Water Resources Management (2025) 39:5415–5442
https://doi.org/10.1007/s11269-025-04210-w

Improving the Accuracy of Groundwater Level Forecasting by Coupling Ensemble Machine Learning Model and Coronavirus Herd Immunity Optimizer

Ahmed M. Saqr[1] · **Veysi Kartal**[2] · **Erkan Karakoyun**[3] · **Mahmoud E. Abd-Elmaboud**[1]

Received: 19 January 2025 / Accepted: 2 April 2025 / Published online: 3 May 2025

Abstract

Groundwater levels are under severe pressure globally due to over-extraction, pollution, and climate change necessitating continuous monitoring for sustainable aquifer management. This study introduces a novel ensemble machine learning (En) model that integrates shallow and deep machine learning (ML) models, optimized through the coronavirus herd immunity optimizer (CHIO), for accurate groundwater level forecasting. This En model was applied to the Ergene River Basin, Türkiye, a region facing severe groundwater depletion and contamination due to intensive agricultural and industrial activities. Groundwater level data spanning 1966 to 2023 on a weekly basis from four wells were used, split into 70% for training and 30% for testing under short- and long-term scenarios. Using the partial autocorrelation function and gamma test the best lag numbers were determined for input data, reflecting aquifer heterogeneity. Score analysis, supported by statistical metrics such as the coefficient of determination (R^2) and root mean square error (RMSE), was employed alongside visual aids to assess the developed En model performance. Results demonstrated that deep ML models outperformed shallow ML models achieving $R^2 \sim 0.99$ and RMSE ~0.5 m. The developed En model outperformed all individual ML models, with score values exceeding 200, and its predictions closely aligned with measured water levels during both testing phases. The findings underscored the developed En model's contribution to achieving sustainable development goals (SDGs) by enhancing water-use efficiency and addressing environmental, economic, and social sustainability challenges. The proposed approach offers a reliable and adaptable solution for groundwater level forecasting, applicable to other aquifers worldwide.

Highlights

- A novel ensemble model integrates shallow and deep learning for groundwater forecasting
- Input lag selection using the partial autocorrelation function test enhances predictive accuracy
- Deep learning models outperform shallow models with $R^2 \sim 0.99$ at different wells

Extended author information available on the last page of the article

- The ensemble model surpasses all individual learning models with score values over 200
- Contribution to SDG 6 and 13 by promoting sustainable groundwater management

Keywords Groundwater level forecasting · Time series analysis · Deep learning · Coronavirus herd immunity optimizer · Ensemble model · Sustainable development goals

1 Introduction

Groundwater is a vital resource essential for sustaining ecosystems, agriculture, and human livelihoods (Saqr et al. 2025). It is a critical water source for domestic, industrial, and irrigation use, especially in areas with limited or unreliable surface water (Jafarzadeh et al. 2021). Groundwater also enhances drought resilience, supports ecosystems, and ensures food security. However, its availability and quality are under growing global threat (Gezici et al. 2024). Over-extraction from agriculture, industrialization, and urbanization has led to aquifer depletion, declining water tables, and reduced storage (Rehman et al. 2024). Climate change further disrupts recharge, increases evaporation, and intensifies droughts (Subbarayan et al. 2025). Pollution from industry, agriculture, and waste disposal worsens water quality, posing health and ecological risks (Paudel et al. 2024). These issues demand continuous monitoring and predictive management frameworks to support thetargets of sustainable development goals (SDGs) (Sathiyamurthi et al. 2025).

Historically, groundwater has been monitored using physical and numerical models (Jafarzadeh et al. 2022; Masria et al. 2024). Physical models can simulate flow based on hydrological principles, while numerical tools like MODFLOW solve mathematical equations to forecast aquifer behavior (Saqr et al. 2025). Though valuable, these methods require extensive data and computing resources and often rely on linear assumptions, limiting their ability to capture complex groundwater behavior (Roy and Datta 2018). To complement these established approaches, machine learning (ML) techniques have emerged as flexible and efficient tools that can work alongside physical models, offering different strengths in handling large datasets and capturing non-linear relationships (Al-Betar et al. 2021; Ezzeldin and Abd-Elmaboud 2024). ML models can handle large datasets, identify nonlinear relationships, and provide accurate forecasts even with incomplete or noisy data.

Shallow ML models such as radial basis function (RBF), adaptive neuro-fuzzy inference system (ANFIS), and artificial neural networks (ANN) are reliable and computationally efficient (Deulkar et al. 2025). Deep ML models, including recurrent neural networks (RNN), long short-term memory (LSTM), and convolutional neural networks (CNN), capture spatial and temporal patterns more effectively (Coutinho et al. 2025). Still, standalone models often fail to fully represent groundwater system complexity. Integrated models combining shallow and deep learning offer higher accuracy and robustness (Jackson et al. 2016).

Optimization methods further enhance prediction performance. Among these, the coronavirus herd immunity optimizer (CHIO), introduced in 2021, has proven effective (Al-Betar et al. 2021). Inspired by herd immunity dynamics, CHIO balances exploration and exploitation, avoids premature convergence, reduces costs, and delivers robust results. It outperforms traditional optimizers like genetic algorithms and particle swarm optimiza-

tion in hydrology, renewable energy, and environmental modeling. For example, CHIO was applied to hybrid renewable energy system sizing (Farouk et al. 2024). However, CHIO has not yet been applied to ensemble machine learning (En) models for groundwater forecasting. Addressing this gap could significantly enhance predictive capacity and support sustainable water resource management (Alsulamy et al. 2025).

This study addresses a key research gap by developing a novel En model that combines shallow and deep learning techniques with CHIO optimization for groundwater level prediction. Shallow models (RBF, ANFIS, ANN) were selected for their interpretability, while deep models (RNN, LSTM, CNN) were chosen for their ability to capture long-term dependencies and spatial features in hydrological data. The framework was applied to the Ergene River Basin, Türkiye, an area facing groundwater challenges from over-extraction, industrial pollution, and climate variability. Four observation wells were analyzed to capture aquifer heterogeneity and spatial variation. The study also evaluates the model's contribution to SDGs by addressing environmental, economic, and social sustainability pillars. Beyond regional focus, the framework can present a scalable, adaptable tool for groundwater management in similar settings, supporting global water security and long-term resilience.

2 Materials and Methods

2.1 Study Area and Data Collection

The Ergene River Basin, located in the Eastern Thrace region of northwest Türkiye, spans an area of approximately 11,000 km² and serves as a critical resource for agricultural irrigation, domestic consumption, and industrial activities (Tokatlı and Varol 2021). Geographically, the basin lies in (40°45′*N* − 42°05′N) latitude, and (26°05′E −28°10′E) longitude, bordered by the Istranca Mountains to the north, the Aegean Sea to the south, and the Maritsa River to the west (Fig. 1). Over recent decades, the basin has experienced significant groundwater level deterioration due to over-extraction, rapid industrialization and climatic variability, making it an ideal location for groundwater level prediction studies to guide sustainable water resource management (Arkoç 2020). Known for its agricultural productivity and industrial activity, the basin has experienced substantial groundwater declines, primarily driven by unregulated extraction and contamination from industrial discharge and agricultural runoff (Ökten and Yazicigil 2005). These issues threaten not only the region's economic productivity but also its ecological balance and availability of potable water for local communities.

The basin's topography features flat plains and gentle hills conducive to extensive agricultural operations. Geologically the region comprises a heterogeneous aquifer system with alternating layers of alluvial sands, clays, and silts that influence recharge and discharge dynamics (Mahmoody et al. 2018). Groundwater recharge occurs primarily through precipitation and surface water infiltration, while discharge is driven by intensive agricultural pumping and industrial water withdrawals. This combination of ecological and economic pressures underscores the necessity of accurate groundwater monitoring and predictive modeling in the Ergene River Basin.

Groundwater data were collected from four observation wells strategically selected to capturethe heterogeneity of the aquifer system and the spatial variability within the Ergene River Basin. These wells are distributed across the Kırklareli and Tekirdağ provinces,

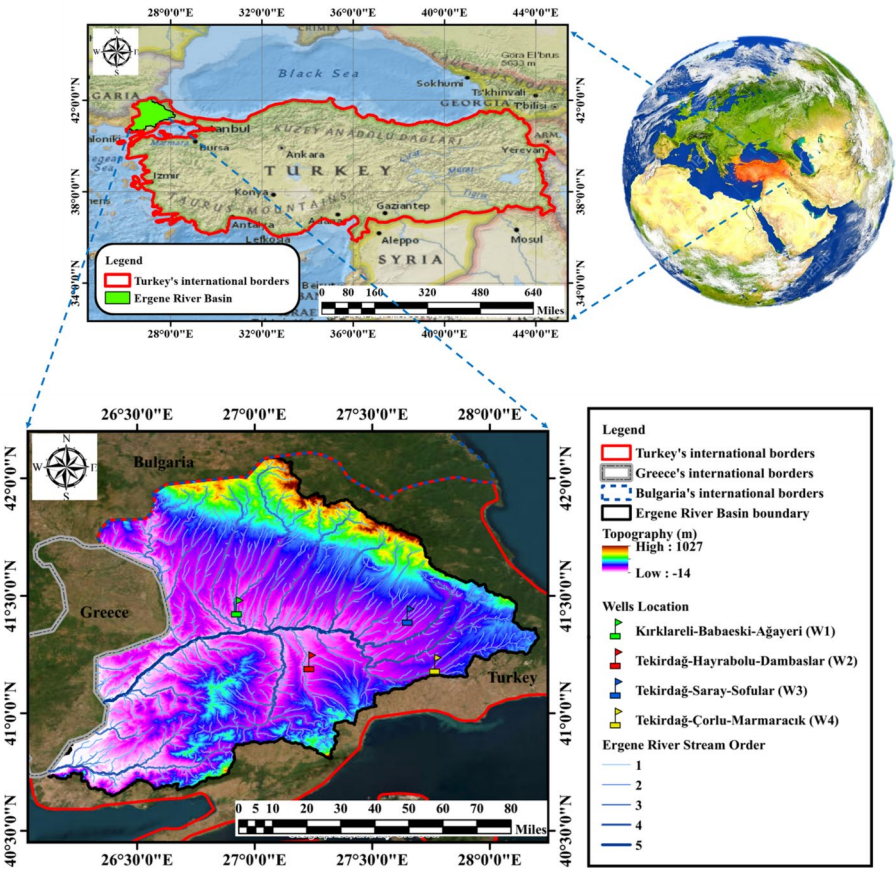

Fig. 1 A map showing the study area at Ergene River Basin, Türkiye, and the location of four observation wells

reflecting the diverse hydrogeological conditions of the basin. Detailed information regarding the coordinates, locations, and statistical properties of the wells is provided in Table 1, while their geographic distribution is shown in Fig. 1.

The dataset, obtained from the General Directorate of State Water Works of Türkiye, spaned from 1966 to 2023, offering a robust temporal range for analysis. Weekly groundwater level measurements were recorded over this period, ensuring comprehensive coverage of long-term fluctuations. The wells vary in depth, ranging from 200 m to 450 m, capturing both shallow and deep aquifer dynamics. This range provided valuable insights into the interactions between surface water and groundwater systems as well as the impacts of extraction activities on aquifer sustainability.

W1, located in Kırklareli, recorded the highest groundwater levels, with a minimum of −14.25 m and an average of −7.04 m. In contrast, W3, situated in Tekirdağ, demonstrated the lowest levels, with a minimum of − 64.75 m and an average of −32.91 m. The other wells, W2 and W4, also exhibited distinct characteristics, highlighting spatial variability and heterogeneity of the aquifer system. These long-term datasets were indispensable for

Table 1 Details of the four observation wells in the study area, used for conducting this research

Well ID		W1	W2	W3	W4
Well location	Province	Kırklareli	Tekirdağ	Tekirdağ	Tekirdağ
	District	Babaeski	Hayrabolu	Çorlu	Saray
	Village	Ağayeri	Dambaslar	Marmaracık	Sofular
Coordinate	Zone	35	35	35	35
	X	493870 E	519719 E	564550 E	554617 E
	Y	4589129 N	4563239 N	4562340 N	4585289 N
Ground elevation (m.a.s.l)		48	64	123	114
Well depth (m)		248	300	450	200
Range of data recording time	From	June 1975	November 1966	September 1969	June 1971
	To	April 2023	April 2023	April 2023	April 2023
Water level ranges (m.a.s.l)	Min.	− 14.25	− 40.7	− 64.75	− 44
	Max.	− 1.57	− 4.68	− 16.5	− 14.95
	Mean	− 7.04	− 18.61	− 32.91	− 28.34

predictive modeling efforts to understand groundwater behavior and develop sustainable water resource management strategies for the Ergene River Basin.

2.2 Research Methodology

Detailed steps of the research methodology are demonstrated in the following subsections (Fig. 2).

2.2.1 Data Processing

Data preparation is essential for building reliable ML models. Groundwater level data from wells W1–W4 were normalized using the minimum-maximum method (Eq. 1) to [0, 1], ensuring consistency and enhancing model training efficiency and prediction accuracy (Singh and Singh 2020).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X_{norm} represents the normalized value of the dataset, X denotes the original value being transformed, and X_{min} & X_{max} correspond to the minimum and maximum values observed within the dataset, respectively.

Following normalization, the groundwater dataset from four wells was split into training (70%) and testing (30%) subsets to support accurate predictions (Fig. 3). This 70/30 division, widely used in ML forecasting, may allow the model to learn from sufficient data

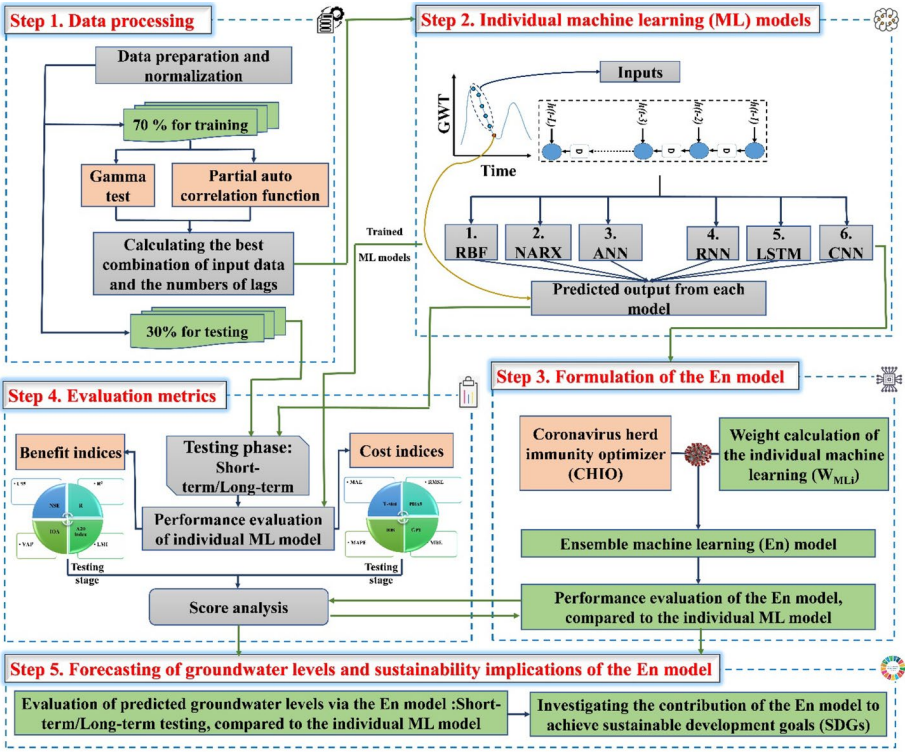

Fig. 2 Methodology flowchart for the current research study

while validating performance on unseen data and reducing overfitting risks. Also, it can ensure effective pattern recognition and robust generalization, particularly for time-series data (Li et al. 2021). Both datasets underwent the same procedures to maintain consistency and comparability. Details of these processes are provided in the following sub-sections.

- Preparation of Training Data

Selecting optimal input data (lag numbers) is crucial for accurate time series forecasting using ML models. Appropriate lags can enhance predictive performance and model relevance (Dong et al. 2023). Since no universal method exists for input selection in data-driven models, partial autocorrelation function (PACF) and gamma test (GT) were used to identify significant historical data points. PACF can measure the direct correlation between a time series X_t and its lagged version X_{t-k}, excluding intermediate lag effects, and is computed recursively via the Yule-Walker formula, as shown in the following equation (Kumar et al. 2022).

$$\varnothing_k = \frac{Cov\left(X_t,\ X_{t-k}\right) - \sum_{j=1}^{k-1} \varnothing_j \times Cov\left(X_{t-j},\ X_{t-k}\right)}{Var\left(X_t\right)} \tag{2}$$

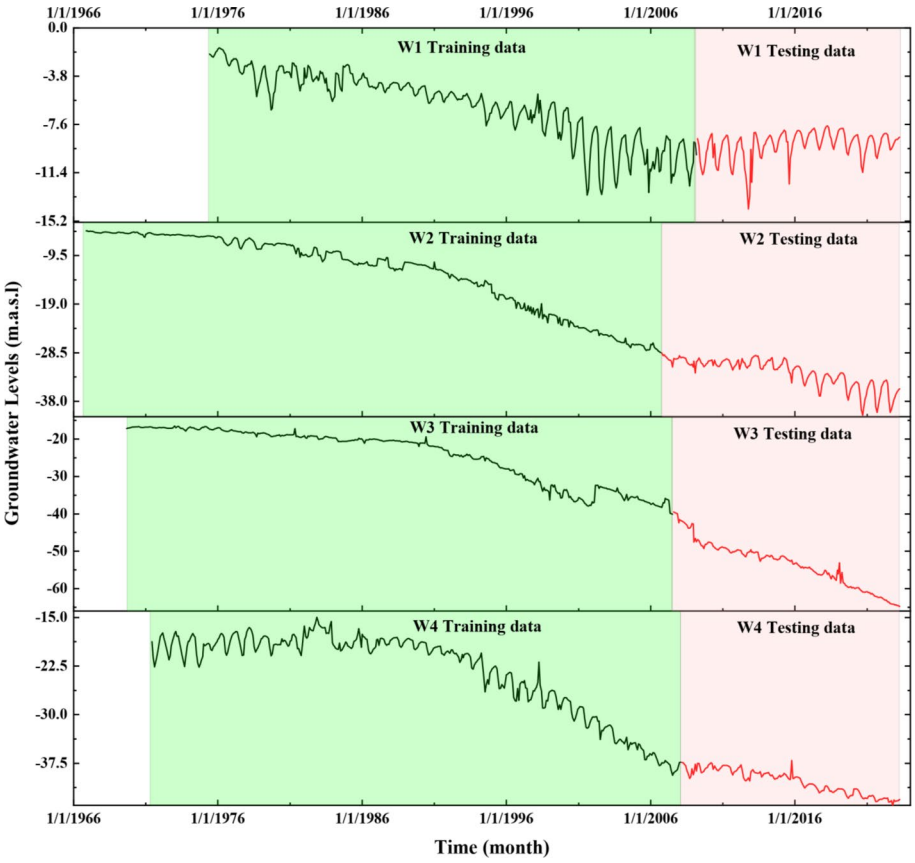

Fig. 3 Temporal data of groundwater levels (70% for training and 30% for testing) at different observation wells (W1, W2, W3, and W4) in the Ergene River basin, Türkiye

where $\varnothing_k$ represents the partial autocorrelation at lag k, Cov (X_t, X_{t-k}) denotes the covariance between X_t and X_{t-k}, $\varnothing_j$ indicate the partial autocorrelation coefficients for intermediate lags (j <k), and Var(X_t) represents the variance of X_t.

Figure 4a shows strong temporal correlations, with lag values exceeding the confidence interval. PACF analysis identified optimal lags of 10 for W1, 12 for W2, and 15 for W3 and W4, reflecting each well's unique hydrogeological dynamics. To validate these findings, GT was applied. GT can estimate data noise and identify optimal lag structures by minimizing unexplained variance. The lag corresponding to the minimum noise level (Γ), computed using Eq. (3), represented the optimal input for forecasting, ensuring accurate modeling of groundwater dynamics across heterogeneous conditions (Evans and Jones 2002).

$$\Gamma = \lim_{n\to\infty} \frac{1}{k} \sum_{i=1}^{k} \left(y_i - \frac{1}{k} \sum_{j\in N(i)} y_i \right)^2 \quad (3)$$

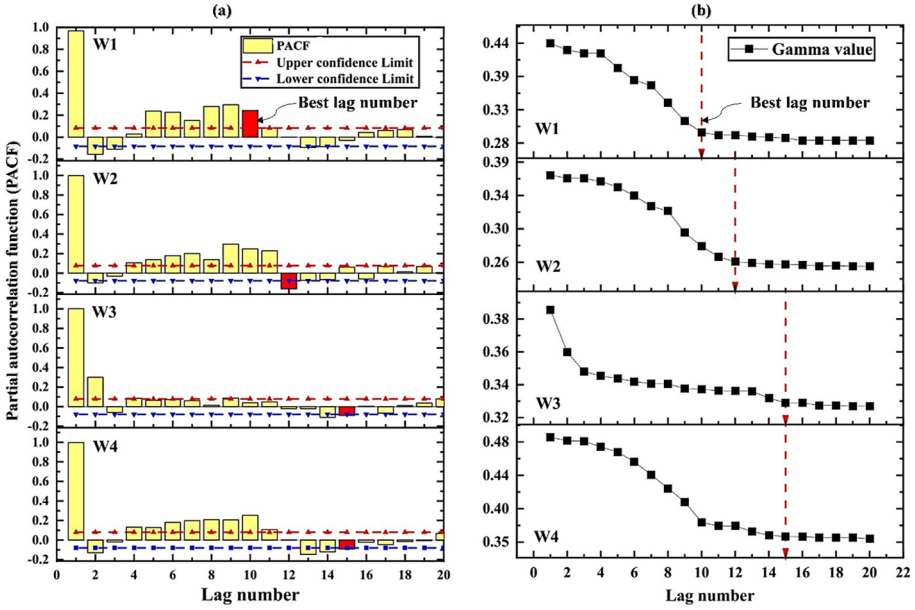

Fig. 4 Determining the best lag number at different observation wells (W1, W2, W3, and W4) in the study area using: (**a**) partial autocorrelation function (PACF), and (**b**) gamma test (GT)

where y_i represents the target value (future value) at the time i, N(i) denotes the set of k-nearest neighbors for the i^{th} point, y_j indicates the target values of these k-nearest neighbors and z denotes the number of time steps.

The GT confirmed PACF results (Fig. 4b), yielding low gamma values at the same lags, e.g., lag 15 for W3. This agreement validated lag selection robustness, ensuring relevant temporal dependencies for accurate ML forecasting. Combining PACF and GT can support precise lag identification (Dong et al. 2023).

- Testing Scenarios

The testing phase evaluated ML model performance using two scenarios: short-term and long-term testing, based on 30% of observed data (Rahmandad et al. 2021; Saini et al. 2023). These scenarios can provide a comprehensive assessment of predictive accuracy and robustness.

- Short-term Testing

Short-term testing uses observed data, where actual historical inputs generate one-step-ahead predictions at each time point, Eq. (4). The entire prediction sequence is computed in a single operation, minimizing error accumulation, and providing a direct comparison with observed data.

$$h(t) = f(h(t-1),\ h(t-2), \ldots\ldots,\ h(t-L)) \tag{4}$$

where h(t) is the predicted groundwater level, h(t- 1), h(t- 2),…, h(t-L) are actual observed historical values, and L represents the optimal number of lag times determined through PACF and GT analyses.

- Long-term Testing

Long-term testing simulates real scenarios by predicting future values without observed data, starting with Eq. (4) and continuing recursively.

$$\begin{aligned} \mathrm{h}(\mathrm{t}+1) &= \mathrm{f}\left(\widehat{\mathrm{h}}(\mathrm{t}),\ \mathrm{h}(\mathrm{t}-1),\ldots,\ \mathrm{h}(\mathrm{t}-\mathrm{L}+1)\right) \\ \mathrm{h}(\mathrm{t}+2) &= \mathrm{f}\left(\widehat{\mathrm{h}}(\mathrm{t}+1),\ \widehat{\mathrm{h}}(\mathrm{t}),\ \mathrm{h}(\mathrm{t}-1),\ldots,\ \mathrm{h}(\mathrm{t}-\mathrm{L}+2)\right) \\ &\cdot \\ \mathrm{h}(\mathrm{t}+\mathrm{n}) &= \mathrm{f}\left(\widehat{\mathrm{h}}(\mathrm{t}+\mathrm{n}-1),\ \widehat{\mathrm{h}}(\mathrm{t}+\mathrm{n}-2),\ldots,\ \widehat{\mathrm{h}}(\mathrm{t}+\mathrm{n}-\mathrm{L})\right) \end{aligned} \tag{5}$$

where $\widehat{\mathrm{h}}()$ represents predicted values used as inputs.

In this approach, the model used its predictions as inputs, updating sequences iteratively. This tested robustness and error propagation over time (Rahmandad et al. 2021). Together, both testing methods can ensure comprehensive performance evaluation under controlled and realistic forecasting conditions.

2.2.2 Individual Machine Learning (ML) Models

To forecast groundwater levels, six ML models were individually applied: three shallow (RBF, ANFIS, ANN) and three deep (RNN, LSTM, CNN). These models were selected for their prior success in hydrological forecasting (Wunsch et al. 2021; Yang and Zhang 2022; Pourmorad et al. 2024) (Fig. 5a–f).

- Radial Basis Function (RBF)

The RBF network, a shallow ML model, can capture nonlinear patterns in time series, making it ideal for groundwater forecasting (Momeneh and Nourani 2022). As shown in Fig. 5a, it can transform lagged inputs h(t −1)…h(t −L) via $\varnothing_i$(h) and weights W_i to predict h(t), as expressed in the following equation.

$$h(t) = \sum_{i=1}^{k} W_i \varnothing_i \left(\| h - c_i \|\right) \tag{6}$$

where L represents the number of lagged groundwater levels used as inputs, and $\| h - c_i \|$ denotes the Euclidean distance calculated between the input vector h and the center C_i.

- Adaptive Neuro-fuzzy Inference System (ANFIS)

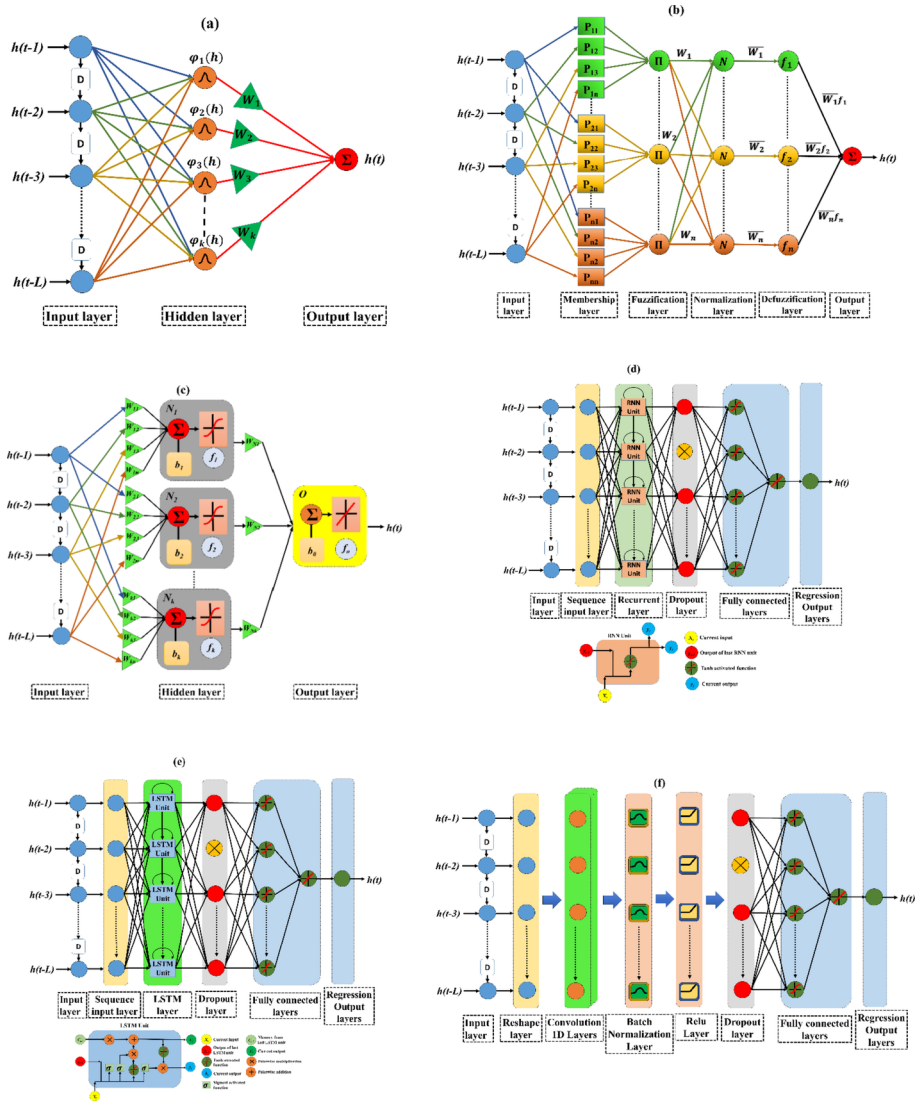

Fig. 5 Different structures of theindividual machine learning (ML) models: (**a**) radial basis function (RBF), (**b**) adaptive neuro-fuzzy inference system (ANFIS), (**c**) artificial neural networks (ANN), (**d**) recurrent neural network (RNN), (**e**) long short-term memory (LSTM), and (**f**) convolutional neural network (CNN)

The ANFIS model can capture nonlinear temporal dependencies in groundwater level forecasting by combining fuzzy logic with neural network learning (Fig. 5b). It can process time-lagged groundwater inputs through membership functions to calculate rule firing strengths, which are then normalized and used to generate partial rule-based outputs. The normalized firing strength $\overline{W}_i$ of each rule can be computed using the product of membership values for each input variable, as shown in Eq. (7). These strengths can be normalized and applied to consequent functions, where the weighted sum of rule outputs can produce

the final groundwater level prediction, as described in Eq. (8). This formulation can allow ANFIS to adaptively optimize both antecedent and consequent parameters, to improve prediction accuracy in hydrological forecasting tasks and get h(t) (Moosavi et al. 2013).

$$\overline{W}_i = \frac{W_i}{\sum_{i=1}^{n} W_i} \quad \& \quad W_i = \prod_{j=1}^{L} \mu_{P_{ij}} (h(t-j)) \tag{7}$$

$$h(t) = \sum_{i=1}^{k} \overline{W}_i f_i = \sum_{i=1}^{k} \overline{W}_i (p_i h(t-1) + q_i h(t-2) + \ldots + r_i h(t-L) + b) \tag{8}$$

where $h(t-j)$ is the lagged groundwater level at time $t-j$, μP_{ij} is the membership function for the j^{th} rule and i^{th} input, f_i is the consequent function of the i^{th} rule, p_i & q_i & r_i are the coefficients of the consequent linear function for the i^{th} rule, and b is the bias term.

- Artificial Neural Networks (ANN)

ANN are computational models capable of learning complex nonlinear relationships in hydrological data (Pourmorad et al. 2024). As shown in Fig. 5c, ANN typically includes an input layer (lagged groundwater levels), a hidden layer(s), and an output layer. In hidden layers, inputs are weighted (W_{kj}), summed (Σ), and transformed using activation functions (e.g., sigmoid). Outputs are aggregated via weights and biases to predict h(t), as expressed below (Shen 2018).

$$h(t) = f\left(\sum_{k=1}^{K} W_k f_k \left(\sum_{j=1}^{N} W_{kj} h(t-j) + b\right) + b\right) \tag{9}$$

where N & K are neuron counts in input and hidden layers, and f_k & f are activation functions. ANN parameters are optimized using backpropagation to minimize loss (Rajaee et al. 2019).

- Recurrent Neural Network (RNN)

RNNs are deep learning models designed to process sequential data by capturing temporal dependencies, making them ideal for groundwater level forecasting (Wunsch et al. 2021; Farouk et al. 2024). As shown in Fig. 5 d, the architecture includes input, sequence input, recurrent, fully connected, and regression output layers. The recurrent layer can process input sequences using RNN units that consider both current input and previous hidden states, maintaining contextual information across time steps, as expressed mathematically below (Mienye et al. 2024).

$$y_t = f(W_{hx} \cdot x_t + W_{hh} \cdot y_{t-1} + b) \tag{10}$$

where, y_t is the hidden state, x_t is the input at time t, W_{hx} & W_{hh} are input-to-hidden and hidden-to-hidden weights. Fully connected layers can refine features before predicting h(t) via regression output.

$$h(t) = f(W_o \cdot y_t + b_o) \tag{11}$$

where W_o is the weight matrix of the output layer.

- Long Short-term Memory (LSTM)

LSTM networks, an advanced RNN variant, are designed to capture long-term dependencies in sequential data and are ideal for groundwater forecasting (Kratzert et al. 2019; Yang and Zhang 2022). As shown in Fig. 5e, the LSTM architecture includes input, sequence input, fully connected, and regression output layers. LSTM units use input, forget, and output gates to manage information flow, as expressed mathematically below (Khorram and Jehbez 2023).

$$\begin{cases} i_t & = \sigma(W_i \cdot [y_{t-1}, x_t] + b_i) \\ f_t & = \sigma(W_f \cdot [y_{t-1}, x_t] + b_f) \\ o_t & = \sigma(W_o \cdot [y_{t-1}, x_t] + b_o) \\ \tilde{c}_t & = \tanh(W_c \cdot [y_{t-1}, x_t] + b_c) \\ c_t & = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ y_t & = o_t \odot \tanh(c_t) \end{cases} \tag{12}$$

where i_t, f_t, and o_t are input, forget, and output gates, c_t is the cell state, W & b are weights and biases, tanh is the sigmoid function, and $\odot$ denotes element-wise multiplication. This structure can allow LSTMs to capture long-term dependencies and avoid vanishing gradients (Wani et al. 2024). Fully connected and regression layers can yield the final prediction h(t).

$$h(t) = W_o \cdot (o_t \odot \tanh(c_t)) + b_o \tag{13}$$

- Convolutional Neural Network (CNN)

CNNs are a class of deep ML models widely used to capture spatial and temporal patterns, making them suitable for groundwater-level forecasting when data show strong spatial dependencies or temporal correlations (Wunsch et al. 2021). As shown in Fig. 5f, the CNN architecture typically includes an input layer, sequence input layer, convolutional and pooling layers, fully connected layers, and regression output layers. The input layer takes (h(t −1), h(t −2),…, h(t −L), which are transformed into sequences. Convolutional layers extract spatial/temporal features using filters while pooling layers down-sample data. Fully connected layers aggregate features and pass them to the regression output layer to produce h(t), mathematically expressed by Eq. (14) in this study.

$$h(t) = f(W_o \times z + b_o) \tag{14}$$

where z is the feature vector, generated by convolutional and pooling layers.

2.2.3 Formulation of the Ensemble Machine Learning (En) Model

The ensemble prediction approach is preferred over individual models due to its ability to capture complex input-output relationships more effectively (Roy and Datta 2018). A weighted En model was developed to combine the six individual ML models to enhance the accuracy of groundwater level prediction in the study area based on their performance:

$$h_{En} = \sum_{i=1}^{n} \omega_i \times h_{IM_i} \tag{15}$$

where h_{En} is the ensemble output, h_{IM_i} represents the output of the i^{th} individual model, ω_i is the weight assigned to the i^{th} model, and n is the number of individual models.

To determine optimal weights for each model in the ensemble model, the CHIO was applied (Supplementary Eqs. S1–S3) (Al-Betar et al. 2021). This approach integrated cost and benefit indices to address trade-offs among performance metrics. Cost indices included mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), T-statistics (T-stat), global performance index (GPI), mean absolute percentage error (MAPE), percent bias (PBIAS), and index of scatter (IOS). Benefit indices comprise coefficient of determination (R^2), uncertainty at 95% confidence interval (U95), Nash–Sutcliffe efficiency (NSE), variance accounted for (VAF), index of agreement (IOA), Legate McCabe's index (LMI), correlation coefficient (R), and A10 index. Detailed descriptions of these indices aregiven in Supplementary Tables S1–S2 and Eqs. S4–S19.

For the benefit and cost indices, the weights $\omega_1 \rightarrow \omega_6$ were assigned for the individual predictive models, i.e., RBF, ANFIS, ANN, RNN, LSTM, and CNN, respectively. The optimization framework addressed two competing objectives: primarily, the maximization of aggregate benefit indices across all prediction models and secondarily, the minimization of their cumulative cost indices. The mathematical representation of the proposed CHIO-based weight assignment methodology can be expressed as follows:

$$\begin{cases} \text{Maximize} : f_1(\text{BI}) = \sum_{i=1}^{N} \omega_i^n \times \sum_{j=1}^{K} \text{BI}^k \\ \text{Minimize} : f_2(CI) = \sum_{i=1}^{N} \omega_i^n \times \sum_{l=1}^{L} \text{CI}^l \end{cases} \tag{16}$$

where $f_1(\text{BI})$ is the objective function for maximizing benefit indexes, $f_2(\text{CI})$ is the objective function for minimizing cost indexes, ω_i^n is the weight coefficient for the i^{th} model, BI^k is the k^{th} benefit index, CI^l is the l^{th} cost index, and N & K & L are the total numbers of models, benefit indexes, and cost indices, respectively.

2.2.4 Evaluation Metrics

To evaluate the predictive accuracy of individual ML models, 16 performance metrics (benefit and cost indices) were used, as reported by Khatti et al. (2024) and Saqr et al. (2025). Score analysis was performed during testing to compare the developed En model with six

individual ML models. This method aggregated ranks from all 16 indices into a single score, widely adopted in hydrological studies. Each model was ranked from 1 to 6 per index based on the proximity to the optimal value, i.e., closer to zero for cost indices and closer to unity for benefit indices. The final score was calculated as the sum of all ranks, and reflected overall performance, with higher scores indicating superior accuracy and robustness (Kutty et al. 2023).

2.2.5 Forecasting of Groundwater Levels and Sustainability Implications Due to the Developed Ensemble Machine Learning (En) Model

The developed En model can result in accurate groundwater prediction, supporting the evaluation of its contribution to the environmental, social, and economic pillars of the SDGs. Also, it can aid in resource conservation, equitable distribution, and efficient use while enhancing environmental stability, promoting social equity, and maximizing economic benefits (Timalsina et al. 2025).

3 Results

3.1 Accuracy Assessment of Individual Machine Learning (ML) Models

The precision of six individual ML models, i.e., RBF, ANFIS, ANN, RNN, LSTM, and CNN, for forecasting groundwater levels at four observation wells (W1–W4) was assessed during training and two testing phases: short-term and long-term. EEvaluation was conducted using cost indices (e.g., RMSE) and benefit indices (e.g., R^2), visualized via Taylor diagrams (Figs. 6a–d and 7a–d, and 8a–d), with detailed metrics in Supplementary Tables S(3–5). During the training phase, all models (shallow and deep) demonstrated similar accuracy, with R^2 approaching 1.0 and RMSE nearing zero, indicating strong agreement with observed data (Fig. 6a–d). At W1, LSTM and CNN achieved $R^2 \sim 0.99$ and RMSE ~0.5 m. Comparable performance was recorded across W2, W3, and W4, where shallow models (RBF, ANFIS) also achieved $R^2 \sim 0.99$ despite greater groundwater variability.

In short-term testing, where observed inputs were used for one-step-ahead predictions, deep ML models consistently outperformed shallow ones across all wells (Fig. 7a–d). At W2, RNN and LSTM achieved $R^2 = 0.99$ and RMSE = 0.5 m, while shallow models showed higher RMSE values, up to ~0.9 m (Fig. 7b). At W3, RNN, LSTM, and CNN reached $R^2 > 0.99$ with RMSE ~0.6 m, outperforming ANFIS and RBF (RMSE = 1.0 m) (Fig. 7c). At W1 and W4, deep models showed marginal advantages: RMSE = 0.4 m compared to 0.6 m at W1 (Fig. 7a), and RMSE = 0.4 m vs. 0.5 m at W4 (Fig. 7d).

In long-term testing, which relied on previous predictions as inputs, deep models again outperformed shallow ones (Fig. 8a–d). At W1, LSTM, RNN, and CNN achieved RMSE = 0.6 m, outperforming ANFIS (1.2 m) and RBF (1.4 m) (Fig. 8a). At W2, deep models held RMSE = 1.5 m, while shallow models ranged from 1.8 m (RBF, ANFIS) to 2.5 m (ANN) (Fig. 8b). At W3, deep models maintained RMSE = 3.0 m, while RBF and ANFIS dreached 7.0 m (Fig. 8c). At W4, LSTM led all models with RMSE = 1.0 m (Fig. 8d).

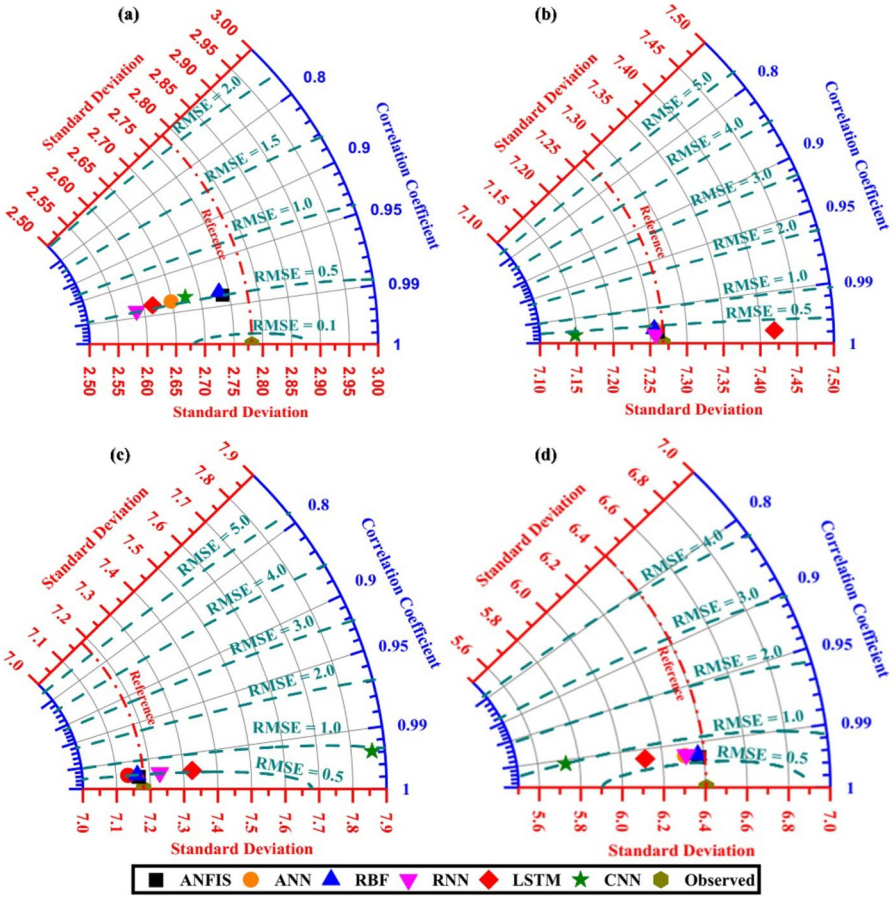

Fig. 6 Performance evaluation using the Taylor diagram for the six individual machine learning (ML) models under investigation during the training phase at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

3.2 Performance Evaluation of the Developed Ensemble Machine Learning (En) Model

The performance of the En model was evaluated against six individual ML models using score analysis. The En model consistently doutperformed all individual models, achieving the highest score values across all wells (W1–W4), as shown in Fig. 9(a–d). At W1, the En model scored 203, surpassing deep ML models: CNN (176), LSTM (165), and RNN (164). In contrast, shallow models showed weaker performance, with scores of 55 (RBF), 63 (ANN), and 134 (ANFIS) (Fig. 9a). At W2, the En model again led with a score of 201. Deep models followed, scoring 146 (CNN), 145 (LSTM), and 162 (RNN), while shallow models performed notably worse: 132 (RBF), 42 (ANN), and 92 (ANFIS) (Fig. 9b). At W3, where high spatial and temporal variability posed forecasting challenges, En model achieved its high performance with a score of 224. Deep models also performed well: 172 (LSTM), 163 (RNN), and 162 (CNN). However, shallow models struggled with the aqui-

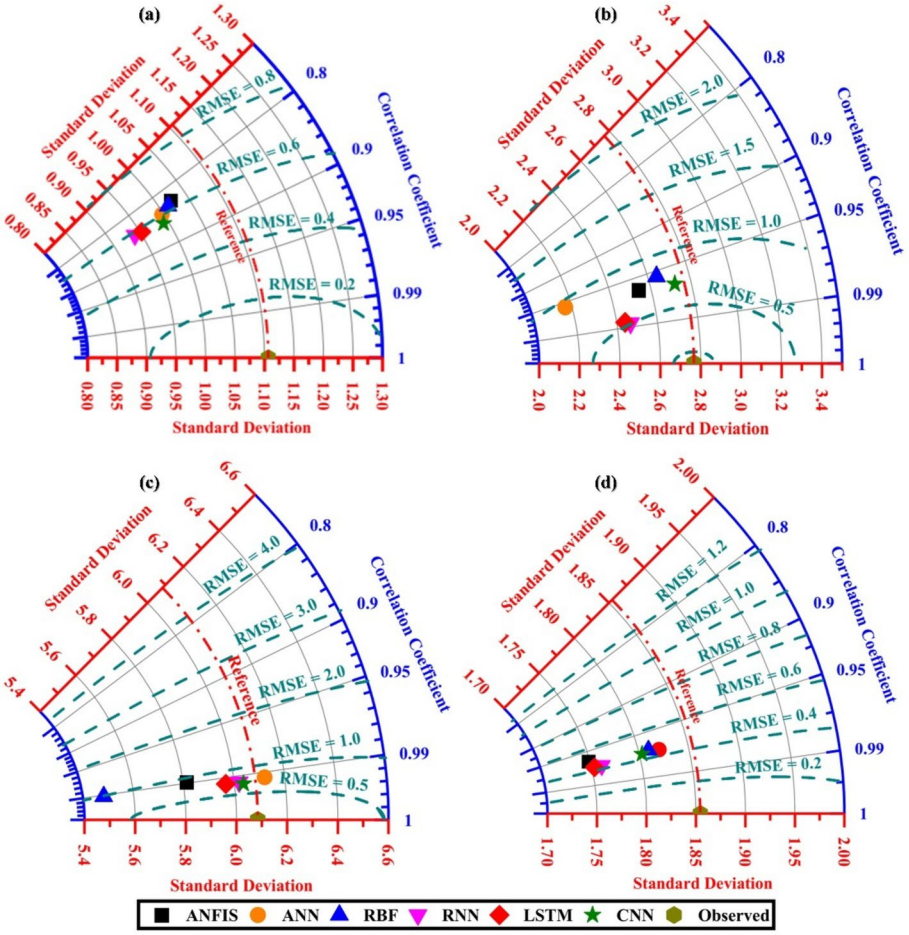

Fig. 7 Performance evaluation using the Taylor diagram for the six individual machine learning (ML) models under investigation during the short-term testing phase at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

fer's heterogeneity, scoring 63 (RBF), 127 (ANN), and 60 (ANFIS) (Fig. 9c). At W4, where groundwater conditions were relatively stable, the En model maintained top performance with a score of 208. Deep ML models followed: LSTM (175), CNN (165), and RNN (137). Shallow models showed improved but still inferior performance, scoring 77 (RBF), 75 (ANN), and 89 (ANFIS) (Fig. 9d).

3.3 Groundwater lEvel Forecasting via the Developed Ensemble Machine Learning (En) Model vs. the Individual Machine Learning (ML) Models

The temporal variation of groundwater levels during short- and long-term testing confirmed thatthe developed En model consistently outperformed individual models (RBF, ANFIS, ANN, RNN, LSTM, and CNN) in the stability and alignment with observed data across all wells (W1–W4), as shown in Figs. 10(a–d) and 11(a–d).

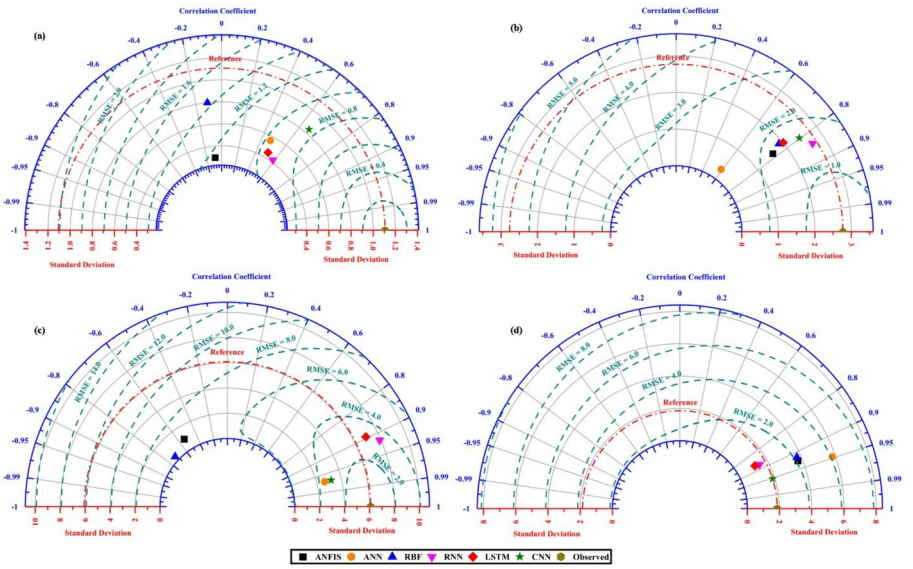

Fig. 8 Performance evaluation using the Taylor diagram for the six individual machine learning (ML) models under investigation during the long-term testing phase at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

In the short-term testing phase, the En model accurately captured fluctuations at all wells. At W1, it closely followed observed data, remaining stable even during abrupt changes. While LSTM and CNN performed well, they showed minor deviations at peaks and troughs (Fig. 10a). Shallow models like ANFIS and RBF had difficulty with sudden variations. At W2, with moderate variability, the En model aligned strongly with observed trends. Deep models such as RNN and LSTM performed reasonably but introduced slight oscillations, while ANN and RBF failed to capture sharp changes effectively (Fig. 10b). At W3, the most variable well, the En model tracked complex patterns accurately, outperforming CNN and RNN, which diverged at abrupt peaks (Fig. 10c). Shallow models underperformed significantly. At W4, under more stable conditions, the En model retained top accuracy, with CNN and LSTM performing comparably, though ANFIS deviated more during changes (Fig. 10d).

In long-term testing, the En model maintained consistent alignment with extended groundwater trends. At W1, it captured both seasonal and sustained trends precisely, while LSTM and CNN slightly underestimated trends, and ANFIS overestimated them (Fig. 11a). At W2, the En model effectively followed the gradual decline, outperforming RNN and CNN, which showed minor deviations (Fig. 11b). Shallow models struggled more with rapid declines. At W3, the En model managed steep trends with minimal error, while LSTM exhibited increasing divergence, and ANN and RBF performed poorly (Fig. 11c). At W4, the En model sustained strong accuracy, while LSTM and CNN remained close but less stable and shallow models still showed notable deviations (Fig. 11d).

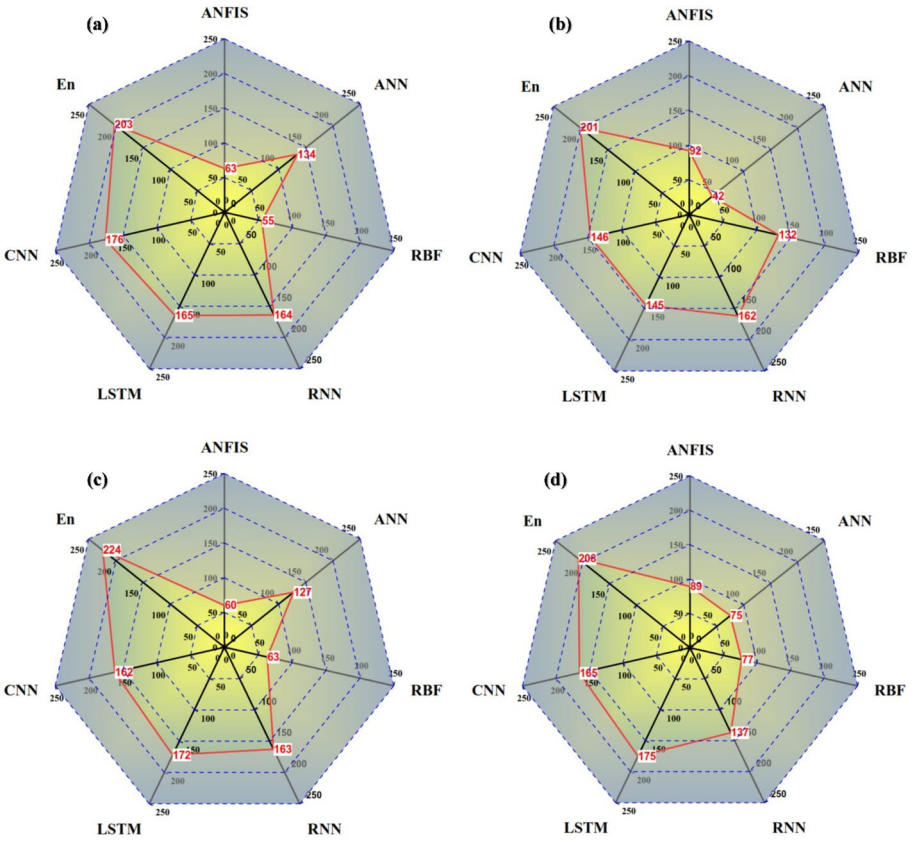

Fig. 9 Radar/spider diagrams showing the score value of the developed ensemble machine learning (En) model versus individual machine learning (ML) models (RBF, ANFIS, ANN, RNN, LSTM, and CNN) during short-term testing at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

3.4 Impact of the Study Findings on Sustainable Development Goals (SDGs)

Based on the accuracy of the En model in capturing spatial and temporal groundwater variations in the Ergene River Basin, it can offer substantial contributions to environmental, economic, and social sustainability dimensions, aligning with several SDG targets (Fig. 12). Environmentally, it can support 50% of SDG 6 by advancing sustainable groundwater management. Specifically, it can contribute to targets 6.1 (ensuring safe and affordable drinking water), 6.3 (improving water quality), 6.4 (enhancing water-use efficiency), and 6.5 (integrated water resources management) through accurate forecasts that inform extraction and quality monitoring (Timilsina et al. 2025). It also can address 40% of SDG 13 by aiding climate resilience through targets 13.1 (adapting to climate-related hazards) and 13.2 (integrating climate considerations into policies). Economically, the model can contribute to 17% of SDG 8 by supporting targets 8.1 (sustaining economic growth) and 8.4 (improving resource efficiency), ensuring reliable groundwater access for agriculture and industry. Under SDG 12, it can support 9% through target 12.2 (sustainable resource management), enabling informed extraction that may respect aquifer recharge rates. Socially, the model

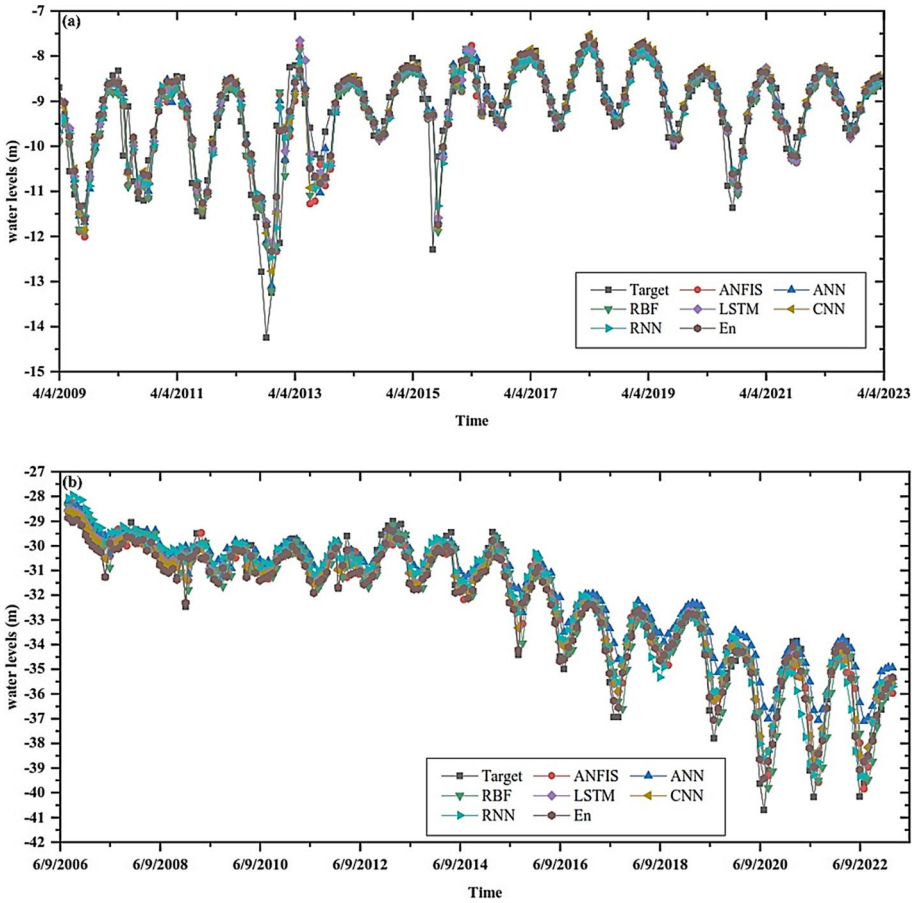

Fig. 10 Temporal variation of groundwater levels during the short-term testing phase of the developed ensemble machine learning (En) model compared to the individual machine learning (ML) models at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

can enhance public health and urban sustainability, contributing to 8% of SDG 3 via target 3.9 (reducing health risks from polluted water) and 10% of SDG 11 through target 11.3 (inclusive and sustainable urbanization). Overall, the En model can serve as a practical decision-support tool that may promote proactive groundwater governance, climate adaptation, and equitable water distribution across communities, reinforcing its alignment with global SDGs.

4 Discussion

The present study introduced a novel En model optimized using CHIO for groundwater level forecasting in the Ergene River Basin, characterized by complex and dynamic aquifer behavior. The En model integrated multiple ML models (RBF, ANFIS, ANN, RNN, LSTM, and CNN) to overcome the limitations of single models and enhance prediction reliability. It

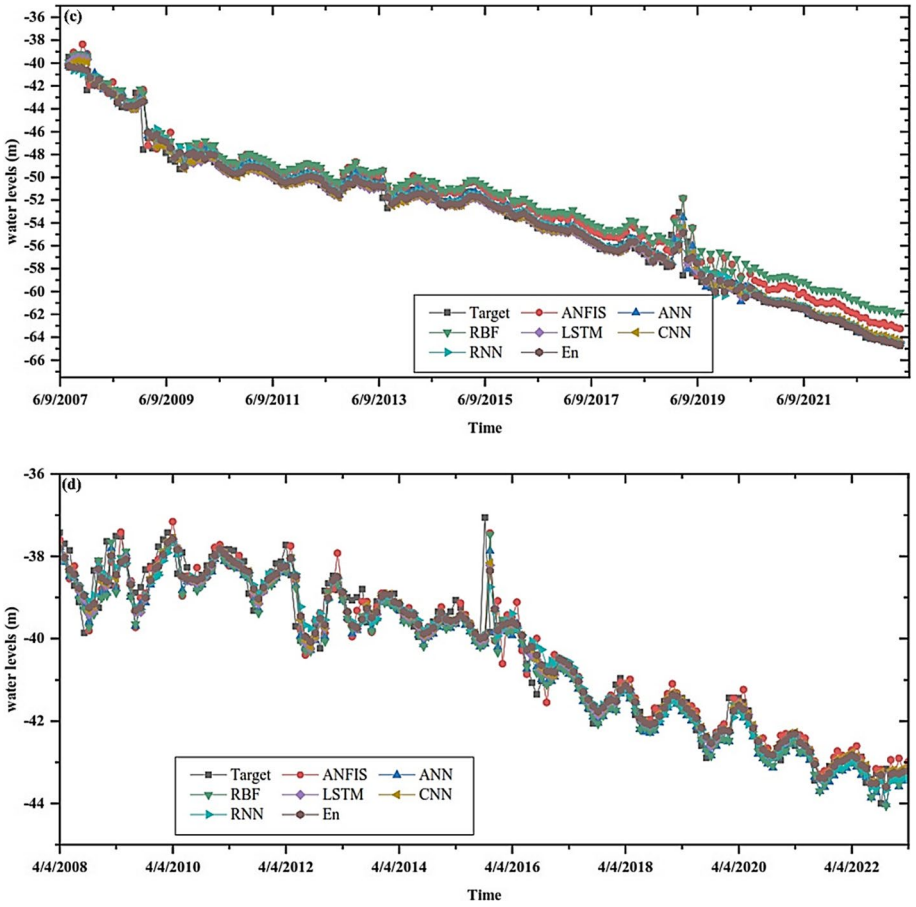

Figure 10 (continued)

demonstrated superior performance in both short- and long-term testing phases, effectively capturing abrupt fluctuations and long-term trends across all wells (W1–W4). The study findings with existing literature, highlighted the advantages of the En models in hydrological forecasting (Ravindran et al. 2021; Mirzania et al. 2024). Roy and Datta (2018) emphasized the improved robustness of En approaches in groundwater management. Gong et al. (2018) showed that the En model outperformed ANN and ANFIS in terms of forecasting accuracy, and similar benefits were observed in studies conducted in -semi-arid and coastal regions (Yoon et al. 2011; El Bilali et al. 2021).

CComparison between shallow and deep ML models revealed that deep models consistently outperformed shallow ones, achieving $R^2 \sim 0.99$ and RMSE ~0.5 m. This is due to deep models' ability to learn temporal dependencies and handle non-linear, complex patterns (Deulkar et al. 2025). Models like LSTM, RNN, and CNN effectively processed spatial temporal variability in groundwater data, as supported by Shen (2018), and Waqas and Humphries (2024). In contrast, shallow models (e.g., ANFIS, RBF) struggled with more dynamic conditions due to their limited structural capacity, as observed particularly at well

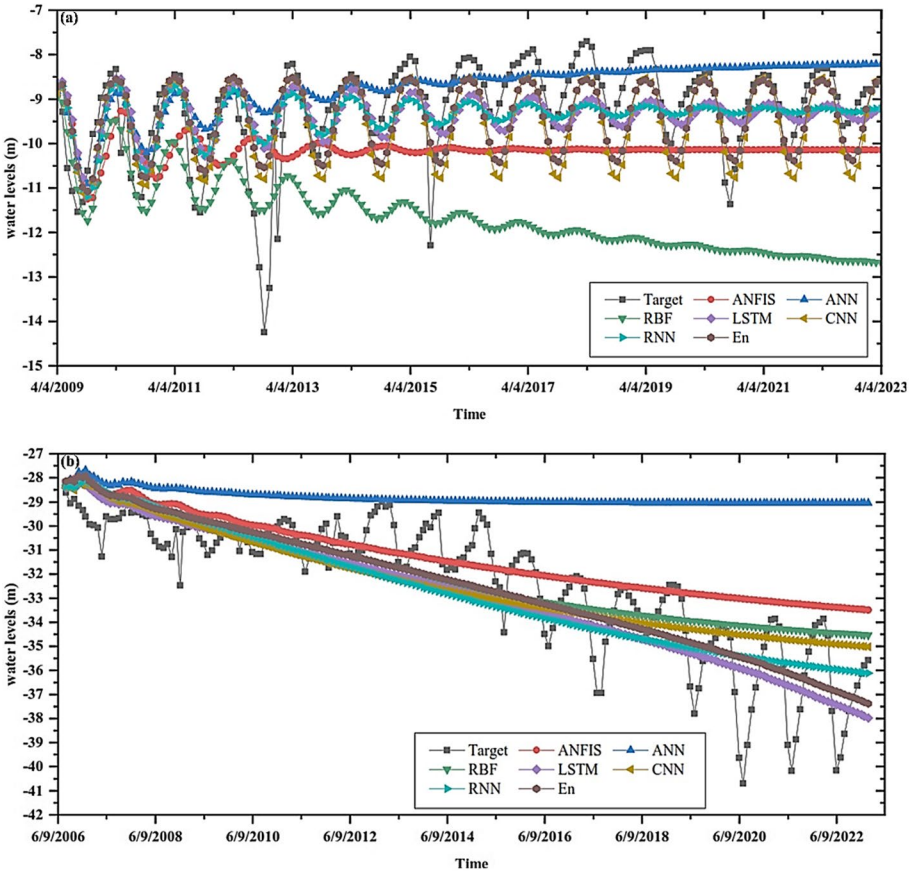

Fig. 11 Temporal variation of groundwater levels during the long-term testing phase of the developed ensemble machine learning (En) model compared tothe individual machine learning (ML) models at different observation wells: (**a**) W1, (**b**) W2, (**c**) W3, and (**d**) W4

W3, where groundwater fluctuations were more intense (Gaffoor et al. 2022; Igwebuike et al. 2025).

The En model outperformed all individual ML models across the board, achieving score values over 200 at each well. During short-term testing, it accurately tracked observed fluctuations with minimal deviation. In long-term testing, it maintained stability and captured extended trends, including those at well W3 where other models showed growing error. The strength of the En model lay in its ability to leverage the strengths of multiple models while mitigating the weaknesses of others (Wei et al. 2023). By aggregating predictions and assigning optimized weights using CHIO, the model reduced bias and enhanced overall robustness (Roy et al. 2023). This technique also can help counteract error propagation often seen in deep models during longer forecast horizons.

The En model's strong performance was a result of effective weight assignments that emphasized high-performing models and downplayed the influence of those with lower accuracy (Abbasi et al. 2022). This approach can ensure more balanced and reliable forecasts, even in heterogeneous groundwater systems. These results highlighted the En model's

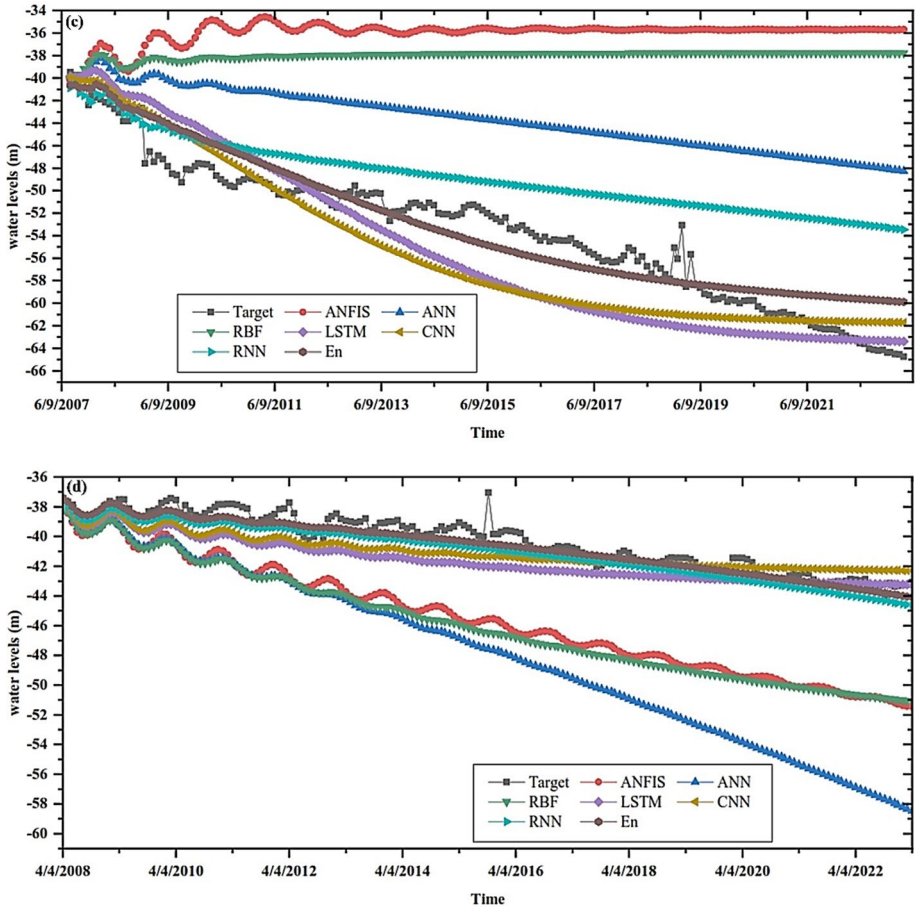

Figure 11 (continued)

consistent superiority in score-based performance, regardless of the wells' conditions. It can effectively handle both high-variability (W3) and more stable (W4) environments while individual models, particularly shallow ones, showed varying degrees of underperformance across all locations. Such consistent superiority across varied well conditions further highlighted the adaptability and practical potential of the developed En model for real-world groundwater forecasting applications (Pham et al. 2021).

Beyond technical performance, the En model can offer substantial contributions to SDGs. It can support SDG 6 by providing accurate groundwater forecasts that may guide sustainable usage, improve water use efficiency, and help preserve drinking water supplies (Timalsina et al. 2025). The model also can align with SDG 13, enhancing climate resilience by addressing variability in recharge and over-extraction (Paudel et al. 2024). These outcomes are consistent with prior studies on the role of ML in supporting climate-adaptive water management (Saqr et al. 2025). Economically, the model may contribute to SDG 8 and SDG 12 by supporting efficient groundwater use in agriculture and industry, which are vital for local productivity (Noman et al. 2024). Accurate forecasts can enable decision-makers to

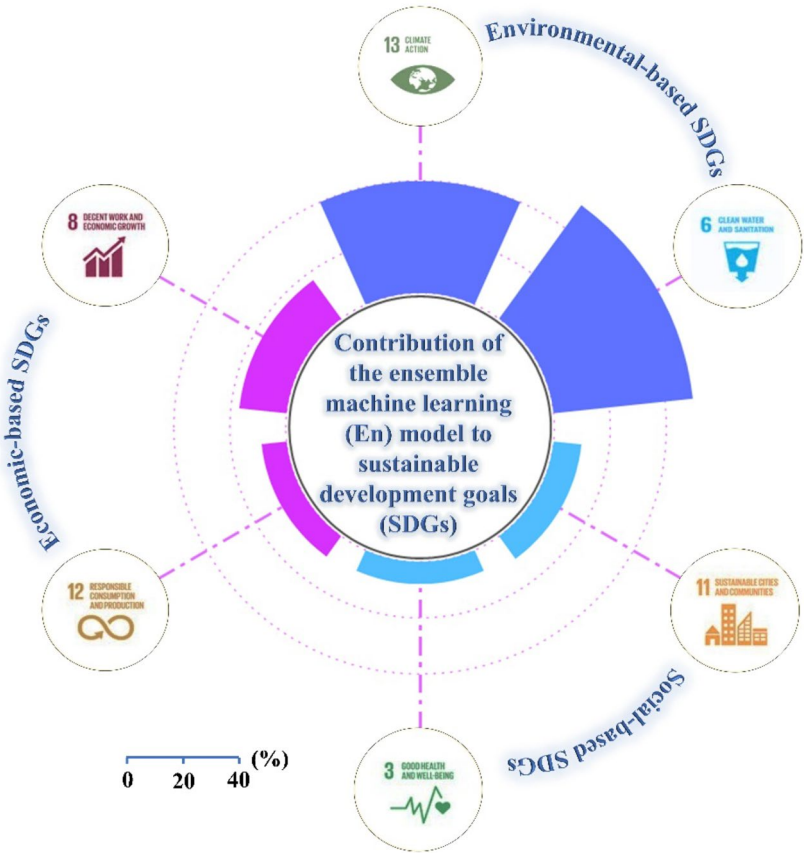

Fig. 12 Radial plot showing the quantitative contributions of the outcomes of the developed ensemble machine learning (En) model to sustainable development goals (SDGs) according to the three pillars of sustainability

optimize extraction practices and minimize wastage (Subbarayan et al. 2025). Socially, the En model may address SDG 3 and SDG 11 by promoting equitable groundwater distribution and safeguarding public health, especially in urban and vulnerable communities. Its role in reducing risks from over-extraction and contamination can align with broader global efforts to ensure water security and resilience (Masria et al. 2023).

5 Conclusion and Future Perspectives

This study developed a novel En model, optimized via CHIO, for accurate groundwater level forecasting in the Ergene River Basin, Türkiye. By integrating six ML models (RBF, ANFIS, ANN, RNN, LSTM, and CNN), the En model overcame limitations of standalone approaches and effectively represented aquifer heterogeneity. Weekly groundwater data (1966–2023) from four wells were split into 70% training and 30% testing under both short- and long-term scenarios. Optimal input lags were determined using PACF and GT,

enhancing model precision. While deep ML models outperformed shallow ones ($R^2 \sim 0.99$, RMSE ~0.5 m), the En model consistently achieved the highest accuracy, with score values exceeding 200 across all wells. It effectively captured both abrupt fluctuations and long-term trends, including at well W3, which exhibited complex temporal variability, demonstrating strong robustness and reliability for groundwater forecasting. The En model also can support multiple SDGs, contributing to 50% of SDG 6 and 40% of SDG 13 by improving water use efficiency and climate resilience. Economically, it can support 17% of SDG 8 and 9% of SDG 12 while socially it can contribute to 8% of SDG 3 and 10% of SDG 11, ensuring equitable water distribution and public health. Future research should incorporate real-time monitoring and test the model across diverse hydrogeological settings to enhance generalizability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11269-025-04210-w.

Acknowledgements The authors would like to thank the General Directorate of State Water Works of Türkiye for supplying data. Also, the authors would like to thank the Irrigation and Hydraulics Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt for facilitating the use of the computer workstation and the software MATLAB to execute the calculations.

Authors' Contributions A.M.S.: Conceptualization, Visualization, Investigation, Formal analysis, Methodology, Writing– original draft, Writing– Review & Editing. V.K.: Data acquisition, Conceptualization, Investigation, Writing– original draft, Writing– Review & Editing. E.K.: Conceptualization, Writing– original draft, investigation. M.E.A.: Conceptualization, Visualization, Investigation, Formal analysis, Methodology, Writing– original draft, Writing– Review & Editing.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).
This research was supported by the Faculty of Engineering, Mansoura University, Mansoura, Egypt. Besides, open access funding is provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data Availability Data will be made available on request.

Declarations

Ethical Approval The manuscript has not been submitted to any other journal. The proposed work is original and has not been published anywhere else.

Consent to Participate The authors declare that they have the consent to participate in this paper.

Consent to Publication The authors declare that they have consent to publish in this journal.

Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbasi E, Alavi Moghaddam MR, Kowsari E (2022) A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency. J Clean Prod 379. https://doi.org/10.1016/j.jclepro.2022.134588

Al-Betar MA, Alyasseri ZAA, Awadallah MA, Abu Doush I (2021) Coronavirus herd immunity optimizer (CHIO). Neural Comput Appl 33:5011–5042. https://doi.org/10.1007/s00521-020-05296-6

Alsulamy S, Kumar V, Kisi O et al (2025) Enhancing water level prediction using ensemble machine learning models: A comparative analysis. Water Resour Manag. https://doi.org/10.1007/s11269-025-04142-5

Arkoç O (2020) Assessment of Temporal variations of groundwater recharge in ergene basin (Northwestern Turkey) in terms of climate change. Kırklareli Üniversitesi Mühendislik Ve Fen Bilim Derg 6:95–118. https://doi.org/10.34186/klujes.771456

Coutinho ER, Madeira JGF, Borges DGF et al (2025) Multi-Step forecasting of meteorological time series using CNN-LSTM with decomposition methods. Water Resour Manag. https://doi.org/10.1007/s11269-025-04102-z

Deulkar AM, Dixit PR, Londhe SN, Jain RK (2025) Comparative assessment of artificial neural networks (ANNs), long short term memory network (LSTM) and hydrologic engineering Centre-Hydrologic modelling system (HEC-HMS) for runoff modelling. Water Resour Manag. https://doi.org/10.1007/s11269-024-04055-9

Dong Y, Yang T, Xing Y et al (2023) Data-driven modeling methods and techniques for pharmaceutical processes. Processes 11. https://doi.org/10.3390/pr11072096

El Bilali A, Taleb A, Brouziyne Y (2021) Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region. J Afr Earth Sci 181. https://doi.org/10.1016/j.jafrearsci.2021.104244

Evans D, Jones AJ (2002) A proof of the gamma test. Proc R Soc Math Phys Eng Sci 458:2759–2799. https://doi.org/10.1098/rspa.2002.1010

Ezzeldin R, Abd-Elmaboud M (2024) Modeling flow resistance and geometry of dunes bed form in alluvial channels using hybrid RANN–AHA and GEP models. Int J Sediment Res. https://doi.org/10.1016/j.ijsrc.2024.08.002

Farouk S, Elsamahy A, Kandil SA (2024) Power management of hybrid system using coronavirus herd immunity optimizer algorithm. J Electr Eng Technol. https://doi.org/10.1007/s42835-024-02026-z

Gaffoor Z, Pietersen K, Jovanovic N et al (2022) A comparison of ensemble and deep learning algorithms to model groundwater levels in a Data-Scarce aquifer of Southern Africa. Hydrology 9. https://doi.org/10.3390/hydrology9070125

Gezici K, Katipoğlu OM, Şengül S (2024) Hybrid machine learning models for groundwater level prediction in a snow-dominated region: an evaluation of EEMD, VMD and EWT decomposition techniques. Hydrol Process 38. https://doi.org/10.1002/hyp.15169

Gong Y, Wang Z, Xu G, Zhang Z (2018) A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. Water (Switzerland) 10. https://doi.org/10.3390/w10060730

Igwebuike N, Ajayi M, Okolie C et al (2025) Application of machine learning and deep learning for predicting groundwater levels in the West Coast aquifer system, South Africa. Earth Sci Inf 18. https://doi.org/10.1007/s12145-024-01623-w

Jackson CR, Wang L, Pachocka M et al (2016) Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model. Hydrol Process 30:3107–3125. https://doi.org/10.1002/hyp.10850

Jafarzadeh A, Pourreza-Bilondi M, Akbarpour A et al (2021) Application of multi-model ensemble averaging techniques for groundwater simulation: synthetic and real-world case studies. J Hydroinformatics 23:1271–1289. https://doi.org/10.2166/hydro.2021.058

Jafarzadeh A, Khashei-Siuki A, Pourreza-Bilondi M (2022) Performance assessment of model averaging techniques to reduce structural uncertainty of groundwater modeling. Water Resour Manag 36:353–377. https://doi.org/10.1007/s11269-021-03031-x

Khatti J, Samadi H, Grover KS (2024) Estimation of settlement of pile group in clay using soft computing techniques. Geotech Geol Eng 42:1729–1760. https://doi.org/10.1007/s10706-023-02643-x

Khorram S, Jehbez N (2023) A hybrid CNN-LSTM approach for monthly reservoir inflow forecasting. Water Resour Manag 37:4097–4121. https://doi.org/10.1007/s11269-023-03541-w

Kratzert F, Klotz D, Herrnegger M et al (2019) Toward improved predictions in ungauged basins: exploiting the power of machine learning. Water Resour Res 55:11344–11354. https://doi.org/10.1029/2019WR026065

Kumar R, Kumar P, Kumar Y (2022) Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms. Int J Inf Technol 14:359–373. https://doi.org/10.1007/s41870-021-00741-8

Kutty AA, Kucukvar M, Onat NC et al (2023) Measuring sustainability, resilience and livability performance of European smart cities: A novel fuzzy expert-based multi-criteria decision support model. Cities 137. https://doi.org/10.1016/j.cities.2023.104293

Li J, Hao J, Feng QQ et al (2021) Optimal selection of heterogeneous ensemble strategies of time series forecasting with multi-objective programming. Expert Syst Appl 166. https://doi.org/10.1016/j.eswa.2020.114091

Mahmoody M, Dogan A, Engin G (2018) Sustainable surface-subsurface water use in ergene river basin Turkey. Int Congr Environ Model Softw

Masria A, Seif AK, Ghareeb M et al (2023) Numerical modeling of vadose zone electrical resistivity to evaluate its hydraulic parameters. Appl Water Sci 13. https://doi.org/10.1007/s13201-023-02024-y

Masria A, Alshammari TO, Ghareeb M et al (2024) 2D and 3D modeling of resistivity and chargeability to identify the type of saturated groundwater for complex sedimentary facies. Hydrology 11. https://doi.org/10.3390/hydrology11080120

Mienye ID, Swart TG, Obaido G (2024) Recurrent neural networks: A comprehensive review of architectures, variants, and applications. Information 15:517. https://doi.org/10.3390/info15090517

Mirzania E, Roshni T, Ghorbani MA, Heddam S (2024) River water temperature prediction using a hybrid model based on variational mode decomposition (VMD) and outlier robust extreme learning machine. Environ Process 11. https://doi.org/10.1007/s40710-024-00716-4

Momeneh S, Nourani V (2022) Forecasting of groundwater level fluctuations using a hybrid of multi-discrete wavelet transforms with artificial intelligence models. Hydrol Res 53:914–944. https://doi.org/10.2166/nh.2022.035

Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N (2013) A Wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour Manag 27:1301–1321. https://doi.org/10.1007/s11269-012-0239-2

Noman M, Ullah I, Khan MA et al (2024) Analysis of overcurrent protective relaying as minimum adopted fault protection for small-scale hydropower plants. Int J Environ Sci Technol 21:4457–4470. https://doi.org/10.1007/s13762-023-05284-y

Ökten Ş, Yazicigil H (2005) Investigation of safe and sustainable yields for the sandy complex aquifer system in the ergene river basin, Thrace region, Turkey. Turkish J Earth Sci 14:209–226

Paudel G, Pant RR, Joshi TR et al (2024) Hydrochemical dynamics and water quality assessment of the Ramsar-Listed Ghodaghodi lake complex: unveiling the water-Environment nexus. Water 16:1–23. https://doi.org/10.3390/w16233373

Pham BT, Jaafari A, Phong T, Van et al (2021) Naïve Bayes ensemble models for groundwater potential mapping. Ecol Inf 64. https://doi.org/10.1016/j.ecoinf.2021.101389

Pourmorad S, Kabolizade M, Dimuccio LA (2024) Artificial intelligence advancements for accurate groundwater level modelling: an updated synthesis and review. Appl Sci 14. https://doi.org/10.3390/app14167358

Rahmandad H, Xu R, Ghaffarzadegan N (2021) Enhancing Long-term forecasting: learning from COVID-19 models. SSRN Electron J. https://doi.org/10.2139/ssrn.3906690

Rajaee T, Ebrahimi H, Nourani V (2019) A review of the artificial intelligence methods in groundwater level modeling. J Hydrol 572:336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Ravindran SM, Bhaskaran SKM, Ambat SKN (2021) A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter. Environ Process 8:1567–1599. https://doi.org/10.1007/s40710-021-00543-x

Rehman A, Xue L, Islam F et al (2024) Unveiling groundwater potential in Hangu district, Pakistan: A GIS-driven bivariate modeling and remote sensing approach for achieving SDGs. Water (Switzerland) 16. https://doi.org/10.3390/w16223317

Roy DK, Datta B (2018) A review of surrogate models and their ensembles to develop saltwater intrusion management strategies in coastal aquifers. Earth Syst Environ 2:193–211. https://doi.org/10.1007/s41748-018-0069-3

Roy DK, Munmun TH, Paul CR et al (2023) Improving forecasting accuracy of multi-scale groundwater level fluctuations using a heterogeneous ensemble of machine learning algorithms. Water (Switzerland) 15. https://doi.org/10.3390/w15203624

Saini VK, Kumar R, Al-Sumaiti AS et al (2023) Learning based short term wind speed forecasting models for smart grid applications: an extensive review and case study. Electr Power Syst Res 222. https://doi.org/10.1016/j.epsr.2023.109502

Saqr AM, Nasr M, Fujii M et al (2025) Integrating MODFLOW-USG and Walrus optimizer for estimating sustainable groundwater pumping in Arid Regions subjected to severe drawdown. In Ali S, Negm A, Groundw Dev Countries Springer Water. Springer, Cham. https://doi.org/10.1007/978-3-031-79122-2_5

Sathiyamurthi S, Youssef YM, Gobi R et al (2025) Optimal land selection for agricultural purposes using hybrid geographic information system–fuzzy analytic hierarchy process–geostatistical approach in Attur Taluk, India: synergies and trade-offs among sustainable development goals. Sustain 17. https://doi.org/10.3390/su17030809

Shen C (2018) A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour Res 54:8558–8593. https://doi.org/10.1029/2018WR022643

Singh D, Singh B (2020) Investigating the impact of data normalization on classification performance. Appl Soft Comput 97. https://doi.org/10.1016/j.asoc.2019.105524

Subbarayan S, Youssef YM, Singh L, et al (2025) Soil and water assessment tool-based prediction of runoff under scenarios of land use/land cover and climate change across Indian Agro-climatic zones: implications for sustainable development goals. Water (Switzerland) 17 https://doi.org/10.3390/w17030458

Timalsina R, Acharya S, Đurin B et al (2025) An assessment of seasonal water quality in Phewa lake, Nepal, by integrating geochemical indices and statistical techniques: a sustainable approach. Water (Switzerland) 17. https://doi.org/10.3390/w17020238

Timilsina B, Kharel M, Shrestha R et al (2025) Assessment of soil quality along the elevation gradient of the Seti river watershed in Pokhara metropolitan City, Nepal. Environ Monit Assess 197. https://doi.org/10.1007/s10661-025-13716-0

Tokatlı C, Varol M (2021) Impact of the COVID-19 lockdown period on surface water quality in the Meriç-Ergene river basin, Northwest Turkey. Environ Res 197. https://doi.org/10.1016/j.envres.2021.111051

Wani OA, Mahdi SS, Yeasin M et al (2024) Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas. Sci Rep 14:27876. https://doi.org/10.1038/s41598-024-77687-x

Waqas M, Humphries UW (2024) A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX 102946

Wei Y, Wang F, Hong B, Yang S (2023) Revealing Spatial variability of groundwater level in typical ecosystems of the Tarim basin through ensemble algorithms and limited observations. J Hydrol 620. https://doi.org/10.1016/j.jhydrol.2023.129399

Wunsch A, Liesch T, Broda S (2021) Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). Hydrol Earth Syst Sci 25:1671–1687. https://doi.org/10.5194/hess-25-1671-2021

Yang X, Zhang Z (2022) A CNN-LSTM model based on a Meta-Learning algorithm to predict groundwater level in the middle and lower reaches of the Heihe river, China. Water (Switzerland) 14. https://doi.org/10.3390/w14152377

Yoon H, Jun SC, Hyun Y et al (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J Hydrol 396:128–138. https://doi.org/10.1016/j.jhydrol.2010.11.002

Authors and Affiliations

Ahmed M. Saqr[1] · Veysi Kartal[2] · Erkan Karakoyun[3] · Mahmoud E. Abd-Elmaboud[1]

✉ Ahmed M. Saqr
ahmedsaqr@mans.edu.eg

✉ Mahmoud E. Abd-Elmaboud
mahmoud_elsayed@mans.edu.eg

Veysi Kartal
vkartal@firat.edu.tr

Erkan Karakoyun
e.karakoyun@alparslan.edu.tr

[1] Present address: Irrigation and Hydraulics Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

[2] Present address: Department of Civil Engineering, Engineering Faculty, Siirt University, Siirt 56100, Türkiye

[3] Present address: Faculty of Engineering and Architecture, Mus Alparslan University, Mus 49250, Türkiye