. H. & Kaiser, M. F. A refined Geospatial approach for groundwater potentiality mapping and optimal retention dams site selection: a case study Wadi Sudr, Gulf of Suez, Egypt. *Environ. Earth Sci.* **84**, 144 (2025).
69. Shams, E. T. M. M. et al. Geospatial intelligence and multi-criteria analysis for mapping groundwater potential zones and sustainable resource management in Wadi Qena Basin, Eastern Desert, Egypt. *Appl. Water Sci.* **15**, 145 (2025).
70. Elmotawakkil, A. et al. Artificial intelligence for groundwater recharge prediction in an arid region: application of tabular deep learning models in the Feija Basin, Morocco. *Frontiers Remote Sensing* **6**, (2025).
71. Moumane, A. et al. Remote Sensing, and machine learning to optimize sustainable groundwater recharge in arid mediterranean landscapes: A case study from the middle Draa Valley, Morocco. *Water (Basel)*. **17**, 2336 (2025). Integrating GIS.
72. Hosseini, F. S., Jafari, A., Zandi, I., Alesheikh, A. A. & Rezaie, F. Groundwater potential mapping using optimized decision Tree-Based ensemble learning model with local and global explainability. *Water (Basel)*. **17**, 1520 (2025).

Author contributions

AE and AM data curated, conceptualize, wrote the main manuscript text ; AZ and AS methodology. visualization, wrote the main manuscript text ; JAK and MB investigated, formal analysis, wrote the main manuscript text ; SKB and NE supervised, investigation, wrote the main manuscript text.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.K.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025