Application of Machine Learning Algorithms for Groundwater Level Prediction in the Najafabad Plain

Sahar Davari[1], Saeid Eslamian[1], Mohammad Jamali[1]*, Hamid Reza Safavi[2]

1. Department of Water Science and Engineering, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Iran. E-mail: sahar.davari.iut@gmail.com, saeid@iut.ac.ir; mohammad.jamali@ag.iut.ac.ir
2. Department of Civil Engineering, Isfahan University of Technology, Isfahan, Iran. E-mail: hasafavi@iut.ac.ir

***Correspondence: mohammad.jamali@ag.iut.ac.ir**

S1: Methodological Framework Flowchart

Figure S1 illustrates the overall methodology adopted in this study. The process begins with the collection of hydrometeorological and groundwater data. During the preprocessing phase, outliers are identified and removed, and missing values are interpolated to ensure data continuity. Following preprocessing, the temporal and spatial variability of groundwater levels is analyzed. Based on hydrogeological and topographic characteristics, the study area is divided into five zones to enable localized modeling. In the modeling phase, three machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost)—are applied independently within each zone. The models are trained and validated using five-fold cross-validation, and hyperparameters are optimized through grid search to enhance predictive performance.
Model performance is assessed using statistical indicators including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Finally, results across the five zones are compared to identify the most effective modeling approach for predicting groundwater level fluctuations.

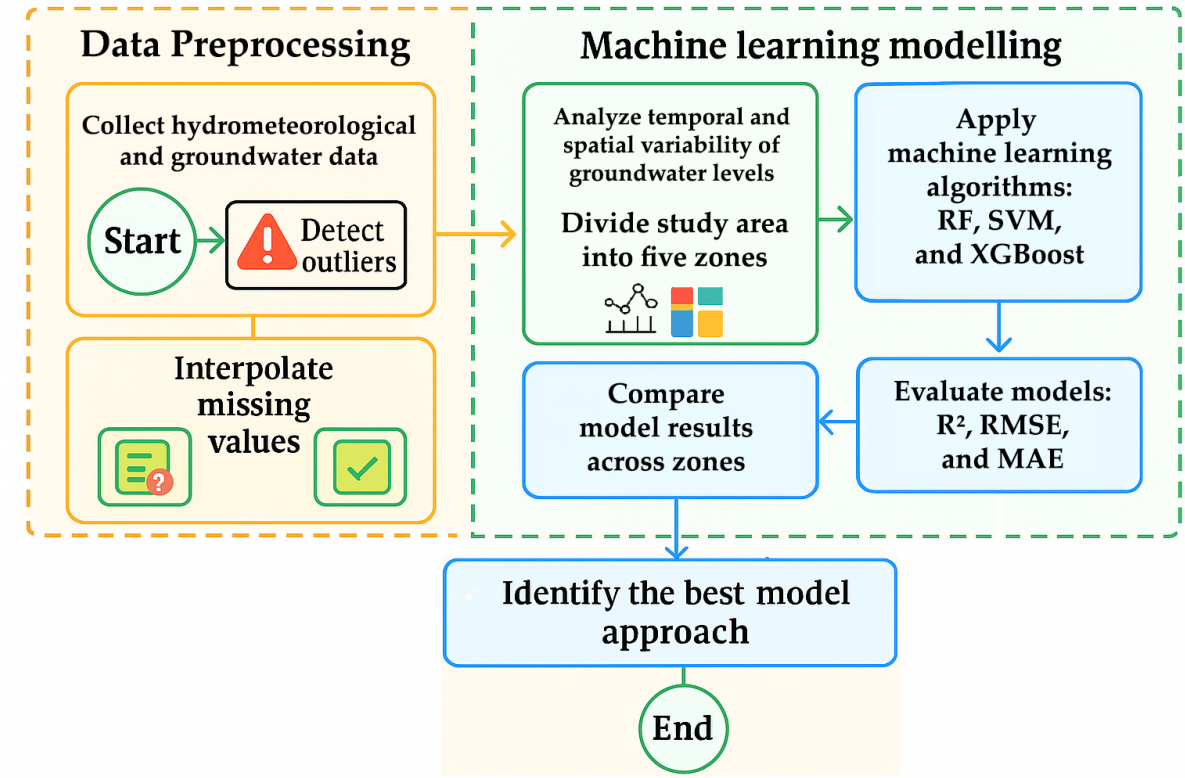

Figure S1. Methodological framework adopted in this study.

S2: Machine Learning Model Implementation Code

The Python source codes used for groundwater level prediction in this study, including model training, evaluation, and zonal analysis, are available at the following GitHub repository:

GitHub Repository – Groundwater-ML-Najafabad

The repository includes:

Data preprocessing scripts

Model development for RF, XGBoost, and SVM

Performance evaluation metrics

Example notebooks and results.