

UDK/UDC: 004.85:627.82(282)(65)

Prejeto/Received: 12.03.2026

Izvirni znanstveni članek – *Original scientific paper*

Sprejeto/Accepted: 14.08.2026

DOI: [10.15292/acta.hydro.2026.03](https://doi.org/10.15292/acta.hydro.2026.03)

Objavljeno na spletu/Published online: 06.10.2026

PIEZOMETRIC LEVEL FORECASTING AT THE BOUHANIFIA DAM IN ALGERIA USING MACHINE LEARNING: GMDH, LSTM, AND XGBOOST

PIEZOMETRIČNO NAPOVEDOVANJE NIVOJA VODE NA JEZU BOUHANIFIA V ALŽIRIJI Z UPORABO STROJNEGA UČENJA: GMDH, LSTM IN XGBOOST

Slimane Benyahia^{1,2}, Bouchrit Rouissat¹, Nadia Smail¹, Fatah Touati²

¹RISAM Lab, Department of Civil Engineering, Faculty of Technology, Abou Bekr Belkaid University, Tlemcen, Algeria.

²Department of Space Geodesy, Centre of Space Techniques, Algerian Space Agency, Oran, Algeria.

Abstract

Dam safety monitoring relies on the integrated analysis of multiple interdependent parameters, including seepage rates, piezometric levels, reservoir water elevation, pore pressures, structural displacements, and climatic variables. Interpreting these complex datasets using advanced modeling techniques is essential for assessing structural integrity and anticipating potential risks. This study proposes a machine learning-based approach to predicting piezometric levels, a key indicator in the hydro-mechanical behavior analysis of embankment dams. Three advanced models were developed and compared: the Group Method of Data Handling (GMDH), Long Short-Term Memory recurrent neural networks (LSTM), and eXtreme Gradient Boosting (XGBoost). The models were trained and tested on a multi-year time series (2017–2025) comprising reservoir water levels, climatic variables (precipitation and temperature), and measurements from four piezometers installed within the Bouhanifia concrete-faced rockfill dam (CFRD) in Algeria. The results demonstrated high predictive accuracy across all piezometers, supported by robust performance metrics: root mean square errors (RMSE) ranging from 0.066 m to 0.153 m in validation, mean absolute relative errors (MARE) between 0.099% and 0.235%, correlation coefficients (R) exceeding 0.90 for all models and piezometers, Loague and Green coefficients (LG) ranging from 0.471 to 0.709, and variance accounted for (VAF) between 81.1% and 95%. Comparative analysis highlighted the distinct strengths of each model: XGBoost achieved the highest accuracy (RMSE as low as 0.066 m for P01), LSTM demonstrated remarkable stability across all monitoring points with the highest R (0.975 for P03), and GMDH offered a balance of accuracy and interpretability with reduced computational complexity (<10 seconds).

¹Stik / Correspondence: slimane.benyahia@univ-tlemcen.dz

© Benyahia S. et al.; This is an open-access article distributed under the terms of the [Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\) licence](https://creativecommons.org/licenses/by-sa/4.0/).

© Benyahia S. et al.; Vsebina tega članka se sme uporabljati v skladu s pogoji [licence Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/).

Keywords: Machine Learning, Piezometric Levels, Dam Safety Monitoring, GMDH, LSTM, XGBoost, Bouhanifia Dam.

Izvleček

Spremljanje varnosti jezov temelji na celostni analizi številnih medsebojno odvisnih parametrov, vključno s hitrostjo pronicanja, piezometričnimi nivoji, višino vode v rezervoarju, pornimi tlaki, strukturnimi premiki in podnebnimi spremenljivkami. Interpretacija teh kompleksnih naborov podatkov z uporabo naprednih tehnik modeliranja je bistvena za oceno strukturne celovitosti in predvidevanje potencialnih tveganj. Ta študija predlaga pristop za napovedovanje piezometričnih nivojev, ki temelji na strojnem učenju, kar je ključni kazalnik pri analizi hidromehanskega vedenja nasipnih jezov. Razviti in primerjani so bili trije napredni modeli: skupinska metoda obdelave podatkov (GMDH), rekurentne nevronske mreže z dolgim kratkoročnim spominom (LSTM) in eXtreme Gradient Boosting (XGBoost). Modeli so bili usposobljeni in preizkušeni na večletnem časovnem nizu (2017–2025), ki vključuje nivoje vode v rezervoarjih, podnebne spremenljivke (padavine in temperaturo) ter meritve štirih piezometrov, nameščenih v skalometni pregradi Bouhanifia (CFRD) v Alžiriji. Rezultati so pokazali visoko natančnost napovedi pri vseh piezometrih, kar podpirajo robustne meritve uspešnosti: koren povprečne kvadratne napake (RMSE) v validaciji od 0,066 m do 0,153 m, povprečne absolutne relativne napake (MARE) med 0,099 % in 0,235 %, korelacijski koeficienti (R) nad 0,90 za vse modele in piezometre, koeficienti po Loagueu in Greenu (LG) od 0,471 do 0,709 ter pojasnjena varianca (VAF) med 81,1 % in 95 %. Primerjalna analiza je izpostavila različne prednosti vsakega modela: XGBoost je dosegel najvišjo natančnost (RMSE le 0,066 m za P01), LSTM je pokazal izjemno stabilnost na vseh merilnih točkah z najvišjim R (0,975 za P03), GMDH pa je ponudil ravnovesje med natančnostjo in interpretabilnostjo z zmanjšano računsko kompleksnostjo (manj kot 10 sekund).

Ključne besede: strojno učenje, piezometrični nivoji, spremljanje varnosti jezov, GMDH, LSTM, XGBoost, jez Bouhanifia.

1. Introduction

Dams are pieces of critical infrastructure for socioeconomic development, ensuring drinking water supply, irrigation, hydroelectric power generation, and flood protection. However, their management throughout their life cycle (design, construction, operation, and maintenance) faces complex challenges related to extreme environments, limited data, and aging structures (Cao et al., 2025). Dam safety fundamentally relies on continuous monitoring of key parameters, among which piezometric levels occupy a prominent position as fundamental indicators of the hydro-mechanical behavior of embankment dams (Khorchani et al., 2025).

Piezometric levels, which reflect seepage flows and pore pressures within the dam body and its foundation, are essential for detecting anomalous seepage and preventing failure risks (Ranković et al., 2014). Early detection of such anomalies is crucial for avoiding catastrophic failures

(Khorchani et al., 2025). Traditionally, predicting these levels has relied on statistical or deterministic approaches, notably the Finite Element Method (FEM). However, these methods struggle to capture the complex, nonlinear relationships between multiple hydrological, climatic, and geotechnical variables, and are limited in their ability to handle the uncertainties inherent in real-world monitoring data (Harbi et al., 2024).

Faced with these limitations, the emergence of artificial intelligence and machine learning has opened new opportunities for digitalization and intelligence in dam engineering (Cao et al., 2025). Numerous studies have demonstrated the effectiveness of AI techniques for predicting piezometric levels and seepage. Ranković et al. (2014) showed that a Feed-Forward Neural Network (FFNN) significantly outperforms Multiple Linear Regression (MLR) for piezometric level prediction. Ziggah et al. (2022) evaluated several AI models, demonstrating that the Group Method of Data Handling (GMDH) offers the best

performance among seven compared models, including LSSVM, M5 Prime, GPR, BPNN, SVM, and RBFNN.

Harbi et al. (2024) compared a Back Propagation Neural Network (BPNN) with its Genetic Algorithm-optimized version (BPNN-GA) and concluded that these AI models outperform the FEM for seepage prediction in earth dams. Khorchani et al. (2025) applied three AI techniques (FFNN, RBF, and SVM) to piezometric level prediction in a homogeneous earth-fill dam, demonstrating the superiority of the FFNN model with an Average Relative Variance (ARV) of 0.094, significantly better than RBF (0.462) and SVM (0.243).

Recent methodological advances have also been proposed to enhance predictive performance. Sharghi et al. (2019) demonstrated that the combined application of pre-processing and post-processing techniques improves the AI model's accuracy by nearly 30% for seepage modeling. Zheng et al. (2025) developed an innovative hybrid framework for permeability parameter inversion in dam foundations, achieving exceptional accuracy (maximum relative error < 1%) while being 129.5 times faster than the conventional FEM approach. Xie et al. (2026) proposed an optimization framework based on residual compensation, integrating TCN-BiLSTM and attention mechanisms, significantly improving the accuracy of existing monitoring models.

In the specific context of CFRD structures, Liao et al. (2025) explored the use of Generative Adversarial Networks (GANs) for intelligent dam profile design, demonstrating a considerable efficiency gain (reducing design time from 1-2 hours to approximately 6 minutes). Zhou et al. (2021) combined the finite element method with Support Vector Machines (SVM) for seismic fragility analysis of CFRDs, significantly reducing computation time while maintaining satisfactory accuracy.

Despite these advances, comparative applications of different AI techniques for piezometric level prediction in Algerian dams remain poorly documented. The Bouhanifia Dam, a strategic

CFRD structure in northwestern Algeria (ANBT, 2025), has been regularly monitored for piezometric levels since 2017, providing a substantial database for evaluating advanced predictive methods. Table 1 summarizes the main characteristics of this dam.

Within this framework, the present study aims to evaluate and compare the predictive performance of three complementary AI methods: the Group Method of Data Handling (GMDH), Long Short-Term Memory networks (LSTM), and eXtreme Gradient Boosting (XGBoost). These three approaches have demonstrated strong potential in predictive hydrology (Ni et al., 2021; Fan et al., 2019; Kratzert et al., 2018; Osman et al., 2021; Özbey et al., 2024; Yazdi et al., 2025; Waqas et al., 2024). The main objective is to identify, among these three methods, the one that offers the best balance between accuracy, robustness, and interpretability for piezometric level prediction in the specific context of the Bouhanifia Dam.

The specific contributions of this study are:

1. A rigorous comparison of three complementary AI methods on the same real-world dataset covering an extended period (2017-2025).
2. A multi-criteria evaluation framework integrating accuracy metrics (RMSE, MARE), linear association (R), skill scores (LG), and explained variance (VAF), following best practices (Ali & Abustan, 2014).
3. An analysis of spatial variability in predictive performance across different piezometers.
4. Practical recommendations for the operational deployment of these methods in dam monitoring systems.

By leveraging multi-year data from the Bouhanifia Dam, this research aims to contribute to the development of more reliable decision support tools for real-time monitoring and preventive safety management of dams, in the face of growing climatic and operational uncertainties.

2. Study Area and Data

2.1 Study Area

The study focuses on the Bouhanifia Dam, a Concrete-Faced Rockfill Dam (CFRD) located in Northwestern Algeria (Figure 2). The Bouhanifia Dam has a height of 54 m and a crest length of 464 m. Its initial capacity was 73 hm³, but according to the latest bathymetric survey (2012), its capacity has been reduced to 34.52 hm³ due to siltation, estimated at 0.73 hm³ per year. The average annual inflow is 80 hm³/year. The dam serves multiple purposes: drinking water supply for the surrounding municipalities, water transfer to neighboring regions, and irrigation of agricultural areas. The geological context consists of marl and limestone formations, with the dam foundation resting on impermeable marl layers. The climate is Mediterranean semi-arid, with average annual precipitation of 350 mm, concentrated mainly between October and April. The main characteristics of the Bouhanifia Dam are summarized in Table 1, and a view of the dam is shown in Figure 1 (ANBT, 2025).

2.2 Data Used

As part of the safety monitoring system for the Bouhanifia Dam (Figure 1), regular observations are conducted to collect the geodetic, geotechnical, and environmental variables essential for dam safety analysis. To evaluate the predictive performance of various artificial intelligence methods, monitoring data collected over a substantial period from October 2017 to April 2025 at a weekly sampling frequency were used to train and test the applied models. The dataset comprises 469 weekly observations and includes the following parameters: reservoir water level, piezometric water levels from four strategically installed piezometers (P01, P02, P03, and P04), temperature, and precipitation. It is important to note that the climatic variables (temperature and precipitation) were obtained from an automatic recording system and are therefore fully continuous, with no missing values or outliers. In contrast, the reservoir water level and piezometric measurements are collected through manual in situ measurements, which inherently contain some gaps and outliers due to measurement constraints and human factors.

Table 1: Main characteristics of Bouhanifia Dam.

Preglednica 1: Glavne značilnosti jezu Bouhanifia.

Dam	Bouhanifia
Type of embankment Dam	Concrete-faced rockfill dam
Coordinates (WGS84)	N35.2789°,W 0.0605°
Construction year	1929
First impoundment year	1940
Height (m)	54
Length (m)	464
Initial capacity (hm ³)	73
Current capacity(hm ³)	34.52
Annual average inflow (hm ³ /an)	80
Annual siltation rate (hm ³ /year)	0.73



Figure 1: Bouhanifia Dam.

Slika 1: Jez Bouhanifia.

The reservoir water level is the primary variable directly influencing pore pressures within the dam. Over the entire observation period, it oscillates between a minimum of 283.70 m and a maximum of 295.00 m, representing a total amplitude of approximately 11.30 m. The mean value is 288.50 m with a standard deviation of 2.80 m, indicating moderate variability. The coefficient of variation, which is low, reflects the relative stability of the water level despite seasonal and interannual fluctuations. Examination of the temporal distribution reveals a well-defined annual cycle, as illustrated in Figure 3, which presents the time series of upstream water level and piezometric levels. The highest levels occur during winter and spring, particularly between January and May, with

peaks often reaching 292 to 295 m. Conversely, the lowest levels occur during the summer and autumn period, between August and October, when the level can drop to 284 m. This seasonal variation is explained by the region's semi-arid Mediterranean hydrological regime, characterized by abundant winter precipitation and a pronounced summer drought, associated with high evaporation and irrigation withdrawals (Figure 4).

The four piezometers, installed at different strategic locations within the dam body and its foundation, exhibit distinct but consistent behaviors according to their positions, as shown in Figure 3. Piezometer P01 records the lowest levels, ranging from 248.55 to 252.10 m, with a mean of 250.30 m and a standard deviation of 0.45 m. Its location, furthest from the reservoir in a less permeable zone, explains this significant attenuation. Piezometer P02, with values ranging from 253.43 to 257.20 m, presents a mean of 255.50 m and a standard deviation of 0.55 m, occupying an intermediate position in the piezometric gradient. Piezometer P03, with levels from 254.32 to 257.81 m and a mean of 256.40 m, shows the strongest correlation with the reservoir level, indicating its location in a particularly permeable and responsive zone. Finally, Piezometer P04, with the highest values from 256.66 to 260.92 m and a mean of 258.80 m, exhibits the largest amplitude of variation (4.26 m), suggesting its position in the downstream area.

Correlation analysis reveals significant relationships between the reservoir level and piezometric levels, confirming the direct hydraulic connection between the reservoir and pore pressures within the dam. The reservoir level is well correlated with Piezometer P03 ($r = 0.83$), followed by P02 ($r = 0.80$), P04 ($r = 0.73$), and P01 ($r = 0.62$). This hierarchy reflects the relative proximity of each piezometer to the reservoir and the permeability of the intervening geological formations. Inter-piezometer correlations are also very high, notably between P02 and P03 ($r = 0.94$) and P03 and P04 ($r = 0.85$), indicating that these instruments respond consistently to the same hydraulic solicitations. The weaker correlation between P01 and the other piezometers suggests that this measurement point is either hydraulically less connected or is located in a zone with distinct drainage characteristics. These correlation results are consistent with the visual observations in Figure

3, where the series of P02, P03, and P04 follow the reservoir level variations more closely than P01.

Time lag analysis between reservoir level variations and piezometric responses reveals notable differences across piezometers, reflecting pressure wave propagation times through the dam body and its foundation. These lags can be observed in Figure 3, where certain piezometric series exhibit delays relative to the reservoir level. Piezometer P03 shows the fastest response, with an estimated lag of one to two weeks, indicating a short and permeable flow path. Piezometer P02 responds within two to three weeks, representing an intermediate propagation distance. Piezometer P04 exhibits a delay of three to four weeks, while Piezometer P01 shows the most significant lag of four to six weeks, suggesting its location in a zone of lower permeability or greater distance from the infiltration zone. These temporal lags are critically important for predictive modeling, as they require the use of lagged variables in artificial intelligence models to properly capture system dynamics.

The hydrological cycle of the Bouhanifia Dam follows a well-established annual rhythm, typical of semi-arid Mediterranean regions, as illustrated in Figures 3 and 4. Four distinct hydrological seasons can be identified: the fill season from January to March is characterized by a progressive rise in water levels due to winter precipitation and snowmelt from the surrounding highlands (Figure 4 clearly shows the increase in precipitation during this period); the plateau period in April and May sees the maximum reservoir level reached, typically between 290 and 295 m; the drawdown season from June to October experiences a regular decline in water levels due to intense evaporation (favored by high temperatures visible in Figure 4), irrigation withdrawals, and the absence of significant precipitation; and the recovery phase in November and December sees early autumn rains initiate a new filling cycle. Monthly mean levels perfectly illustrate this cycle: the lowest level is observed in September (285.5 m on average), while the highest is reached in February or March (approximately 290.8 m), corresponding to a mean seasonal amplitude of about 5 m.

Interannual analysis reveals a declining trend in the mean reservoir level over the study period, estimated at approximately 0.4 m per year, as can be

observed in Figure 3. This trend is explained by several combined factors: progressive siltation of the reservoir estimated at $0.73 \text{ hm}^3/\text{year}$ (Table 1), which reduces storage capacity; recurrent droughts in the region, notably in 2020 and 2022; and increasing withdrawals to meet the water needs of a growing population and expanding agriculture. The years 2018 and 2019 were relatively wet, with mean levels of 290 m, while the period from 2020 to 2022 experienced lower levels around 287 m, reflecting drier climatic conditions. Since 2023, a slight recovery has been observed, with mean levels approaching 289 m. These trends are visible in Figure 3, where a gradual decrease in piezometric and reservoir levels can be observed over the years.

Precipitation, although irregular, plays a fundamental role in reservoir recharge and consequently influences piezometric level evolution. In the Bouhanifia region, mean annual precipitation is approximately 350 mm,

concentrated mainly between October and April, as shown in Figure 4. Heavy rainfall episodes, though rare, can cause rapid rises in both reservoir and piezometric levels, with delays ranging from a few days to a few weeks depending on the permeability of the encountered formations. Precipitation series analysis shows high interannual variability, with very wet years (such as 2018) and years of marked drought (such as 2020), directly reflected in water levels and piezometric pressures, as observed in Figures 3 and 4. Temperature, while less directly influential than precipitation, acts as a modulator of the hydrological cycle through its effect on evaporation, which can reach high rates during summer months, as shown in Figure 4. Mean temperatures range from 0 to 5°C in winter to 35 – 40°C in summer, with an annual mean of approximately 18.5°C . High summer temperatures promote reservoir evaporation (estimated at several millimeters per day) and thus contribute to the seasonal decline in water levels, visible in Figure 3.

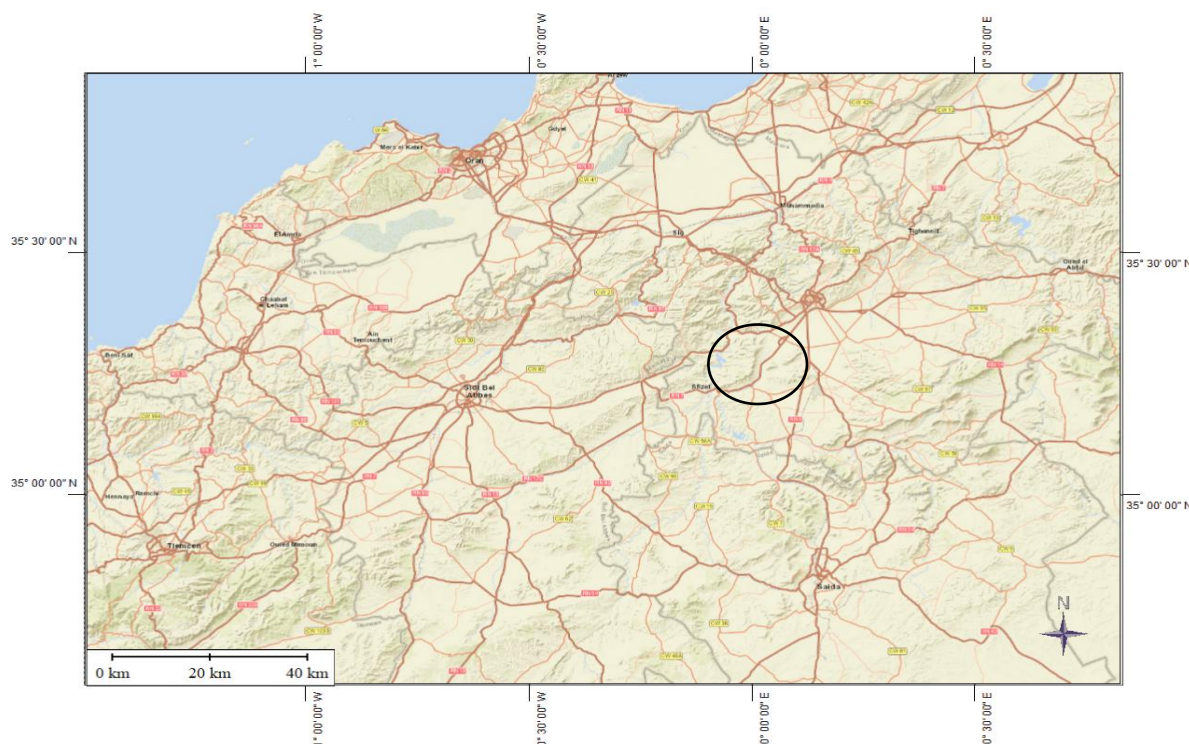


Figure 2: The geographic location of Bouhanifia Dam.

Slika 2: Geografska lokacija jezua Bouhanifia.

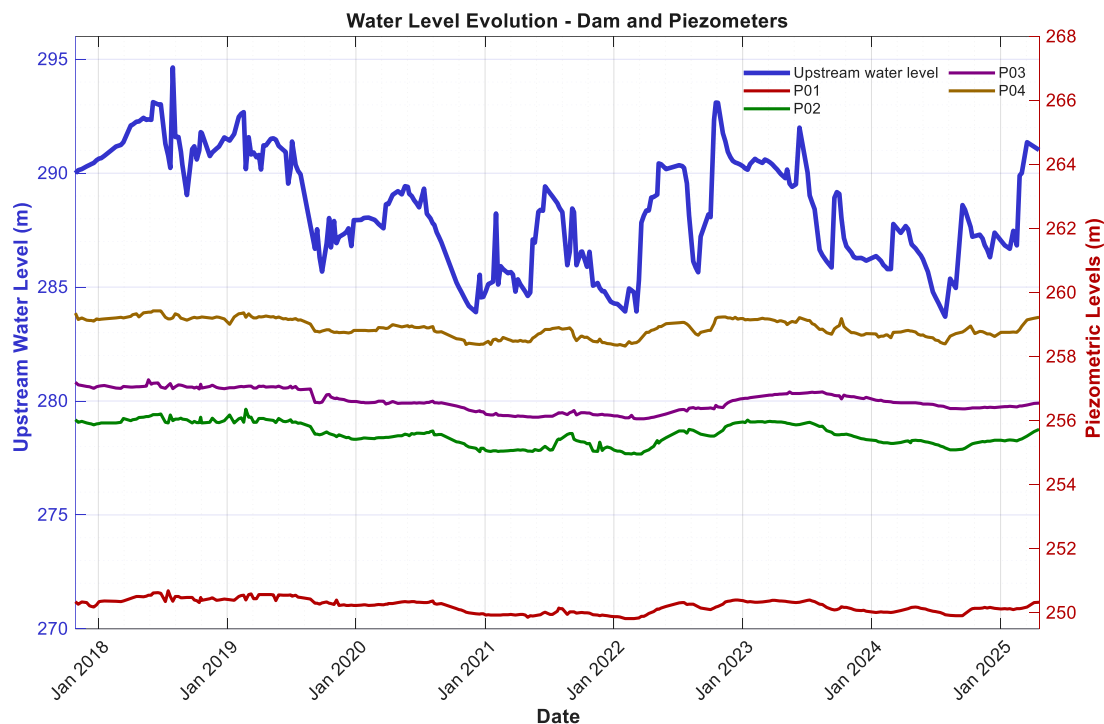


Figure 3: Time series of upstream water level and piezometric levels at P01, P02, P03, and P04 at Bouhanifia Dam from October 2017 to April 2025.

Slika 3: Časovni niz gladine vode gorvodno in piezometričnih gladin na P01, P02, P03 in P04 pri jezu Bouhanifia od oktobra 2017 do aprila 2025.



Figure 4: Temporal variation of temperature (°C) and precipitation (mm) at Bouhanifia Dam from October 2017 to April 2025.

Slika 4: Časovno spreminjanje temperature (°C) in padavin (mm) na jezu Bouhanifia od oktobra 2017 do aprila 2025.

2.3 Data Preprocessing

The raw dataset comprised 469 weekly observations (2017-2025). Data preprocessing included the following steps:

1. Missing data treatment: Missing values (approximately 2.1% of manual in situ measurements) were imputed using linear interpolation for gaps ≤ 2 weeks and seasonal decomposition for longer gaps.
2. Outlier detection: Extreme values were identified using the interquartile range method and validated against measurement logs; confirmed outliers (0.8% of manual in situ measurements) were replaced with median values from adjacent weeks.
3. Normalization: Input variables were normalized using min-max scaling to a range of $[-1, 1]$ to enhance model convergence and prediction stability:

$$a_i = \frac{h_i - h_{\min}}{h_{\max} - h_{\min}} \times (f_{\max} - f_{\min}) + f_{\min} \quad (1),$$

Where a_i is the normalized value, h_i is the measured predictor variable, $h_{\min}$ and $h_{\max}$ are its minimum and maximum values, and $f_{\max}$ and $f_{\min}$ are set to 1 and -1, respectively (Mueller & Hemond 2013).

To forecast piezometric levels, three complementary machine learning approaches were implemented and compared: the Group Method of Data Handling (GMDH), Long Short-Term Memory networks (LSTM), and eXtreme Gradient Boosting (XGBoost). Each method offers distinct advantages for capturing the nonlinear, sequential, and multivariate nature of hydro-geotechnical processes in embankment dams. A schematic overview of the proposed methodology is provided in Figure 5.

3. Methodology

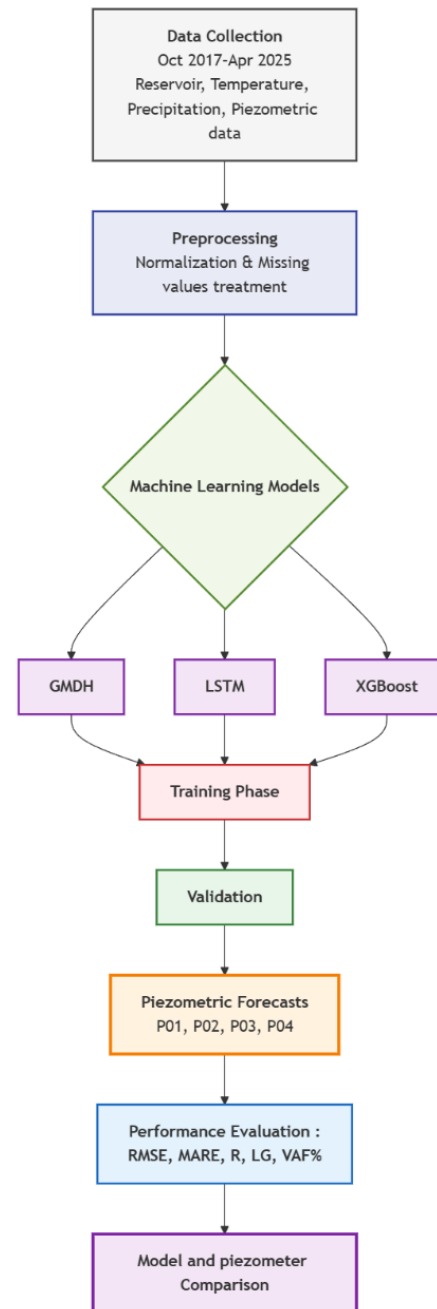


Figure 5: Schematic of the proposed methodology.

Slika 5: Shematični prikaz predlagane metodologije.

3.1 Group Method of Data Handling (GMDH)

The Group Method of Data Handling is a self-organizing inductive modeling algorithm that constructs a hierarchical polynomial network to identify complex input-output relationships (Anastasakis & Mort 2001; Kumi-Boateng & Ziggah, 2020; Shen et al., 2019; Yazdi et al., 2025; Ziggah et al., 2022). For piezometric forecasting, GMDH is particularly effective in modeling interactions between multiple influencing variables such as dam water level, historical piezometric levels, and climatic factors without requiring a predefined functional form (Yazdi et al., 2025).

The algorithm is based on the Kolmogorov–Gabor polynomial (Anastasakis & Mort 2001):

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (2),$$

where $x=(x_1, x_2, \dots, x_m)$ is the input variable vector, y is the output variable, m is the number of input variables, and $a_0, a_i, a_{ij}, a_{ijk}, \dots$ are coefficients determined during training.

Each neuron in the GMDH network computes a bivariate quadratic transfer function:

$$G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \quad (3),$$

The network grows layer by layer, with outputs from one layer serving as inputs to the next. The self-organizing structure allows the model to automatically select the most relevant variables and polynomial terms, optimizing architecture based on predictive performance (Amiri-Doumari et al., 2022).

The GMDH model's optimization relies on the automatic and inductive selection of structural hyperparameters that govern the growth of its hierarchical polynomial network. For the transfer function between layers, the polynomial order was configured to values of 1 or 2. To stabilize training and prevent overfitting, the regularization parameter was tested among the values 0.01, 0.1, and 1. Feature selection was capped at a maximum of 100 input variables (max = 100), allowing the

self-organizing architecture to select only the most significant polynomial terms for the prediction.

For this model, the input features include the reservoir water level at lags specific to each piezometer, as well as climatic variables. For piezometer P03, the reservoir level at $t-1$, $t-2$, and $t-3$ is used, while for P02, P04, and P01, the optimal lags are $t-2$ to $t-4$, $t-3$ to $t-5$, and $t-5$ to $t-7$, respectively. Precipitation is included as current precipitation at time t and cumulative precipitation over the preceding 1, 2, and 3 weeks. Temperature is included as current temperature at time t and average temperature over the preceding 1, 2, and 3 weeks.

3.2 Long Short-Term Memory Neural Network (LSTM)

Long Short-Term Memory networks are a specialized form of recurrent neural networks designed to capture long-term dependencies in sequential data (Hochreiter & Schmidhuber 1997; Kratzert et al., 2018). In piezometric forecasting, LSTM models are capable of learning temporal patterns from historical time series of water levels, piezometric data, and meteorological inputs (Waqas et al., 2024).

The LSTM cell employs three gating mechanisms – forget gate, input gate, and output gate – to regulate information flow:

$$\begin{cases} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{cases} \quad (4),$$

In these equations, x_t represents the input at time step t , and h_t is the hidden state (the short-term memory/output of the cell). The variable C_t denotes the cell state, which functions as the long-term memory of the network. The gates are defined as follows: f_t is the forget gate, controlling what information from the previous cell state C_{t-1} is discarded; i_t is the input gate, determining which new information is stored; and o_t is the output gate, deciding what information from the current cell

state C_t is output to the hidden state h_t . The term $\tilde{C}_t$ represents the candidate cell state the new information proposed to be added, $\tanh(\cdot)$ is the hyperbolic tangent. The parameters W (weight matrices) and b (bias terms) are learnable, with subscripts indicating the specific gate they belong to (e.g., W_f for the forget gate). The function σ is the sigmoid function, which compresses values between 0 and 1 to regulate information flow, and $\odot$ denotes element-wise multiplication (the Hadamard product), which is used by the gates to selectively filter information. This architecture enables the model to retain relevant historical information while adapting to new inputs, making it suitable for forecasting piezometric responses to seasonal and operational changes (Kratzert et al., 2018).

To robustly capture the long-term sequential dependencies present in the piezometric data, the architecture of the LSTM network was refined through a meticulous calibration of its hyperparameters. The model integrates a hidden layer composed of 128 neurons. Network training is governed by an initial learning rate set at 0.001. Furthermore, to control model sensitivity and prevent overfitting, dropout regularization with rates alternating between 0.4 and 0.2 was applied to the network connections during the structural adjustment phase.

The LSTM model was configured with the same input feature set as described for GMDH, including piezometer specific reservoir lags and climatic variables. The LSTM architecture enables the network to capture long-term temporal dependencies in the sequential piezometric data.

3.3 Extreme Gradient Boosting (XGBoost)

XGBoost is an ensemble learning method that builds sequential decision trees to minimize a regularized objective function (Chen & Guestrin, 2016; Kozelj & Fernández, 2025; Osman et al., 2021). It is well-suited for piezometric prediction due to its ability to model nonlinear relationships and handle heterogeneous input features, including

lagged variables and meteorological data (Bonkougou et al., 2024).

The objective function $\mathcal{L}_{xgb}$ combines a loss term and a regularization term:

$$\mathcal{L}_{xgb} = \sum_{i=1}^N L(y_i, \hat{F}(x_i)) + \sum_{m=1}^M \Omega(h_m) \quad (5),$$

where N is the number of samples; L is a differentiable loss function (e.g., mean squared error) that measures the difference between the observed and predicted values; y_i represents the observed target value (e.g., actual piezometric level) for the i -th sample; $\hat{F}(x_i)$ is the predicted value generated by the ensemble model for the i -th sample; M is the number of trees in the ensemble; and $\Omega(h_m)$ is a regularization term that penalizes the complexity of the m -th individual tree h_m to prevent overfitting (Chen & Guestrin 2016).

The complexity of each tree is further defined by the regularization term:

$$\Omega(h) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (6),$$

In this equation, T is the number of leaves in the tree, representing its structural complexity; w is the vector of leaf weights (also called leaf scores), which are the output values assigned to each leaf; γ (gamma) is the minimum loss reduction hyper parameter required to make a further partition on a leaf node, controlling tree pruning; and λ (lambda) is the L2 regularization hyperparameter on leaf weights, which helps smooth the output values and reduce sensitivity to individual observations (Chen & Guestrin 2016; Lainer & Wolfinger 2022).

XGBoost's built-in regularization reduces overfitting and improves generalization, making it a robust choice for operational forecasting in dam safety applications (Osman et al., 2021; Bonkougou et al., 2024).

The configuration of the XGBoost model requires precise tuning of its hyperparameters to balance loss function minimization and tree complexity control. The total number of decision trees was explored within a range from 100 to 500, while the maximum depth of each tree was limited to between 3 and 10 to effectively capture non-linear interactions without creating overly complex structures. The

learning rate, which weights the contribution of each new tree to the global model during the boosting process, was adjusted within an interval between 0.01 and 0.3.

In addition to its predictive capabilities, XGBoost provides built-in feature importance scores, which quantify the contribution of each input variable to the model's predictions. These scores are derived from three complementary metrics: the frequency of feature usage across all decision trees, the average gain of splits involving each feature, and the average coverage of each feature. In this study, the input variables for which feature importance will be analyzed include the reservoir water level, precipitation, and temperature. This analysis will identify the most influential variables driving piezometric variations, thereby enhancing model interpretability and providing insights into the physical processes governing groundwater flow within the dam.

The XGBoost model was configured with the same input feature set as described for GMDH, including piezometer-specific reservoir lags and climatic variables. The XGBoost hyperparameters were optimized to balance predictive accuracy and generalization.

3.4 Performance Evaluation Metrics

A comprehensive multi-metric framework was adopted to rigorously evaluate the predictive performance of each model. The following statistical indices were calculated: Root Loague Mean Square Error (RMSE), Mean Absolute Relative Error (MARE), Correlation Coefficient (R), Loague and Green's Criterion (LG), and Variance Accounted For (VAF). These metrics collectively assess accuracy, relative error, linear

association, model skill, and explanatory power (Ali & Abustan 2014).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{obs,i} - P_{pred,i})^2} \quad (7),$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_{obs,i} - P_{pred,i}}{P_{obs,i}} \right| \quad (8),$$

$$R = \frac{\sum_{i=1}^N (P_{obs,i} - \bar{P}_{obs})(P_{pred,i} - \bar{P}_{pred})}{\sqrt{\sum_{i=1}^N (P_{obs,i} - \bar{P}_{obs})^2 \sum_{i=1}^N (P_{pred,i} - \bar{P}_{pred})^2}} \quad (9),$$

$$LG = 1 - \frac{\sum_{i=1}^N |P_{obs,i} - P_{pred,i}|}{\sum_{i=1}^N |P_{obs,i} - \bar{P}_{obs}|} \quad (10),$$

$$VAF = \left[1 - \frac{\text{var}(P_{obs,i} - P_{pred,i})}{\text{var}(P_{obs,i})} \right] \times 100 \quad (11),$$

where P_{obs} , and $P_{pred,i}$ represent the observed and predicted piezometric levels, respectively; $\bar{P}_{obs}$ and $\bar{P}_{pred}$ are their means; var denotes variance; N is the number of test observations; and i ranges from 1 to N . This evaluation framework ensures a robust and multi-dimensional comparison of the GMDH, LSTM, and XGBoost models in forecasting piezometric levels.

4. Results and Discussion

4.1 Model Performance Evaluation

Model performance was examined using a standard data split, with 80% of the data allocated to model training and 20% reserved for validation (Güneş, 2025). This approach legitimately assesses the models' generalization capacity on data never encountered during training. The performance of the various predictive models developed for each piezometer is presented in detail for each method (GMDH, LSTM, and XGBoost).

The evaluation of the execution times reveals significant variations across the implemented models, stemming from the structural complexity inherent to each algorithm. When deployed on a hardware configuration featuring an Intel Core i7-8700 processor (5 logical threads) and 8 GB of RAM, the GMDH approach proves to be the fastest, completing training and inference in less than 10 seconds. This high efficiency is primarily due to the upstream feature selection phase, which restricts the network's inputs to the 100 most predictive characteristics. Despite its recurrent nature, the

LSTM model exhibits remarkably strong performance, achieving full convergence in approximately 40 seconds. This rapid execution is favored by the compact size of the dataset and the integration of an early stopping mechanism, which halts training as soon as the validation loss stabilizes. Conversely, the XGBoost algorithm demands the highest computational workload, requiring 4 minutes to complete. This extended duration is mainly driven by the iterative hyperparameter optimization process which, although parallelized to maximize CPU core utilization, imposes a critical and sustained load on the system memory.

The results summarized in Table 2 and Figures 6 and 7 show close agreement between predicted and observed values for most models, with generally low prediction residuals across the test set. Statistical indicators confirm the high precision of the predictions, with narrow confidence intervals and robust performance across different hydrological conditions. The quantitative performance metrics for each model and piezometer are presented in Table 2, while Figures 6 and 7 provide complementary visual assessments of the predictions and residuals during the validation phase.

4.2 Analysis of Results

The analysis of the results obtained for the piezometric level prediction at the Bouhanifia Dam reveals overall exceptional performance for the three artificial intelligence methods evaluated (LSTM, GMDH, and XGBoost). The correlation coefficients (R) systematically exceed 0.90 in validation for all piezometers, with maximum values reaching 0.975 for the LSTM model on piezometer P03. This remarkable performance demonstrates the ability of machine learning-based approaches to capture the complex and nonlinear relationships governing piezometric variations within the dam.

The root mean square errors (RMSE) in validation range from 0.066 m to 0.153 m depending on the models and piezometers, representing a highly satisfactory accuracy for operational monitoring

applications. The mean absolute relative errors (MARE) are also very low, ranging from 0.099% to 0.235%, indicating that the models successfully reproduce piezometric levels with exceptional fidelity. The variance accounted for (VAF) exceeds 81% for all models and piezometers, with maximum values reaching 95% for the LSTM model on piezometer P03. Figure 6 presents the time series of observed and predicted piezometric levels during the validation phase, illustrating the close agreement between predictions and observations across all piezometers and models. Figure 7 shows the corresponding residual time series, confirming the low prediction errors and the absence of systematic bias across all piezometers.

4.3 Model Comparison

The XGBoost model stands out for its superior performance across the majority of evaluation metrics. It exhibits the lowest RMSE for piezometers P01 (0.066 m) and P03 (0.105 m), as well as the lowest MARE for these same piezometers (0.099% and 0.169%, respectively). The Loague and Green coefficient (LG) of XGBoost is systematically higher than that of the other models, with validation values ranging from 0.560 to 0.631. This superiority in LG is particularly significant for dam monitoring applications, as this criterion evaluates the model's ability to provide reliable predictions with adequate safety margins.

The performance of XGBoost is explained by its capacity to effectively model nonlinear interactions between input variables through its ensemble decision tree structure. The algorithm also benefits from its built-in regularization, which reduces overfitting and improves generalization, a crucial advantage for time series exhibiting marked seasonal cycles such as those of the Bouhanifia Dam.

The analysis of feature importance scores from the XGBoost model reveals a clear hierarchy of influential variables across all piezometers. The reservoir water level at lags 1 and 2 consistently ranks first, accounting for approximately 70–80% of the predictive power, reflecting the direct hydraulic connection between the reservoir and pore

pressures within the dam body. This dominance of reservoir level is expected, as it is the primary driver of piezometric variations. This finding reinforces the physical consistency of the model, confirming that the piezometric responses are primarily governed by fluctuations in the reservoir level, with the lag structure capturing the propagation time of pressure waves through the dam body and its foundation.

The LSTM model demonstrates remarkable stability across all piezometers, with homogeneous performance and low variance between different measurement points. It produces the highest validation correlations (R) for P01 (0.965) and P04 (0.947), and the highest VAF for P02 (92.6%) and P04 (84%). LSTM particularly excels on piezometer P03, where it achieves an R of 0.975 and a VAF of 95% in validation, the best performances recorded for the entire study. The gaps between training and validation performances are remarkably small for LSTM, with an average RMSE difference of only 0.022 m, indicating excellent generalization capability and no significant overfitting.

This exceptional stability of LSTM is attributed to its ability to capture long-term temporal dependencies through its long-term memory structure. The network can thus retain relevant information over long periods, which is essential for modeling the annual hydrological cycles and interannual trends observed in the Bouhanifia Dam data. The LSTM architecture, with its gating mechanisms (forget, input, and output gates), enables fine regulation of information flow, ensuring remarkable stability even in the face of pronounced seasonal fluctuations.

The GMDH model, although generally ranking third, offers entirely acceptable performances with VAF systematically exceeding 91.9% in training and surpassing 81.1% in validation. It presents a definite advantage in terms of computation time, executing in less than 10 seconds on a standard hardware configuration (Intel Core i7-8700 processor), compared to approximately 40 seconds for LSTM and 4 minutes for XGBoost. This rapid execution is due to the upstream feature selection

phase, which limits network inputs to the 100 most predictive characteristics.

The GMDH method, based on a hierarchical construction of polynomial networks, offers a good compromise between accuracy and interpretability. Its self-organizing structure allows for automatic selection of the most relevant variables and identification of significant polynomial relationships, facilitating the physical interpretation of results. However, GMDH exhibits the largest gaps between training and validation performances, particularly for P03 (validation RMSE = 0.105 m vs 0.059 m in training) and P04 (0.149 m vs 0.052 m), suggesting that the simplicity of its structure may limit its generalization capability on complex dynamics.

4.4 Spatial Variability of Performance

The analysis of performance by piezometer reveals significant spatial variations that provide important insights into the dam's hydraulic behavior.

Piezometer P03 exhibits the best performance across all models, with R values exceeding 0.975 for LSTM and VAF reaching 95% in validation. This exceptional performance can be explained by several factors: P03 shows the strongest correlation with the reservoir level ($r = 0.83$), the shortest response time (1 to 2 weeks), and a regular and well-defined dynamics. Its position, probably close to the upstream concrete face in a permeable zone, offers ideal conditions for predictive modeling. LSTM provides the best performance on this piezometer (RMSE = 0.068 m, $R = 0.975$, LG = 0.709), demonstrating the superiority of recurrent neural networks for modeling rapid and well-defined dynamics.

P02 presents the most balanced and highest performance in terms of VAF (92.6% for LSTM). The RMSE values are low (0.077 to 0.095 m) and the R values are excellent (0.951 to 0.970). LSTM particularly stands out on this piezometer with an R of 0.970 and a VAF of 92.6%, confirming the network's ability to capture intermediate dynamics between fast zones (P03) and slow zones (P01). The gaps between training and validation are also the smallest for this piezometer, indicating particularly well modelable dynamics.

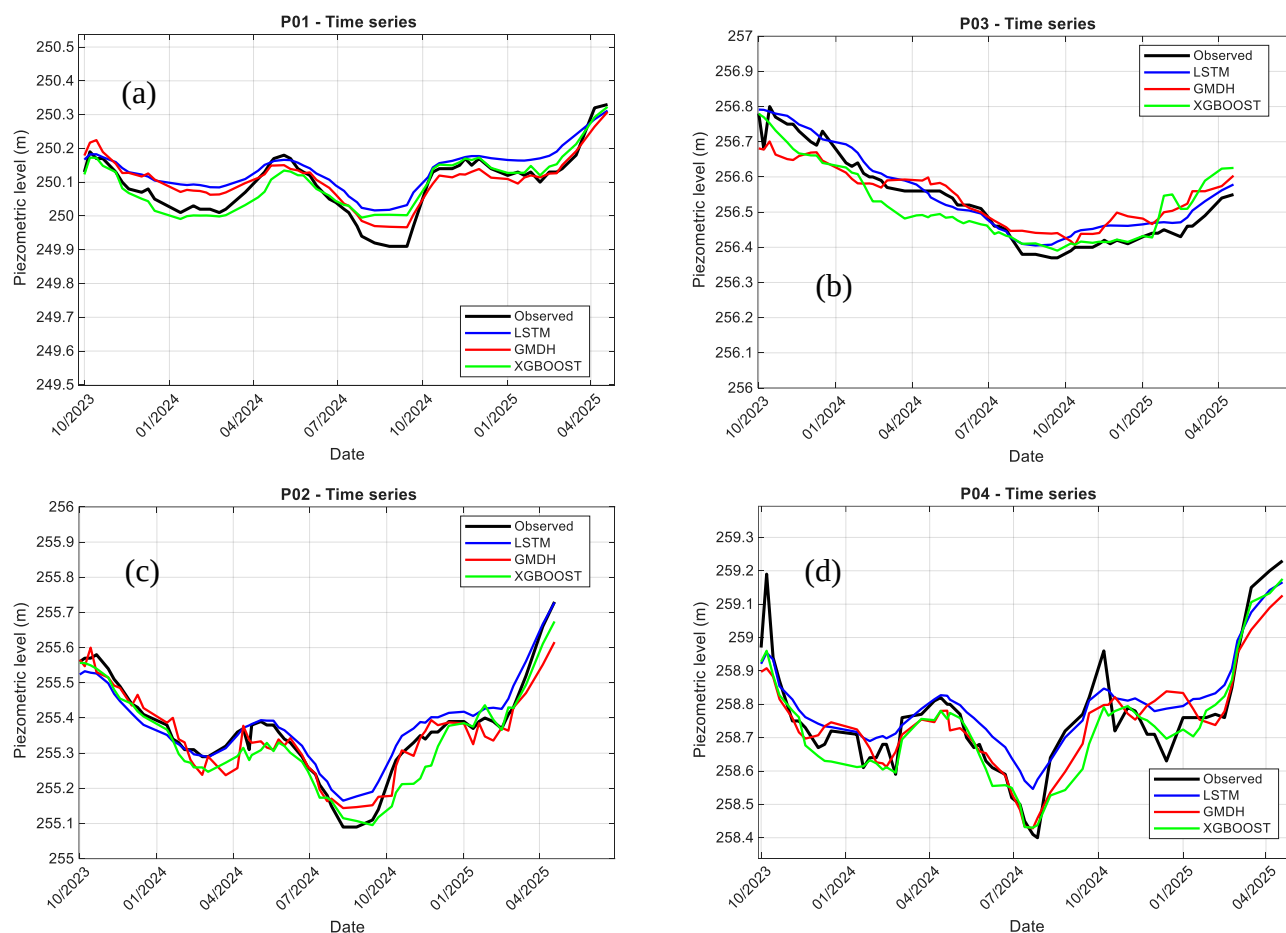


Figure 6: Time series of observed and predicted piezometric levels during the validation phase at piezometers P01–P04 for LSTM (blue), GMDH (red), and XGBoost (green) models, with observed values shown in black. Subfigures: (a) P01, (b) P02, (c) P03, (d) P04.

Slika 6: Časovne vrste opazovanih in napovedanih piezometričnih ravni v validacijski fazi pri piezometrih P01–P04 za modele LSTM (modra), GMDH (rdeča) in XGBoost (zelena), z opazovanimi vrednostmi v črni barvi. Pod slike: (a) P01, (b) P02, (c) P03, (d) P04.

P04 presents the most variable performance, with RMSE ranging from 0.137 m (XGBoost) to 0.153 m (LSTM). R values range from 0.905 (GMDH) to 0.947 (LSTM), and VAF from 81.1% to 87%. The slightly lower performance on this piezometer is explained by its large amplitude of variation (4.26 m) and its position in the downstream area, where dynamics are more complex and influenced by a multitude of factors (drainage, geological heterogeneity, etc.). XGBoost offers the best performance on this piezometer (RMSE = 0.137 m, LG = 0.563), suggesting that the ensemble tree structure is better adapted to the complex dynamics of the downstream area.

4.5 Analysis of Training-Validation Gaps

The analysis of the gaps between training and validation performances provides valuable insights into the generalization capability of the models. LSTM stands out as the model offering the best compromise between accuracy and generalization, with minimal gaps between training and validation phases, including an average RMSE increase of only 0.022 m and an average VAF decrease of 3.6%. This characteristic is particularly important for operational applications where prediction reliability on new data is crucial.

XGBoost, despite its superior training accuracy, shows larger gaps, suggesting slight overfitting. The average RMSE increase is 0.036 m with a VAF decrease of 8.3%. This phenomenon is particularly pronounced for P03, where RMSE increases from 0.044 m in training to 0.105 m in validation, representing an increase of 138%. This sensitivity to overfitting must be taken into account during operational deployment.

GMDH also exhibits significant gaps, with an average RMSE increase of 0.031 m and a VAF decrease of 7.4%, particularly for P03 (0.105 vs 0.059) and P04 (0.149 vs 0.052), indicating that the simplicity of its structure may limit its generalization capability on more complex dynamics. However, its performance remains acceptable for non-critical applications.

4.6 Analysis of Validation Time Series

Examination of the validation time series (Figure 6) and the corresponding residual time series (Figure 7) reveals several important observations. Hydrological transition periods, particularly late summer and early autumn as well as spring, show larger gaps between predictions and observations, reflected in higher residual values. For example, in October 2023, all three models exhibit more pronounced deviations, coinciding with the transition period between the dry and wet seasons. This observation suggests that the models capture general trends well but struggle slightly during rapid transitions.

The models correctly reproduce the declining trend in piezometric levels observed over the validation period, as evidenced by the low residual values

As shown in Figure 7, the predictions follow the general trend of the observations, indicating good capture of interannual evolutions. Systematic biases are observed in some models: LSTM tends to slightly overestimate low levels, particularly in 2024–2025, while GMDH sometimes underestimates peaks. These biases, although small, could be corrected by post-processing techniques. Performance during the January–April 2025 period is particularly good, with very small gaps between predictions and observations, as visible in

both Figures 6 and 7. This suggests that the models generalize well on recent data not seen during training.

4.7 Synthesis and Recommendations

4.7.1 Model Hierarchy by Piezometer

For piezometer P01, XGBoost provides the best performance with the lowest RMSE of 0.066 m and the highest LG of 0.631. For piezometer P02, LSTM achieves the highest R of 0.970 and the highest VAF of 92.6%. For piezometer P03, LSTM again excels with the highest R of 0.975 and the highest VAF of 95%. For piezometer P04, XGBoost offers the lowest RMSE of 0.137 m and the highest LG of 0.563.

4.7.2 Operational Recommendations

For real-time monitoring, the XGBoost model is recommended, as it offers the best accuracy and the highest safety margins (LG). Its ability to provide reliable predictions with low RMSE makes it a valuable tool for early anomaly detection.

For applications requiring long-term robustness, the LSTM model is preferable, as it offers remarkable stability across all piezometers and excellent generalization, with limited gaps between training and validation performances. Its ability to capture long-term temporal dependencies makes it particularly suited for continuous monitoring applications.

For preliminary analyses or resource-constrained applications, the GMDH model can be used as it offers a good balance between accuracy and computational simplicity. Its fast execution time of less than 10 seconds makes it good for exploratory analyses and frequent model updates.

A hybrid approach combining the three models is recommended for optimal monitoring, allowing it to benefit from XGBoost's accuracy, LSTM's stability, and GMDH's speed, while providing redundancy that reinforces the monitoring system's reliability. An alert system based on consensus among the three models, triggering when two or more models detect an anomaly, would provide increased robustness for dam safety.

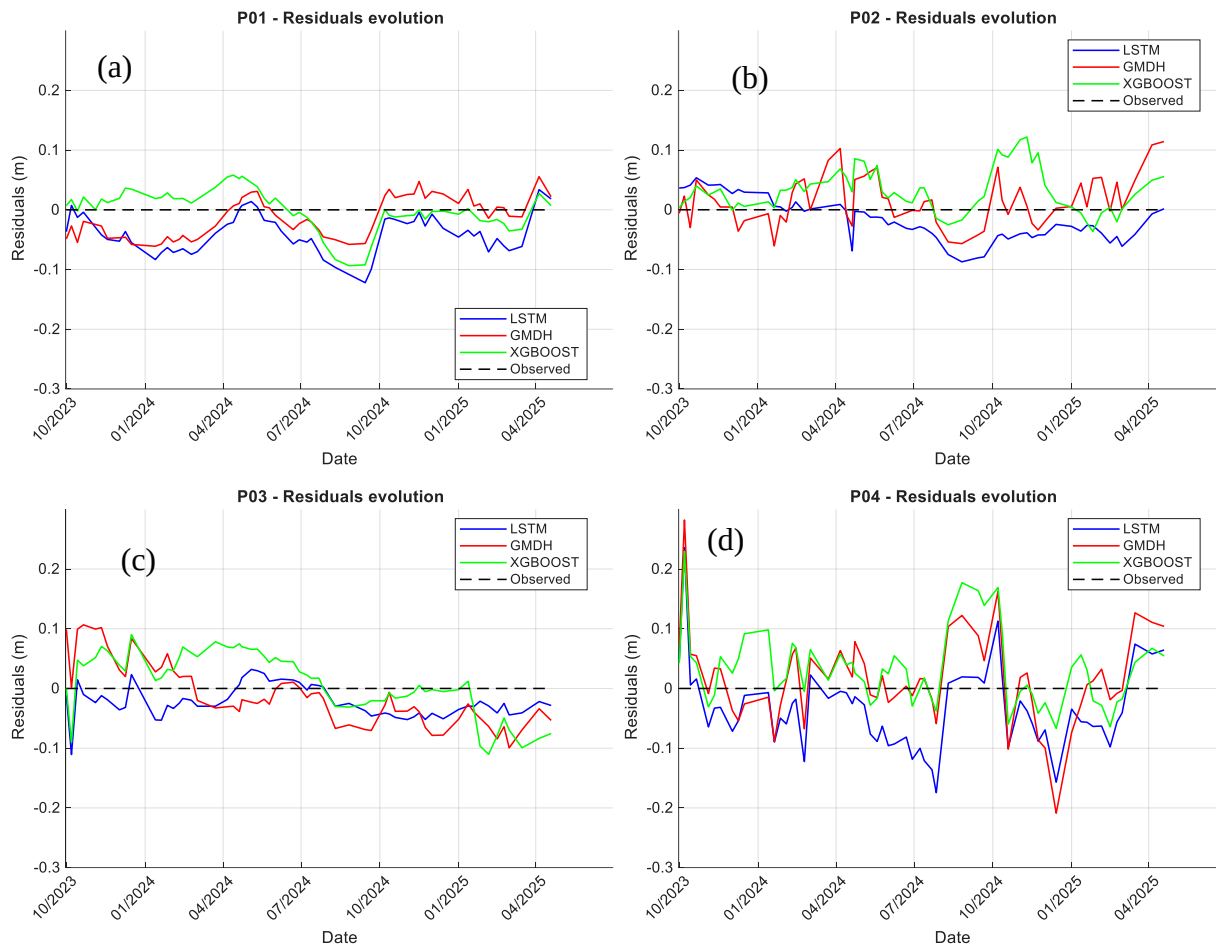


Figure 7: Residual time series (observed – predicted) during validation at piezometers P01–P04 for the LSTM (blue), GMDH (red), and XGBoost (green) models. Subfigures: (a) P01, (b) P02, (c) P03, (d) P04.

Slika 7: Časovne vrste ostankov (opazovano – napovedano) med validacijo pri piezometrih P01–P04 za modele LSTM (modra), GMDH (rdeča) in XGBoost (zelena). Pod slike: (a) P01, (b) P02, (c) P03, (d) P04.

4.8 Analysis of Computational Time and Operational Efficiency

Computational times vary considerably between models, which has important implications for operational deployment. GMDH, with an execution time of less than 10 seconds, is particularly suitable for applications requiring frequent updates or limited computational resources. LSTM, with a time of approximately 40 seconds, offers an excellent compromise between accuracy and computational efficiency, thanks to the integration of an early stopping mechanism that optimizes the training process. XGBoost, although the most accurate, requires approximately 4 minutes of training due to iterative hyperparameter optimization, making it better for offline applications or less frequent updates.

4.9 Physical Interpretation of Results

The results obtained confirm the validity of the hydraulic assumptions underlying the modeling. The hierarchy of performance by piezometer, with P03 showing the best performance, followed by P02, P01, and P04, directly reflects the classification based on response times and correlations with the reservoir level. P03, with its response time of 1 to 2 weeks and correlation of 0.83, indeed exhibits the most modelable dynamics. P02, with its intermediate response of 2 to 3 weeks, offers balanced performance. P01, with its response time of 4 to 6 weeks, exhibits more attenuated but still modelable dynamics. Finally, P04, with its large amplitude and downstream position, presents the most significant modeling challenges.

Piezometer	Method	RMSE		MARE(%)		R		LG		VAF(%)	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation	Training	Validation
P01	LSTM	0.061	0.074	0.051	0.144	0.98	0.965	0.766	0.522	92.7	85.2
	GMDH	0.064	0.070	0.061	0.121	0.972	0.939	0.748	0.547	91.9	85.2
	XGBoost	0.055	0.066	0.039	0.099	0.97	0.921	0.808	0.631	93.7	84.7
P02	LSTM	0.074	0.077	0.057	0.125	0.98	0.97	0.829	0.65	95.4	92.6
	GMDH	0.082	0.085	0.052	0.126	0.973	0.951	0.796	0.648	94.5	90.2
	XGBoost	0.079	0.095	0.059	0.144	0.973	0.962	0.834	0.596	94.7	92.3
P03	LSTM	0.057	0.068	0.053	0.112	0.986	0.975	0.819	0.709	96.7	95
	GMDH	0.059	0.105	0.053	0.169	0.984	0.946	0.806	0.561	96.5	81.5
	XGBoost	0.044	0.105	0.034	0.169	0.99	0.903	0.865	0.56	98	81.4
P04	LSTM	0.046	0.153	0.036	0.235	0.987	0.947	0.77	0.471	97.2	84
	GMDH	0.052	0.149	0.04	0.209	0.982	0.905	0.816	0.531	96.4	81.1
	XGBoost	0.052	0.137	0.053	0.194	0.983	0.933	0.84	0.563	96.7	87

Table 2: Comparative performance evaluation of LSTM, GMDH, and XGBoost models for piezometers P01, P02, P03, and P04 using five performance metrics (RMSE, MARE, R, LG, and VAF) calculated for both training and validation phases.

Preglednica 2: Primerjalna ocena učinkovitosti modelov LSTM, GMDH in XGBoost za piezometre P01, P02, P03 in P04 z uporabo petih metrik uspešnosti (RMSE, MARE, R, LG in VAF), izračunanih za fazi usposabljanja in validacije.

4.10 Perspectives and Future Work

The promising results obtained in this study open several perspectives for future work. Extending the analysis to other dams in Algeria would allow us to validate the generalizability of the developed approaches and identify possible specificities related to local geological and climatic conditions. The integration of additional data, such as seepage flow rates, settlements, or seismic data, could further improve prediction accuracy and broaden the scope of model applications.

Finally, the development of early warning systems based on these models, with alarm thresholds defined from prediction confidence intervals, would constitute a major advancement for dam safety.

5. Conclusion

This study evaluated the performance of three artificial intelligence methods, namely Long Short-Term Memory (LSTM), Group Method of Data Handling (GMDH), and Extreme Gradient Boosting

(XGBoost) for predicting the piezometric level of the Bouhanifia Dam, with particular emphasis on their predictive accuracy, generalization capability, and operational reliability.

The results demonstrate that all three methods yield competitive and highly accurate predictions, confirming their practical applicability for dam safety monitoring. Correlation coefficients consistently exceeding 0.90 in validation and variance explained above 81.1% attest to the robustness of AI-based approaches for modeling complex hydrological systems. Validation RMSE values ranging from 0.066 m (XGBoost, P01) to 0.153 m (LSTM, P04) represent a highly satisfactory accuracy for operational monitoring applications, with relative errors below 0.24%.

The XGBoost model distinguished itself through systematically superior performance on piezometers P01 and P04, with the lowest RMSE (0.066 m and 0.137 m, respectively) and the highest LG (0.631 and 0.563). Its exceptional robustness is attributed to its ensemble learning architecture,

which effectively handles nonlinear relationships and captures feature interactions. However, XGBoost shows some sensitivity to overfitting, with RMSE increasing by up to 138% between training and validation for P03.

The LSTM model proved to be the most stable and consistent across all monitoring points, achieving the highest R for P01 (0.965), P03 (0.975), and P04 (0.947), and the highest VAF for P02 (92.6%) and P04 (84%). With an average RMSE gap of only 0.022 m between training and validation, LSTM demonstrates excellent generalization capability, making it particularly well suited for continuous, long-term monitoring applications.

The GMDH model, while ranking third, offers a computationally efficient alternative with execution times under 10 seconds, achieving VAF exceeding 81.1% in validation. Its self-organizing polynomial structure facilitates the physical interpretation of results, making it suitable for preliminary assessments and resource-constrained applications.

Spatial analysis reveals a clear hierarchy in predictive performance: P03 (fastest response, 1-2 weeks) > P02 (intermediate, 2-3 weeks) > P01 (slow, 4-6 weeks) > P04 (complex downstream dynamics). This hierarchy directly reflects the hydraulic response times and correlations with the reservoir level, validating the physical consistency of the modeling approach.

A hybrid multi-model approach combining the three methods is recommended for optimal monitoring, leveraging XGBoost's accuracy, LSTM's stability, and GMDH's speed, while providing redundancy through consensus-based alert systems. Future work should focus on validating these methods on additional dam sites, exploring transfer learning capabilities, and integrating these predictive models into real-time early warning systems for enhanced infrastructure resilience.

Acknowledgments

The authors would like to express their gratitude to the staff of the National Dams and Water Transfers Agency (ANBT) for their support and for providing the data necessary for this research.

CRedit authorship contribution statement

Conceptualization: B.S.; Methodology: B.S., B.R., S.N., and T.F.; Software: B.S.; Validation: B.S. and T.F.; Formal analysis: B.S.; Investigation: B.S.; Resources: B.S.; Data curation: B.S.; Writing – original draft: B.S. and T.F.; Writing – review & editing: B.S., B.R., S.N., and T.F.; Visualization: B.S.; Supervision: B.S., B.R., S.N., and T.F.; All authors have read and agreed to the published version of the manuscript.

Data availability

The data that support the findings of this study are available from ANBT, which were used under license for the current research.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

References

- Anastasakis, L., Mort, N. (2001). The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH). Research Report – University of Sheffield Department of Automatic Control and Systems Engineering. [Research Report](#)
- Ali, M., H., Abustan, I. (2014). A new novel index for evaluating model performance. JNRD-Journal of Natural Resources and Development, 4, 1-9. <https://doi.org/10.5027/jnrd.v4i0.01>
- Agence Nationale des Barrages et Transferts (ANBT): Caractéristiques techniques de barrage Bouhanifia. Technical report, Algeria (2025).
- Amiri-Doumari, S., Karimipour, A., Nayeypour, S., N., Hatamiafkoueieh, J. (2023). Integration of group method of data handling (GMDH) algorithm and population-based metaheuristic algorithms for spatial prediction of potential groundwater. *Environmental Earth Sciences*, **81**(20), 485. <https://doi.org/10.1007/s12665-023-10931-1>
- Bonkounou, A., A., Zio, S., Sabane, A., Kafando, R., Kabore, A. K., Bissyande, T. F. (2024, June). A comparison of AI methods for Groundwater Level Prediction in Burkina Faso. In IFIP International Conference on Artificial Intelligence Applications and

- Innovations (pp. 3-16). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63219-8_1
- Cao, W., Wu, X., Li, J., Kang, F. (2025). A review of artificial intelligence in dam engineering. *Journal of Infrastructure Intelligence and Resilience*, **4**(1), 100122. <https://doi.org/10.1016/j.iintel.2024.100122>
- Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural Water Management*, **240**, 106272. <https://doi.org/10.1016/j.agwat.2019.106272>
- Güneş Şen, S. (2025). Machine Learning-Based Water Level Forecast in a Dam Reservoir: A Case Study of Karaçomak Dam in the Kızılırmak Basin, Türkiye. *Sustainability*, **17**(18), 8378. <https://doi.org/10.3390/su17188378>
- Harbi, L., Smail, N., Rouissat, B., Charrak, H. (2024). Assessing single and hybrid AI approaches in conjunction with FEM to enhance seepage prediction in earth dams. *Modeling Earth Systems and Environment*, **10**(2), 2421-2433. <https://doi.org/10.1007/s40808-023-01903-2>
- Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, **9**(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Khorchani, M., Rhayma, N., Pereira, S., Riahi, H. (2025). Exploring the role of artificial intelligence in predicting piezometric levels in homogeneous earth-fill dams. *Modeling Earth Systems and Environment*, **11**(1), 33. <https://doi.org/10.1007/s40808-024-02196-9>
- Kozelj, D., Fernández, D., A. (2025). Predicting water distribution pipe failures using machine learning and cross-infrastructure data. *Acta hydrotechnica*, **38**(68), 53-64. <https://doi.org/10.15292/acta.hydro.2025.05>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, **22**(11), 6005-6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Kumi-Boateng, B., Ziggah, Y., Y. (2020). Feasibility of using Group Method of Data Handling (GMDH) approach for horizontal coordinate transformation. *Geodesy and Cartography*, **46**(2), 55-66. <https://doi.org/10.3846/gac.2020.10486>
- Lainder, A., D., Wolfinger, R., D. (2022). Forecasting with gradient boosted trees: augmentation, tuning, and cross-validation strategies: Winning solution to the M5 Uncertainty competition. *International Journal of Forecasting*, **38**(4), 1426-1433. <https://doi.org/10.1016/j.ijforecast.2021.12.003>
- Liao, W., Zhang, Z., Liu, B., Lu, X., Liu, D., Liu, Q., Liu, C. (2025). Intelligent zoning design of concrete-faced rockfill dams using image-parameter fusion enhanced generative adversarial networks. *Engineering Structures*, **339**, 120662. <https://doi.org/10.1016/j.engstruct.2025.120662>
- Mueller, A., V., Hemond, H., F. (2013). Extended artificial neural networks: incorporation of a priori chemical knowledge enables use of ion selective electrodes for in-situ measurement of ions at environmentally relevant levels. *Talanta*, **117**, 112-118. <https://doi.org/10.1016/j.talanta.2013.08.045>
- Ni, L., Wang, D., Singh, V., P., Wu, J., Wang, Y., Tao, Y. (2021). Streamflow forecasting using extreme gradient boosting and long short-term memory coupled with wavelet transform. *Journal of Hydrology*, **595**, 126033. <https://doi.org/10.1016/j.jhydrol.2020.124901>
- Özbey, V., Ergintav, S., Tari, E. (2024). GNSS time series analysis with machine learning algorithms: a case study for Anatolia. *Remote Sensing*, **16**(17), 3309. <https://doi.org/10.3390/rs16173309>
- Osman, A., I., A., Ahmed, A., N., Chow, M., F., Huang, Y., F., El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, **12**(2), 1545-1556. <https://doi.org/10.1016/j.asej.2020.11.011>
- Ranković, V., Novaković, A., Grujović, N., Divac, D., Milivojević, N. (2014). Predicting piezometric water level in dams via artificial neural networks. *Neural Computing and Applications*, **24**(5), 1115-1121. <https://doi.org/10.1007/s00521-012-1334-2>
- Sharghi, E., Nourani, V., Behfar, N., Tayfur, G. (2019). Data pre-post processing methods in AI-based modeling of seepage through earthen dams. *Measurement*, **147**, 106820. <https://doi.org/10.1016/j.measurement.2019.07.048>
- Shen, C., Asante-Okyere, S., Yevenyo Ziggah, Y., Wang, L., Zhu, X. (2019). Group method of data handling (GMDH) lithology identification based on wavelet analysis and dimensionality reduction as well log data pre-processing techniques. *Energies*, **12**(8), 1509. <https://doi.org/10.3390/en12081509>
- Waqas, M., Humphries, U., W., Hlaing, P., T., Ahmad, S. (2024). Seasonal WaveNet-LSTM: a deep learning framework for precipitation forecasting with integrated

large scale climate drivers. *Water*, 16(22), 3194. <https://doi.org/10.3390/w16223194>

Xie, Z., Chen, L., Li, Y. (2026). Optimization of dam safety monitoring models based on residual compensation: A multidimensional temporal framework integrating TCN-BiLSTM-attention mechanism. *Engineering Structures*, **348**, 121869. <https://doi.org/10.1016/j.engstruct.2025.121869>

Yazdi, S., H., Robati, M., Samani, S., Hargalani, F. Z. (2025). Prediction of two groundwater sustainability indicators in semi-arid aquifers using machine learning. *Environmental Earth Sciences*, **84(11)**, 294. <https://doi.org/10.1007/s12665-025-12253-w>

Zheng, C., Cen, W., Liu, B., Qian, J., Ding, Y., Mo, C. (2025). Hybrid optimization and AI-driven surrogate model for seepage parameters inversion in complex dam foundations. *Journal of Hydrology*, **134484**. <https://doi.org/10.1016/j.jhydrol.2025.134484>

Zhou, Y., Zhang, Y., Pang, R., Xu, B. (2021). Seismic fragility analysis of high concrete faced rockfill dams based on plastic failure with support vector machine. *Soil Dynamics and Earthquake Engineering*, **144**, 106587. <https://doi.org/10.1016/j.soildyn.2021.106587>

Ziggah, Y., Y., Issaka, Y., Laari, P., B. (2022). Evaluation of different artificial intelligent methods for predicting dam piezometric water level. *Modeling Earth Systems and Environment*, **8(2)**, 2715-2731. <https://doi.org/10.1007/s40808-021-01263-9>