Applied Computing and Geosciences 28 (2025) 100303

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

# A systematic review of machine learning models for groundwater level prediction

Gilbert Jesse [a], Cyril D. Boateng [b,c,*], Jeffrey N.A. Aryee [a], Marian A. Osei [a,d,e], David D. Wemegah [b], Solomon S.R. Gidigasu [f], Akyana Britwum [b], Samuel K. Afful [g], Haoulata Touré [f], Vera Mensah [f], Prinsca Owusu-Afriyie [a]

[a] Department of Meteorology and Climate Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[b] Department of Physics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[c] Caburu Company Ltd, P.O.BOX MD 2046, Madina, Accra, Ghana
[d] School of Earth and Environment, University of Leeds, Leeds, UK
[e] Centre for Ecology and Hydrology, Crowmarsh Gifford, Oxfordshire, UK
[f] Department of Geological Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[g] Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

ARTICLE INFO

*Keywords:*
Machine learning
Groundwater level prediction
PRISMA
ANN
Feature selection
Optimization
Hyperparameter tuning

ABSTRACT

This study presents a comprehensive synthesis of machine learning (ML) techniques applied to groundwater level (GWL) prediction, focusing on model architectures, feature selection methods, hyperparameter tuning, optimization algorithms, and clustering techniques. A total of 223 peer-reviewed articles were systematically reviewed using the PRISMA framework to guide study identification, inclusion, and exclusion. Widely used models include artificial neural networks (ANN), support vector machines (SVM), long short-term memory networks (LSTM), and random forests (RF). More recent studies increasingly employ hybrid approaches that integrate wavelet transforms, signal decomposition, and optimization techniques such as particle swarm optimization (PSO), genetic algorithms (GA), and ant colony optimization (ACO). Transformer-based models have also begun to emerge as promising tools in this domain. A central focus of this review is feature selection, which remains one of the most underdeveloped areas in GWL modeling. Most studies rely on simple filter methods like autocorrelation and mutual information. While SHapley Additive exPlanations (SHAP) has gained some traction, more advanced techniques, such as recursive feature elimination (RFE), forward feature selection (FFS), factor analysis (FA), and self-organizing maps (SOM), are rarely used. Notably, no study systematically compared multiple feature selection strategies, limiting insights into their impact on model performance. Scientometric analysis shows that Iran, China, India, and the United States contribute the most impactful research. Despite strong predictive outcomes, trial-and-error remains the dominant approach to hyperparameter tuning. The review emphasizes the need for more systematic, interpretable, and generalizable ML approaches to support robust groundwater level (GWL) forecasting.

## 1. Introduction

Groundwater is a critical global resource that supplies clean water to more than two billion people (Famiglietti, 2014). However, increasing demand and unsustainable exploitation, particularly in developing regions, have strained aquifers, with projections indicating that nearly 20% are already overexploited and most could face the same fate by 2050 (Piesse, 2020). Groundwater levels (GWL), the depth from the surface to the saturated zone, are vital indicators of aquifer health and are measured through monitoring wells. Analyzing GWL fluctuations is crucial for managing groundwater resources amidst growing demand and climate change impacts (Butler et al., 2013).

Traditional groundwater modeling often relies on conceptual or physically based models such as MODFLOW and HydroGeoSphere (Brunner and Simmons, 2012; Chakraborty et al., 2020; Dehghani et al., 2022). These physically based models require extensive data on aquifer properties like transmissivity, hydraulic conductivity, and recharge rates, datasets that are frequently unavailable in data-scarce or poorly instrumented regions. Machine learning (ML) has emerged as a robust alternative for modeling GWL dynamics due to its ability to learn

* Corresponding author at: Department of Physics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
*E-mail address:* cyrilboat@knust.edu.gh (C.D. Boateng).

https://doi.org/10.1016/j.acags.2025.100303
Received 11 January 2025; Received in revised form 3 October 2025; Accepted 27 October 2025
Available online 4 November 2025
2590-1974/

complex, non-linear relationships directly from observed data (Nourani et al., 2011; Kalu et al., 2022; Maiti and Tiwari, 2014). A wide array of ML techniques, ranging from traditional methods like Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) to deep learning architectures such as Long Short-Term Memory (LSTM) and Deep Belief Networks (DBNs), have been deployed with promising results. For instance, Nourani et al. (2011) used a hybrid ANN model to predict GWLs in East Azerbaijan, Iran, while Kalu et al. (2022) applied a DBN to model groundwater fluctuations in southern Africa.

While our focus is on ML for GWL forecasting, ML has also been increasingly applied in related domains such as groundwater potential mapping and vulnerability assessment (e.g., delineating recharge zones using ML-MCDM models (Kanji and Das, 2025) and improving DRASTIC-based vulnerability indices (Dasgupta et al., 2024)). In parallel, non-ML statistical and GIS-based methods have also been employed for groundwater stress mapping, such as the GIS-statistical workflows presented by John and Das (2020) and John et al. (2023). Together, these applications highlight the breadth of data-driven approaches being explored in hydrogeology.

However, despite these advances, ML applications in GWL modeling still face critical limitations. Overfitting, underfitting, and poor generalization, particularly in deep learning frameworks, remain recurring issues, often exacerbated by the inclusion of noisy or irrelevant input features (Kuhn et al., 2013a; Sharghi et al., 2022). This highlights the vital role of feature selection: identifying a minimal yet informative subset of input variables that preserves predictive performance while reducing model complexity. Inadequate feature selection not only inflates computational costs, but can also obscure physical interpretability and worsen generalization in unseen conditions.

In other scientific fields such as genomics, remote sensing, and finance, systematic feature selection has been shown to substantially improve model performance and transparency (Pudjihartono et al., 2022; Dokeroglu et al., 2022; Iranzad and Liu, 2024). However, in GWL modeling, no prior review has comprehensively examined the methodologies and implications of feature selection despite its proven importance. Existing reviews in hydrology and groundwater modeling (Solomatine, 2006; Wu et al., 2014; Rajaee et al., 2019; Tao et al., 2022; Boo et al., 2024) tend to provide broad overviews of machine learning applications but often overlook critical methodological aspects. Specifically, they do not categorize or analyze feature selection methods, rarely discuss how time lags are determined, and do not distinguish between filter, wrapper, and embedded feature selection approaches.

To address this critical gap, this review provides a systematic and detailed assessment of ML-based groundwater level modeling, with a specific emphasis on the role of feature selection, model architecture, optimization strategies, and hyperparameter tuning. Using the PRISMA framework, we analyze recent peer-reviewed studies to identify trends, innovations, and research gaps.

This review is guided by the following core inquiries:

1. What are the most commonly employed ML models in GWL modeling?
2. What strategies are used for feature selection, and how do they affect model performance?
3. How are model architectures structured and optimized in current research?
4. How are hyperparameters selected and tuned across different ML models?

By focusing on these aspects, our review not only synthesizes existing knowledge but also provides actionable insights into constructing more reliable, interpretable, and efficient models for GWL prediction. Unlike prior reviews that focus broadly on ML adoption in hydrology, our work uniquely foregrounds feature selection as a methodological pillar, setting a new benchmark for future research in this domain. We do not delve into the mathematical derivations of the models; instead, we focus on their design principles, input structuring, optimization strategies, and performance implications, offering practical value to both researchers and practitioners.

## 2. Methodology

This systematic review adopts the PRISMA framework (Liberati et al., 2009; Page et al., 2021) to ensure a comprehensive, unbiased evaluation of the research literature. Following the PRISMA flow diagram, we minimized researcher bias and maintained a traceable review process.

The review aimed to identify the most common ML models used for GWL prediction, methods to determine optimal input variables, and techniques to fine-tune key hyperparameters to enhance model performance. Through this exploration, we sought to provide insights into factors influencing the accuracy of GWL predictions.

Key stages of the systematic review process included identification, screening, eligibility assessment, and full-text evaluation, as outlined in Fig. 1. By adhering to the PRISMA methodology, we ensured transparency, rigor, and replicability, enabling a thorough analysis of the existing literature and meaningful conclusions.

1. **Identification**: Following the PRISMA conceptual framework, our systematic review utilized a comprehensive search strategy to capture relevant articles from prominent electronic databases, including Scopus, Science Direct, and Google Scholar. The search was limited to English-language publications between 2010 and 2024, aiming to examine the contemporary applications of ML models in predicting groundwater levels and availability. The search strategy employed a carefully designed search string, incorporating specific terms such as (“groundwater level prediction” OR “water table prediction”) AND (“machine learning” OR “artificial intelligence” OR “AI”) AND (forecasting OR modeling). In total, our search yielded 426 articles in all selected databases, demonstrating the wide scope of literature available on this topic. The details of the search string can be found in Table 1.
2. **Screening**: Following the identification of relevant publications, key metadata, including title, keywords, abstract, DOI, publication year, and author names, was extracted and recorded in a Microsoft Excel spreadsheet. All articles were imported into Mendeley, a reference management software used to streamline literature organization and deduplication (Mendeley, 2022). After removing 68 duplicate records, 358 articles were retained for screening. These articles were evaluated against predefined inclusion and exclusion criteria, resulting in the exclusion of 62 papers whose titles or abstracts were determined to be out of scope. To ensure objectivity and consistency, two independent reviewers assessed the eligibility of the remaining articles. In cases of disagreement or ambiguity, article titles were discussed with external experts in groundwater or machine learning research. Final inclusion decisions were based on consensus between reviewers or expert resolution when needed.
3. **Eligibility criteria**
   Studies were included if they:
   (a) Focused on the application of machine learning algorithms for simulating groundwater levels and availability.
   (b) Employed meteorological parameters as inputs in the models.
   (c) Reported original research (e.g., case studies, simulations, empirical analyses) with sufficient methodological detail and performance metrics.
   (d) Were published in peer-reviewed journals in English between 2010 and 2024.

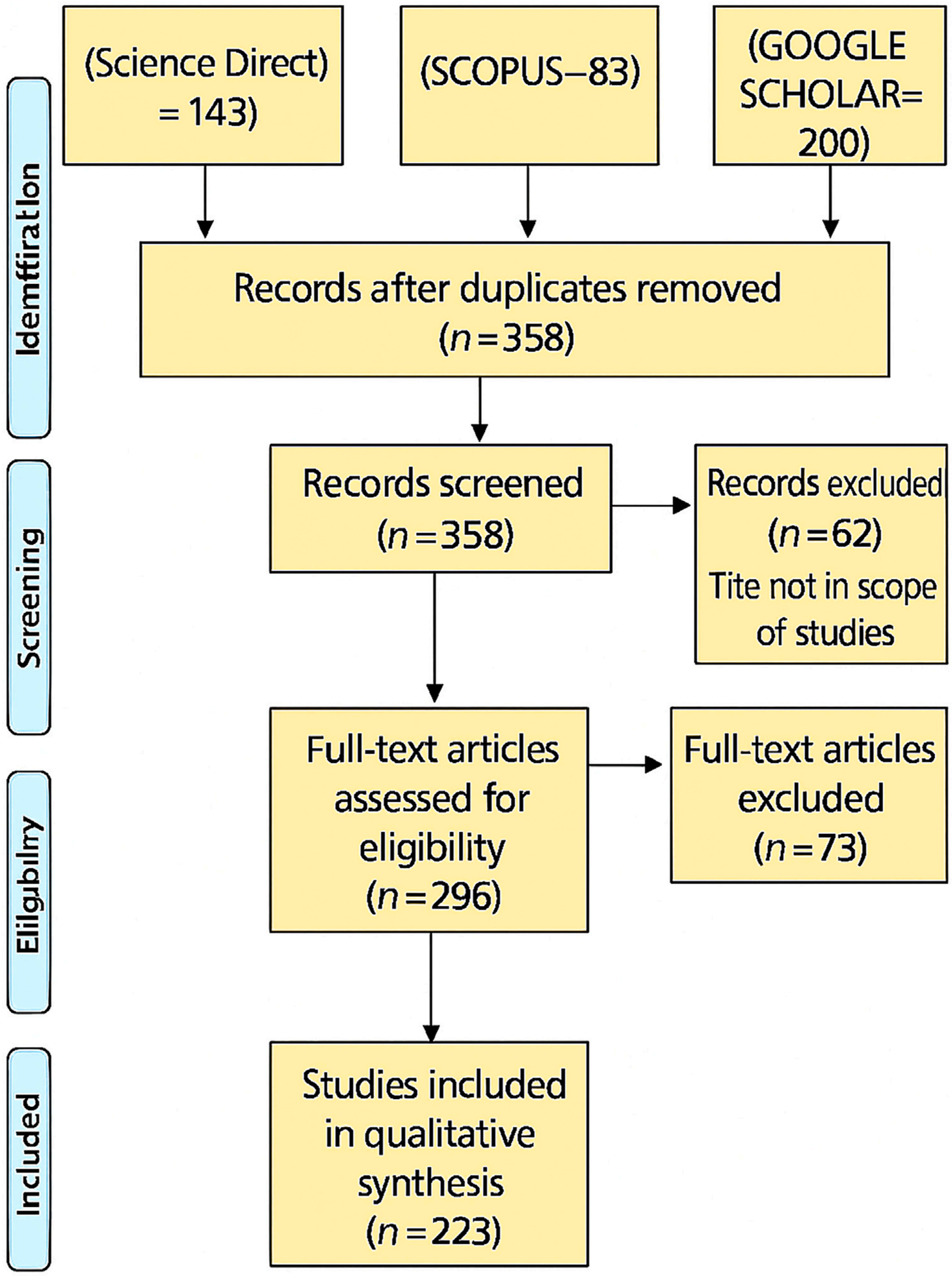


**Fig. 1.** The PRISMA workflow diagram.

Studies were excluded if they:

(a) Relied solely on traditional statistical or physically based models without ML components.
(b) Were secondary sources (reviews, editorials, opinions, conference abstracts, or book chapters).
(c) Did not provide sufficient methodological or results information to assess model development and performance.

## 3. Results and discussion

### 3.1. Publication trends

As shown in Fig. 2, the application of ML for GWL prediction has seen a marked upward trend, particularly after 2019. Between 2010 and 2018, annual publication counts remained relatively modest, fluctuating between 2 and 13 papers per year and collectively accounting for less than 30% of the total reviewed studies. However, from 2020, research activity intensified significantly. In 2022 alone, 34 studies were published, representing approximately 15.1% of all 223 studies reviewed, while 2023 saw a comparable count of 39 papers (17.5%), and 2024 (year-to-date) currently leads with 43 publications (19.3%). This steady growth reflects a growing academic and practical interest in leveraging ML techniques to address groundwater-related challenges. These findings align with recent bibliometric reviews (Afrifa et al., 2022; Tao et al., 2022; Ahmadi et al., 2022; Boo et al., 2024), which also report a sharp increase after 2020 in data-driven groundwater modeling studies. Fig. 3 presents the distribution of total and average citations among the top academic publishers that contribute to the groundwater level (GWL) modeling literature. Elsevier emerged as the most influential publisher, with a total of 3191 citations, resulting in

**Table 1**
The detailed search query.

| Search query |
|---|
| TITLE-ABS-KEY ("roundwater AND level AND prediction" OR "groundwater AND level AND forecasting" AND "machine AND learning") AND PUBYEAR > 2009 AND PUBYEAR ≤ 2024 AND (EXCLUDE (LANGUAGE, "Chinese") OR EXCLUDE (LANGUAGE, "Korean")) AND (LIMIT-TO (EXACTKEYWORD, "Forecasting") OR LIMIT-TO (EXACTKEYWORD, "Groundwater") OR LIMIT-TO (EXACTKEYWORD, "Machine Learning") OR LIMIT-TO (EXACTKEYWORD, "Groundwater Resources") OR LIMIT-TO (EXACTKEYWORD, "Artificial Neural Network") OR LIMIT-TO (EXACTKEYWORD, "Rain") OR LIMIT-TO (EXACTKEYWORD, "Hydrology") OR LIMIT-TO (EXACTKEYWORD, "Evapotranspiration") OR LIMIT-TO (EXACTKEYWORD, "Temperature") OR LIMIT-TO (EXACTKEYWORD, "Groundwater Level Fluctuation") OR LIMIT-TO (EXACTKEYWORD, "Aquifer") OR LIMIT-TO (EXACTKEYWORD, "Evaporation")) |

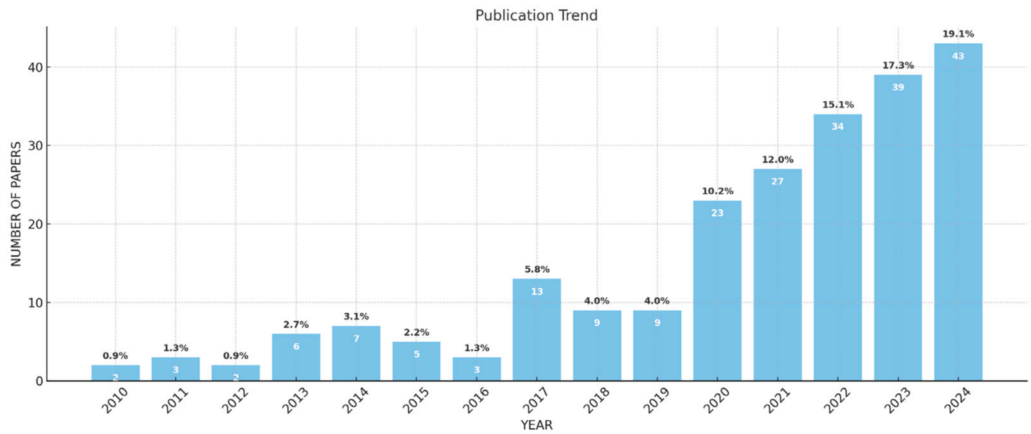


**Fig. 2.** Trends in publications within the reviewed articles from 2010 to 2024.

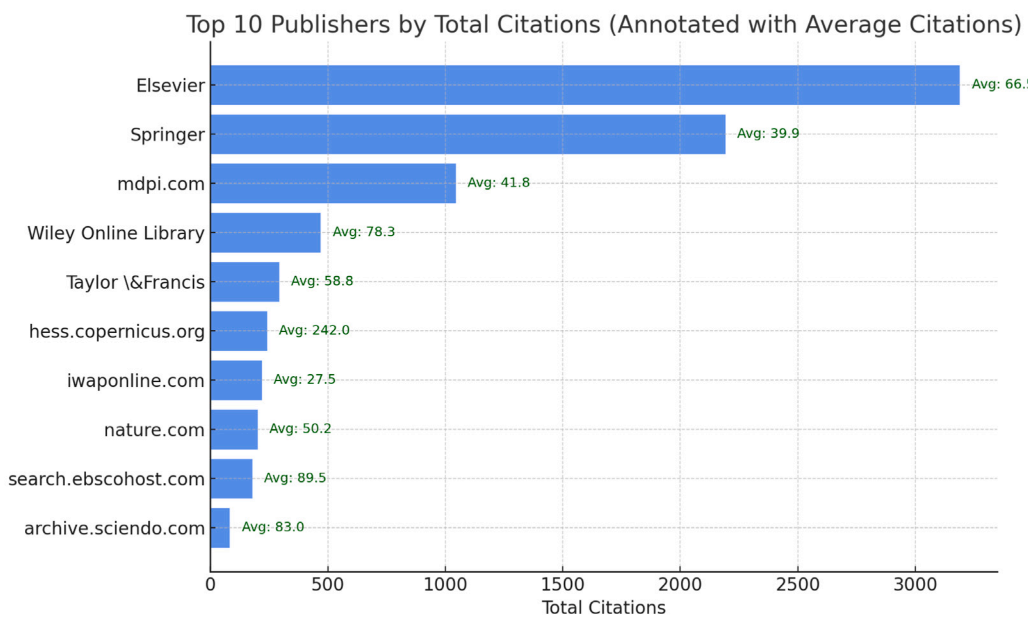


**Fig. 3.** Publication outlet average normal citations.

an average of approximately 66.5 citations per study. Springer accumulated 2192 citations with an average of 39.9 citations per article. MDPI followed with 1045 total citations (41.8 average citations per study). Wiley Online Library, while contributing fewer studies (6), achieved a higher average impact with 78.3 citations per article, indicating a strong influence relative to volume. Taylor & Francis, IWA Publishing (iwaponline.com), and Nature also demonstrated moderate to high average citation metrics. Notably, some specialized or lower-volume platforms such as Copernicus (hess.copernicus.org) and EBSCOhost showcased high average citations per article, although based on limited publication counts.

Our systematic review analyzed 223 articles, revealing distinct authorship patterns and geographic distribution. Iran led with 39% of publications, followed by China (24%) and the United States (13%), as shown in Fig. 4. Many studies in Iran address its dependence on groundwater due to limited surface water resources, highlighting concerns about the depletion of aquifers (Sharafati et al., 2020; Motagh et al., 2017; Milan et al., 2023; Arabameri et al., 2019; Moravej et al., 2020). In general, Asia dominated the research landscape, contributing 81.24% of the studies. In contrast, the use of ML for groundwater modeling in Africa remains underrepresented, with little literature highlighting a significant research gap in applying ML techniques to groundwater dynamics on the continent. Fig. 4 summarizes the geographical distribution. Fig. 5 presents a country-level co-authorship and a normalized citation network derived from VOSviewer. The graph reveals clear regional and international collaboration clusters, with India,

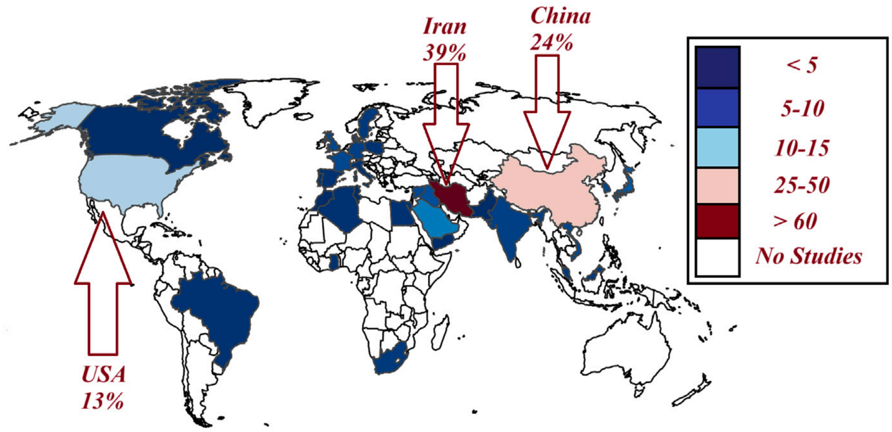


**Fig. 4.** Distribution of the reviewed papers based on the country of origin.

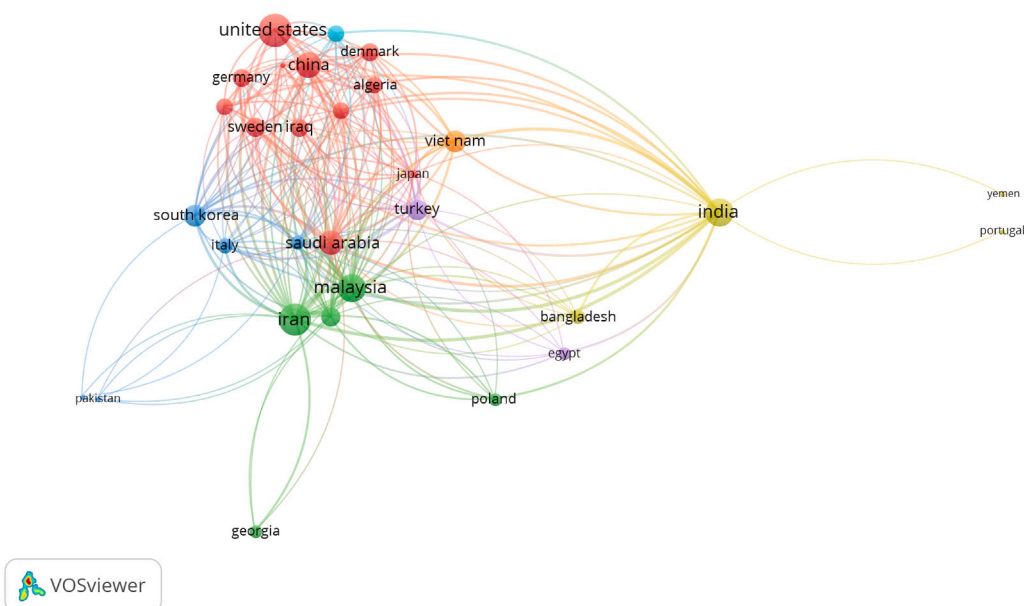


**Fig. 5.** Country-level co-authorship and normalized citation network for studies on machine learning-based groundwater level modeling, generated using VOSviewer.

Iran, and Malaysia serving as major hubs. In particular, India forms a distinct cluster of cooperation with countries such as Portugal, Yemen, and Bangladesh, while Iran exhibits strong ties within the Middle East and Central Asia. The United States and China are central to a western collaboration cluster, linking with Germany, Sweden, and Algeria. The size of the node reflects the strength of the citation, indicating that countries such as the United States, Iran, China, and India not only publish frequently, but also produce highly cited research in GWL modeling.

### 3.2. Algorithms used in groundwater level prediction

This review explores techniques for predicting GWL, with a focus on the most widely applied machine learning algorithms. As illustrated in Fig. 6, five dominant models: ANN (51%), SVM (13%), LSTM (12%), tree-based ensembles (8%) and ANFIS (5%) collectively account for 89% of the models reviewed. These approaches have been used independently and in hybrid configurations in diverse case studies. The following sections critically examine methodological elements including feature selection, model design, data pre-processing, and parameter tuning.

### 3.3. ANN (Standalone and hybrid) models

#### 3.3.1. Bibliographic review

Experts worldwide have conducted numerous studies on groundwater level prediction across various geographical regions using ANNs, and these studies have consistently reported the efficacy of these models. For example, Dash et al. (2010) developed a hybrid ANN-GA (Genetic Algorithm) model to predict GWLs in the Mahanadi River Basin, India, outperforming standalone ANN models trained with Levenberg–Marquardt (LM), gradient descent, and Bayesian Regularization (BR) algorithms. Hyperparameters and activation functions were optimized via trial and error, with the hybrid model demonstrating superior predictive performance. Chen et al. (2010) employed self-organizing maps (SOM) to predict GWLs in the Zhuoshuixi River Basin, Taiwan, using single-site and multisite models. ACF and PACF determined input

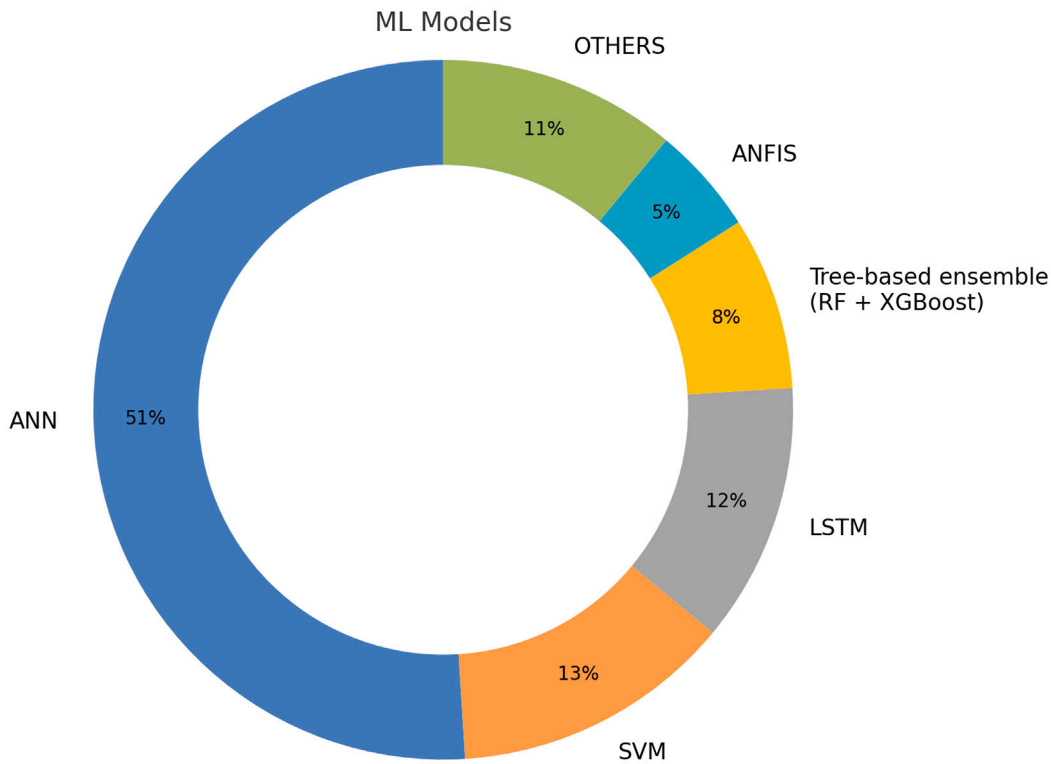


**Fig. 6.** ML models utilized in GWL modeling.

features, and the multisite model outperformed single-site configurations. Building on this, Chen et al. (2011) combined SOM with backpropagation networks (BPN), developing six models. The improved multisite SOM-BPN model outperformed ARIMA and single-site ANN models. Rakhshandehroo et al. (2012) compared FFNN, Radial Basis Function (RBF), Elman (ELNN), and generalized regression neural networks (GRNN) for monthly GWL prediction in the Shiraz plain, Iran. All models performed effectively, with FFNN achieving the best results. Adamowski and Chan (2011) integrated ANN with discrete wavelet transform (DWT) for GWL forecasting in the Chateauguay watershed, Quebec. DWT decomposed time series data, and the WA-ANN model, trained with the LM algorithm, outperformed standalone ANN and ARIMA models.

Nourani et al. (2011) integrated ANN and geostatistics to predict GWL in eastern Azerbaijan, Iran. Sensitivity analysis guided input selection, and FFNNs with optimized hidden neurons, trained using the LM algorithm, achieved superior performance. Taormina et al. (2012) applied FFNNs for hourly GWL simulations in the Lagoon of Venice, Italy, utilizing AutoRegressive with eXogenous inputs (ARX) models and the Akaike Information Criterion for input selection. Optimized via trial and error, the models effectively simulated GWL over extended periods.

A few years after Chen et al. (2010) used SOM for determining hidden layer neurons in RBFN models, Nourani et al. (2015) combined SOM-based clustering and wavelet transform (WT) to model one- and multi-step-ahead GWLs in the Aradabil plain, Iran. SOM identified homogeneous clusters, and WT extracted multi-scale features from non-stationary GWL, runoff, and rainfall time series. The FFNN model, trained via backpropagation with feature selection and lagged values guided by mutual information (MI), demonstrated improved accuracy and reduced model complexity. Bahaa et al. (2015) compared MLR, ANN, wavelet MLR, wavelet ANN, and a wavelet ensemble ANN for GWL forecasting in Quebec, Canada. The "à Trous" wavelet transform was used for de-noising, and MI guided feature selection. The wavelet ensemble ANN consistently outperformed other models across all lead times. Chang et al. (2015) developed ANN models to simulate supra-permafrost GWLs in the Qinghai-Tibet Plateau, China, using two and three input variable configurations. A three-layer feedforward network trained with the LM algorithm showed higher accuracy with three inputs, while the two-input model performed reasonably well without field GWL data.

Gholami et al. (2015) used a multilayer perceptron (MLP) network with dendrochronology and precipitation data to simulate GWLs in the Caspian Sea's alluvial aquifer (1912–2013). A standard three-layer feedforward neural network, trained using LM backpropagation, showed that the sigmoid transfer function provided the best results during the growing season.

Sun et al. (2016) employed a standard FFNN with a quasi-Newton training algorithm to predict GWLs in a swamp forest in Singapore. Inputs included reservoir levels and rainfall. Using logistic and threshold activation functions, the ANN achieved accurate forecasts up to 7 days ahead, though performance decreased with longer lead times.

Choubin and Malekian (2017) compared ANN and ARIMA models for simulating GWLs in Iran's Shiraz basin using precipitation, streamflow, temperature, evaporation, and GWL data. The LM algorithm trained the ANN with logistic-sigmoid and purelinear activation functions. ARIMA outperformed ANN based on the evaluation metrics. Wen et al. (2017) also compared a wavelet analysis-based ANN (WA-ANN) with a traditional ANN model for forecasting GWLs in an arid inland river basin in northwestern China. Input variables included GWLs, total precipitation, evaporation, and average temperature. Both models used a three-layer architecture with a single-neuron output layer, sigmoid activation for hidden layers, and a linear activation for the output layer. The WA-ANN consistently outperformed the traditional ANN model across all evaluation metrics.

Sahoo et al. (2017) developed a hybrid artificial neural network (HANN) model for GWL prediction, incorporating innovative covariate processing methods. Singular spectrum analysis decomposed time series data, while mutual information and genetic algorithms identified key components. The model used logistic sigmoid and linear activation functions in a single hidden layer, optimized through trial and error, and trained with the LM algorithm. HANN outperformed hybrid linear and nonlinear regression models. Similarly, El Ibrahimi et al. (2017) integrated DWT with ANN-PMC for GWL prediction in Morocco, using a three-layer architecture with precipitation, temperature, and GWLs as covariates. Systematic parameter optimization showed the DWT-ANN-PMC model exceeded standalone ANN-PMC and MLR models in accuracy.

Guzman et al. (2017) used the Nonlinear autoregressive exogenous model (NARX) network to predict daily groundwater levels (GWLs) in the Mississippi River Valley Alluvial (MRVA) aquifer in the southeastern United States. The model, trained with both LM and BR algorithms, used precipitation and GWL as covariates. Sensitivity analysis identified historical GWL and precipitation as key variables. The optimal network consisted of two hidden nodes with sigmoid transfer functions and one output node with a linear function. The NARX-BR model achieved the highest predictive accuracy among the tested models. Wunsch et al. (2018) employed the NARX model for GWL prediction in southwest Germany, identifying influential time lags using STL decomposition and auto/cross-correlation. The study demonstrated NARX's accuracy and adaptability for GWL modeling. Similarly, Kouziokas et al. (2018) implemented multilayer FFNN models with varying architectures to forecast GWLs in Montgomery County, Pennsylvania. Using the LM algorithm with Tanh-Sigmoid transfer functions in hidden layers, the models achieved optimal predictive performance compared to other training algorithms and transfer functions.

A year later, Lee et al. (2019) developed FFNN models to predict GWLs in South Korea's Yangpyeong riverside area. Correlation analysis identified surface water level as the most influential covariate. The models used logistic-sigmoid and linear activation functions in the hidden and output layers, respectively, and were trained using backpropagation, achieving strong predictive performance. Chen et al. (2020) compared MLP, RBF, MODFLOW, and SVM models for GWL prediction in the Heihe River Basin. The MLP employed a single hidden layer with optimized neurons, using backpropagation and gradient descent. Results showed that SVM and RBF outperformed MODFLOW in accuracy. Banadkooki et al. (2020) assessed RBF neural networks with Whale Algorithm (WA), MLP-WA, and GP models for GWL prediction in Yazd–Ardakan, Iran. Cross-correlation and partial autocorrelation guided feature selection, with MLP-WA demonstrating superior predictive performance. Similarly, Yadav et al. (2020) combined singular spectrum analysis (SSA), MI, GA, ANN, and SVM models to analyze groundwater fluctuations in India, considering climatic and non-climatic factors. Hybrid models (SSA-MI-GA-ANN and SSA-MI-GA-SVM) consistently outperformed standalone models.

Müller et al. (2021) compared LSTM, MLP, RNN, and CNN models for GWL prediction in Butte County, California, using surrogate-based algorithms and random sampling for hyperparameter tuning. MLP, optimized with ADAM and ReLU activation, outperformed the other models, emphasizing the importance of precise hyperparameter tuning. Sharghi et al. (2022) evaluated FFNN, ANFIS, LSTM, and GMDH models for multi-step GWL forecasting in Iran. Pre- and post-processing techniques like COMUSA and NAE improved accuracy, with GMDH outperforming others, highlighting the benefits of clustering and ensemble approaches. Yin et al. (2021) introduced a Bayesian ensemble model integrating ANN, SVM, and Response Surface Regression (RSR) to predict groundwater storage fluctuations. The ANN model, trained with the LM algorithm, demonstrated low uncertainty, particularly at sub-regional scales. Bayesian model averaging provided the most accurate predictions. Kalu et al. (2022) utilized a Deep Belief Network (DBN) to simulate monthly GWLs in southern Africa, using hydrological variables and global climate indices. Variable selection employed correlation analysis, and the architecture was optimized through trial and error. Results highlighted DBN's efficacy in GWL prediction. Collados-Lara et al. (2023) applied NAR, NARX, and Elman Neural Networks for short-term GWL prediction in Spain. Effective precipitation emerged as a key predictor, with NARX and Elman Neural Networks outperforming others. Van Thieu et al. (2023) proposed the Augmented Artificial Ecosystem Optimization-based Multi-Layer Perceptron (AAEO-MLP) model for monthly GWL prediction in India, using ELU activation and mutual information for input selection. The AAEO-MLP consistently outperformed other MLP models. Panahi et al. (2023) compared Radial Basis Function Neural Network (RBF), ANFIS, SVM, and MLP for GWL prediction under future climate conditions in Iran. RBFNN exhibited superior performance, proving reliable for climate-related GWL forecasts. Wei et al. (2023a) proposed a hybrid WT-PSR-ANN model for GWL forecasting by combining wavelet decomposition, phase space reconstruction (PSR), and ANN. GWL time series were decomposed using three types of mother wavelets, and PSR was applied to select relevant sub-series as input. Lyapunov exponent analysis confirmed chaotic behavior in the data. The WT-ANN outperformed standard ANN, while WT-PSR-ANN yielded the best results overall. The study highlighted PSR as a valuable enhancement to wavelet-based preprocessing for improving model accuracy. Ghafoor et al. (2024) applied autoregressive moving average (ARIMA) and ANN models to predict GWLs across four wells in Cheyenne and Delta counties, Colorado, USA. Monthly GWL data were preprocessed via interpolation and up-sampling to enhance model training. The ANN employed a feedforward architecture with rectified linear unit (ReLU) and linear activation functions. Model evaluation revealed that ANN consistently outperformed ARIMA in accuracy across all sites. This result was attributed to ANN's robustness against nonlinearity and noise, making it more adaptable to the irregular GWL fluctuations observed. Fahim et al. (2024) evaluated MLR, SVM, Gaussian process regression (GPR), regression trees, bagged and boosted ensembles, and ANN models for GWL prediction in Bangladesh. Weekly GWL observations, with missing values addressed using cubic spline interpolation. Inputs included Global Land Data Assimilation System (GLDAS) groundwater storage, rainfall, temperature, elevation, irrigation, population, and GRACE-based water storage variability. ANN performed best, effectively modeling spatial groundwater variability. Seifi et al. (2024) proposed a hybrid BFSA-MVMD-GRU-RVM model for GWL prediction in Iran's Bastam Plain. Boruta was used for feature selection, multivariate variational mode decomposition (MVMD) for decomposition, gated recurrent unit (GRU) for sequence learning, and relevance vector machine (RVM) for prediction. Inputs included lagged rainfall, temperature, pumping, irrigation return flow, and humidity. The model outperformed other MVMD-based models and showed reduced prediction uncertainty across short- to long-term forecasts. Moradi et al. (2023) compared a numerical model (GMS) with AI models including GA-ANN, ICA-ANN, Extreme Learning Machine (ELM), Outlier Robust ELM (ORELM), and Group Method of Data Handling (GMDH) for predicting GWL in the Lur Plain, Iran. Initial inputs were GWL, rainfall, temperature, and evaporation, but only GWL was retained due to stronger correlations. ORELM outperformed all other models and the numerical approach. Model evaluation used RMSE, NRMSE, NASH, R, and a Taylor diagram. Abdi et al. (2024) applied CNN and DNN models for GWL prediction in Iran, using 34 wells. Missing data were handled via interpolation (Kriging, SVM, M5P), with M5P proving most accurate. Inputs included spatial coordinates and groundwater levels. CNN outperformed DNN across scenarios, especially when interpolated data were used to expand input coverage. Ghafoor et al. (2024) applied ARIMA and ANN models to forecast GWL at four wells in Colorado, USA. Auto-ARIMA identified optimal parameters for ARIMA. A feedforward ANN with one hidden layer used Adam optimizer and ReLU activation. ANN consistently outperformed ARIMA across all wells. Despite data limitations and a small number of wells, both models showed effectiveness for GWL prediction. Seidu et al. (2023) evaluated four ANN models (BPNN, RBFNN, GRNN, GMDH) for groundwater level prediction using data from 13 boreholes in Ghana. Input features included rainfall, evaporation, and temperature. Min–Max normalization was applied, and model performance was assessed across five train–test splits (90–10 to 50–50). The 70–30 and 80–20 splits produced the most accurate predictions. RBFNN performed best in six out of thirteen boreholes. Akbari Majd et al. (2024) developed ANN-based models hybridized with three metaheuristic algorithms (Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO)) to predict GWL in Iran's Ardabil Plain. Unlike most studies, historical GWL was not used. Inputs included rainfall, temperature, runoff, and discharges. Three preprocessing stages were tested, with Stage 3 involving time series decomposition (trend, seasonality removal) showing the best performance. ANN-GA performed best for some wells, ANN-PSO for others. Overall, Stage 3 improved prediction accuracy by 76%, making the method suitable for data-scarce basins. Feng et al. (2024) evaluated decision tree (DT), RF, SVM, convolutional neural network (CNN), recurrent neural network (RNN), and generative adversarial network (GAN) models for GWL prediction in Izeh City, Iran. The study used extraction rate, rainfall, and river flow as inputs. Feature importance was assessed using Pearson and Spearman correlation, revealing that river flow and extraction had stronger associations with GWL than rainfall. CNN outperformed other models, demonstrating robustness and high accuracy in capturing groundwater fluctuations.

Refer to supplementary information for a detailed description of ANN.

#### 3.3.2. Results

Based on the reviewed papers, we found that

1. The Levenberg–Marquardt (LM) algorithm emerged as the primary optimization method for training ANN models, often enhanced with Bayesian Regularization to improve generalization. The Backpropagation algorithm was utilized to compute gradients for weight updates. The LM algorithm is recognized for its computational efficiency, balancing Newton's method, which converges rapidly near minima but risks divergence, with gradient descent, which ensures convergence but at a slower rate if step sizes are chosen carefully (Tyagi et al., 2022). Researchers have highlighted the LM method's computational advantages and its reduced likelihood of becoming trapped in local minima, making it a robust and effective choice for training ANN models (Daliakopoulos et al., 2005).
2. A typical ANN structure consists of three layers, incorporating the sigmoid activation function in the hidden layer and a linear activation function in the output layer. Some studies also employed ReLU and ELU. Notably, in most of the reviewed papers, the determination of the ANN structure, the number of hidden neurons, and hyper-parameter tuning were achieved using a trial-and-error approach. However, genetic programming (GP) was used for hyperparameter optimization in specific cases. The reported learning rates mainly ranged between 0.001 and 0.009, with the highest recorded as 0.01 (Emamgholizadeh et al., 2014).

   Activation function plays a crucial role in the successful training of deep neural networks. It introduces non-linearity into the neural network model, enabling the network to learn more effectively by capturing and understanding the intricate non-linear patterns present in input datasets. Table 2 summarizes the general output range, along with some advantages and disadvantages of the various activation functions used in GWL modeling. Further details on these functions can be in the works of LeCun et al. (2015), Banerjee et al. (2019), Ramachandran et al. (2017), Jamel and Khammas (2012), Narayan (1997), Kalaiselvi et al. (2022), Ding et al. (2018), Nwankpa et al. (2018) and Rasamoelina et al. (2020)
3. The selection of optimal input variables primarily relied on correlation analyses, including Partial Auto-correlation, Auto-correlation, and Cross-Correlation function. In some cases, MI was used to improve model performance through enhanced feature selection.

   Auto-correlation gauges how similar a time series is to a lagged version of itself, helping spot repeating patterns or trends at different time points in the same series. Partial Auto-correlation is similar but removes the influence of intermediate lagged values.

   Cross-correlation compares two different time series at the same time points, identifying relationships and potential lead–lag patterns between them.

   MI quantifies the mutual information shared between two variables, offering an indirect means to identify pertinent time lags by evaluating information content across various temporal shifts (Van Thieu et al., 2023). Table 3 provides a summary of the formulas, advantages, and disadvantages of the feature selection methods employed. Further details on these methods can be found in the works (Kraskov et al., 2004; Veyrat-Charvillon and Standaert, 2009; Batina et al., 2011; Latham and Roudi, 2009; Bourke, 1996; Yoo and Han, 2009; Ramsey, 1974)
4. Among the hybrid models, the combinations consistently yield optimal performance involving ANN coupled with WT.

**Table 2**
Comparison of activation functions (LeCun et al., 2015; Banerjee et al., 2019; Ramachandran et al., 2017; Jamel and Khammas, 2012; Narayan, 1997; Kalaiselvi et al., 2022; Ding et al., 2018; Nwankpa et al., 2018; Rasamoelina et al., 2020).

| Activation function | Sigmoid | Tanh | ReLU | ELU |
|---|---|---|---|---|
| Output range | (0, 1) | (−1, 1) | [0, ∞) | [−1, ∞) |
| Smooth gradient | Yes | Yes | No | No |
| Vanishing gradient issue | Yes | Yes | Yes (for negative inputs) | Reduced |
| Zero-centered output | No | Yes | No | Yes |
| Computational efficiency | Yes | No | Yes | No |
| Dying neurons problem | No | No | Yes | Reduced |

### 3.4. ANFIS (Standalone and hybrid) models

#### 3.4.1. Bibliographic review

When it comes to GWL modeling with Adaptive Neural Networks Fuzzy Inference Systems (ANFIS), Jalalkamali et al. (2011) evaluated the predictive performance of ANFIS and ANN models for groundwater level (GWL) prediction in two neighboring wells in Kerman Plain, Iran. Input variables included rainfall, air temperature, and GWLs. Hyperparameters for both models were optimized through trial and error, with the ANFIS model utilizing the Gaussian membership function (gaussmf). Performance evaluations indicated that ANFIS outperformed ANN in accuracy. Emamgholizadeh et al. (2014) compared ANFIS and ANN for GWL prediction in Bastam Plain, Iran, concluding that ANFIS, with trapezoidal input membership functions and a hybrid learning algorithm, consistently outperformed ANN. Similarly, Maiti and Tiwari (2014) evaluated ANN, Bayesian Neural Network (BNN), and ANFIS for modeling GWL fluctuations in Dindigul, Southern India. The study found ANFIS excelled with noise-free data, while BNN was more effective for noisy hydrological series. The following year, Mirzavand et al. (2015) evaluated the ANFIS model against the support vector regression (SVR) model for predicting monthly GWL fluctuations in the Kashan plain, Iran, concluding that ANFIS with a Bell-shaped MF outperformed SVR. Similarly, Gong et al. (2016) compared ANFIS to ANN and support vector machine (SVM) for forecasting GWLs near Lake Okeechobee, Florida. Input factors were selected using partial and autocorrelation coefficients, and models were tuned via trial-and-error. Results indicated that ANFIS and SVM consistently outperformed ANN.

Seifi et al. (2020) combined six meta-heuristic methods (e.g., grasshopper optimization algorithm (GOA), cat swarm optimization

**Table 3**
Comparison of feature selection methods employed for ANN (Kraskov et al., 2004; Veyrat-Charvillon and Standaert, 2009; Batina et al., 2011; Latham and Roudi, 2009; Bourke, 1996; Yoo and Han, 2009; Ramsey, 1974).

| Method | Formula | Advantages | Disadvantages |
|---|---|---|---|
| Partial Autocorrelation (PACF) | $\mathrm{PACF}(k) = \phi_{kk}$ | - Measures the direct effect of a lagged variable<br>- Helps determine appropriate lag in time series models | - Assumes linearity<br>- Sensitive to noise and outliers |
| Autocorrelation (ACF) | $\rho_k = \frac{\mathrm{Cov}(X_t, X_{t+k})}{\mathrm{Var}(X_t)}$ | - Detects repeating patterns and serial correlation<br>- Easy to compute and interpret | - Only captures linear dependence<br>- Can be misleading for non-stationary or non-linear data |
| Cross-Correlation (CCF) | $\mathrm{CCF}(k) = \frac{\mathrm{Cov}(X_t, Y_{t+k})}{\sigma_X \sigma_Y}$ | - Identifies lead–lag relationships across variables<br>- Useful in multivariate time series | - Does not account for indirect effects<br>- Assumes stationarity and linearity |
| Mutual Information (MI) | $I(X;Y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$ | - Captures non-linear dependencies<br>- More general than correlation<br>- Model-agnostic | - Computationally expensive<br>- Hard to interpret for high-dimensional data |

**Notes:**
$X_t$, $Y_t$ represent time series variables at time $t$; $k$ is the lag; $\mathrm{Cov}(\cdot)$ is covariance; $\mathrm{Var}(\cdot)$ is variance; $\sigma_X$, $\sigma_Y$ are standard deviations of $X$ and $Y$; $\phi_{kk}$ is the $k$th lag partial autocorrelation coefficient.
$p(x)$ and $p(y)$ are marginal probabilities; $p(x,y)$ is the joint probability distribution.

(CSO), and genetic algorithm (GA)) with ANN, ANFIS, and SVM for monthly GWL predictions. Principal component analysis (PCA) reduced time series data, and the Taguchi model optimized parameters. ANFIS-GOA achieved the best accuracy, while SVM was less effective. Kayhomayoon et al. (2022) applied ANFIS with meta-heuristic algorithms (e.g., Genetic Algorithm (GA) and Ant Colony Optimization (ACO)) to predict GWLs in the Urmia aquifer. Using a Sugeno-type function with Gaussian membership functions, ANFIS-ACOR outperformed the base model and other hybrids. Roy et al. (2023) developed a Bayesian model averaging (BMA)-based ensemble model for GWL forecasting in Bangladesh, combining seven ML models including ANFIS, RF, GPR, Bi-LSTM, and SVR. Past GWL values were used as inputs, with Minimum Redundancy Maximum Relevance (MRMR) applied for lag selection. Lag-1 was most predictive. MARS and RF handled feature selection internally. The ensemble consistently outperformed standalone models across all wells and forecast horizons.

Refer to supplementary information for a detailed description of ANFIS.

#### 3.4.2. Results

Based on the reviewed papers, we found that

1. Gaussian Membership Functions (MF) were the most frequently employed, followed by Trapezoidal MF. Membership functions are a fundamental component of fuzzy logic systems, and they define how each input variable's value is associated with different fuzzy sets. These functions are essential to ANFIS's fuzzy inference procedure. In a study by Talpur et al. (2017), the influence of four common membership function shapes on the effectiveness of ANFIS in tackling diverse classification tasks was investigated. Their findings indicated that the Gaussian membership function, due to its superior accuracy and lower computational demands, emerged as the most promising choice. It is worth noting that several reviewed papers did not specify the MF utilized.
2. Hyperparameter tuning was performed primarily using a trial-and-error approach. However, in select cases, algorithms such as ACOR, GOA, CSO, WA, GA, KA, and PSO were employed, producing favorable results. Section 5.1, provides more information about these algorithms.
3. Partial Autocorrelation and Autocorrelation analyses were the prevalent methods for selecting optimal features for model inputs.
4. Both standalone ANFIS models and hybrid ANFIS models consistently outperformed ANN models. This superior performance can likely be attributed to ANFIS models integrating both neural networks and fuzzy logic, making them more adept at handling non-stationary time series data.

### 3.5. SVM/SVR (Standalone and hybrid) models

#### 3.5.1. Bibliographic review

Behzad et al. (2010) compared SVM and ANN for simulating and forecasting GWLs in the Towaco aquifer, Morris County, N.J., across multiple timeframes (daily to bimonthly). The study employed the radial basis function kernel for SVM and concluded that SVM outperformed ANN in both training and testing phases, showcasing its reliability for GWL prediction. In a similar study, Yoon et al. (2011) evaluated the performance of SVM and ANN models in predicting GWL fluctuations in a coastal aquifer in Korea. The study employed the cross-correlation analysis to identify the most influential features. The SVM model was trained using the Sequential Minimal Optimization (SMO) algorithm, and parameters were fine-tuned through trial and error. Results showed that the SVM model outperformed the ANN model in forecasting GWL fluctuations. Suryanarayana et al. (2014) introduced an integrated wavelet-SVR (WA-SVR) model with an RBF kernel for predicting monthly GWL fluctuations in Visakhapatnam, India. Optimized through trial and error, the WA-SVR outperformed SVR, ANN, and ARIMA models. Zhou et al. (2017) combined discrete wavelet transform (DWT) with SVM (WSVM) for GWL forecasting in the Huai River Basin, China. Using PSO-based hyperparameter tuning and lag optimization via Partial Autocorrelation Function, WSVM achieved superior performance over standard ANN, SVM, and Wavelet Preprocessed ANN (WANN) models. In the same year, Ebrahimi and Rajaee (2017) evaluated the impact of wavelet analysis on SVR, MLR, and ANN models for one-month-ahead GWL predictions in the Qom Plain, Iran. Using auto-correlation analysis for feature selection, wavelet-enhanced models (WNN, WLR, WSVR) outperformed standalone counterparts. The RBF kernel was used for SVR with trial-and-error optimization. Nie et al. (2017) applied SVM and RBF-ANN models to forecast monthly GWL in Jilin province, China, concluding that SVM effectively predicts GWL while analyzing uncertainties through confidence intervals. Huang et al. (2017) employed chaos theory for variable selection, bypassing linear correlation analyses. They developed SVM and BP-ANN models for predicting GWLs in China's Three Gorges Reservoir Area. The chaotic PSO-SVM model, optimized with RBF kernel and PSO, achieved superior accuracy over its linear counterpart and chaotic BP-ANN models. Mukherjee and Ramachandran (2018) explored the relationship between terrestrial water changes from GRACE data ($\Delta$TWS) and GWLs in India using SVR, ANN, and linear regression models. SVR outperformed the others, highlighting $\Delta$TWS as a valuable input for modeling irregular GWL time series. Hussein et al. (2020) compared SVR with MLR, MLP, RF, and XGB for GWL prediction, with XGB also used for feature selection. Feature engineering was performed using the Gaussian Mixture Model on GRACE data. SVR consistently outperformed all other models. Yadav et al. (2020) integrated MI theory, SSA, and GA with SVM and ANN models to analyze climatic and non-climatic impacts on GWL fluctuations in India. Using the Kernel Basis Function, hybrid models (SSA-MI-GA-ANN and SSA-MI-GA-SVM) outperformed standalone models after hyperparameter fine-tuning. Yin et al. (2021) developed a Bayesian ensemble modeling approach using SVM, ANN, and Response Surface Regression to predict groundwater storage fluctuations. The SVM model, employing a Gaussian kernel, demonstrated low uncertainty and strong regional-scale predictions. Yu et al. (2021) combined Grey Relational Analysis (GRA) and Factor Analysis (FA) with SVM to predict GWL in Minqin County, China. The GRA-FA-SVM, utilizing an RBF kernel, outperformed standalone SVM, BPNN, and RBFNN models. Liu et al. (2021) compared standalone SVM and SVM with data assimilation (SVM-DA) for predicting short- to medium-term GWL changes (1–3 months) in the northeastern United States. Using correlation analysis for feature selection and the RBF kernel with fine-tuned hyperparameters, SVM-DA achieved higher accuracy than standalone SVM. Dehghani et al. (2022) evaluated hybrid models (BWO–SVR, WSVR, and AIG–SVR) for predicting GWL changes under the RCP8.5 scenario in Iran's Khorramabad plain. Parameter tuning employed black widow and rifile algorithms, with WSVR consistently outperforming other models. Kajewska-Szkudlarek et al. (2022) compared SVR and MLP for monthly GWL prediction in northern Poland, using the Hellwig method for predictor selection. The RBF kernel in SVR slightly outperformed MLP. Sarkar et al. (2024) evaluated five nonlinear ML models: Polynomial Regression, Random Forest, XGBoost, KNN, and SVM-RBF for GWL prediction at a well in IIT Roorkee, India. Cross-validation was used for hyperparameter tuning. XGBoost achieved the highest accuracy and was best suited for capturing GWL changes linked to gravity variations. Niu et al. (2023) used a SVM model to predict GWL in the North China Plain, focusing on regions with varying human activity. Cross-correlation analysis and SHapley Additive exPlanations (SHAP) analysis were used to select input variables (e.g., precipitation, temperature, population, GDP) and determine time lags. Bayesian optimization tuned SVM parameters. The model had a commendable overall performance. Kayhomayoon et al. (2023) evaluated GWL prediction in Iran's Dehgolan aquifer using MODFLOW and machine learning models—SVR, least-square SVR (LSSVR), and a hybrid SVR–WOA. Inputs included past GWL and monthly recharge/withdrawal. LSSVR outperformed other ML models, and the combined LSSVR-MODFLOW model gave the best results. Wu et al. (2023) compared SVM, Long-Short Term Memory (LSTM), MLP, and Gated Recurrent Units (GRU) models for GWL prediction in China's Hebei Plain using data from six monitoring stations (2018–2020). RBF kernel function selected for SVM. GRU showed the highest accuracy, especially for fluctuating or increasing trends. SVM had the weakest performance.

Refer to supplementary information for a detailed description of SVM/SVR

#### 3.5.2. Results

Based on the reviewed papers, we found that

1. The RBF that is the Radial Basis Function kernel was the most frequently employed, followed by the Polynomial and Linear kernels. The RBF kernel formulated by Broomhead and Lowe (1988) in 1988 is a mathematical function commonly used in machine learning, particularly in SVM and other kernel-based algorithms. It is a type of kernel function that helps transform data into a higher-dimensional space, making it easier to classify or separate non-linear data. it is important to note that there is no one-size-fits-all kernel function. The choice of kernel relies on the particular problem, dataset, and its inherent characteristics. In some cases, other kernels like linear, polynomial, or sigmoid kernels may perform better. Table 4 provides a comparison between these kernels.
2. Hyper-parameters were predominantly fine-tuned through a trial-and-error approach. However, in specific instances, alternative methods such as the Creative Rifle and Black Widow (Dehghani et al., 2022), and the Taguchi Model (Seifi et al., 2020) were utilized to optimize the tuning process.
3. In addition to correlation analysis, Principal Component Analysis (PCA), Chaos theory, Grey Relational Analysis (GRA), Factor Analysis (FA), and the Hellwig method were utilized to facilitate optimal feature selection. Table 5 provides a general overview as well as some advantages and disadvantages.
4. SVMs, when tuned to optimal hyperparameters, consistently outperformed ANN and ANFIS models.

### 3.6. LSTM (Standalone and hybrid) models

#### 3.6.1. Bibliographic review

Zhang et al. (2018) assessed the LSTM model's performance in five sub-areas of the Hetao Irrigation District in arid northwestern China, comparing it with a traditional FFNN and a double-layered LSTM model. Monthly data on evaporation, temperature, precipitation, water diversion, and time were used as inputs. Key hyperparameters were fine-tuned through trial and error. The results indicated that the LSTM model outperformed both the FFNN and the double-layered LSTM model. A few years later, Solgi et al. (2021) employed the LSTM-NN model to predict GWLs using historical GWL data as the only input. The study compared its performance with a basic neural network (NN) for predicting short- and long-term GWLs in the Edwards aquifer, Texas. The LSTM-NN model was trained with the Adam optimizer, and input variable selection was refined through trial and error. Results consistently demonstrated that the LSTM-NN outperformed the basic NN across all evaluation scenarios. In that same year, Haq et al. (2021) applied LSTM networks for real-time tracking and prediction of Terrestrial Water Storage Change (TWSC) and Groundwater Storage Change (GWSC) across five Saudi Arabian basins using GRACE datasets from 2003 to 2025. Correlation analysis was employed to evaluate the influence of input variables. The LSTM model, trained with the ADAM optimizer, outperformed the autoregression model in accuracy and computational efficiency. Wu et al. (2021a) proposed the WT-multivariate LSTM (WT-MLSTM) method for simulating and predicting GWLs, tested in the Liangshui River Basin, China, and the Cibola National Wildlife Refuge, USA. The model, trained using the Adam optimizer, demonstrated superior prediction accuracy compared to standard LSTM, MLSTM, and WT-LSTM models. Ao et al. (2021) compared the LSTM model, the kernel-based nonlinear extension of the Arps decline model (KNEA), and the GRU model for estimating GWL in the Hetao Irrigation District, China. Hyperparameters were optimized using grid search, and the study concluded that the LSTM model outperformed the other methods.

In some of the more recent studies, Vu et al. (2023) utilized the Bidirectional LSTM (BiLSTM) model to predict GWL dynamics in a Normandy karst massif in eastern France. Feature relevance was determined through correlation analysis, and the ADAM optimizer was used for training. The study concluded that the BiLSTM model outperformed the standard LSTM model. Manna and Anitha (2023) developed the Double-Edge Bi-Directed Long Short-Term Memory (DEBi-LSTM) model, a deep ensemble learning approach for simulating and forecasting groundwater levels in India. The model was enhanced using the Randomized Low-Ranked Approximation (RLRA) algorithm, with feature selection guided by the Variance Inflation Factor (VIF) and Multi-Collinearity Test. The study concluded that the DEBi-LSTM model outperformed existing models, including LSTM, bagging ensemble, and general ensemble models. Foroumandi et al. (2023) employed ConvLSTM, FFNN, and RF models to downscale monthly GRACE-derived Terrestrial Water Storage Anomaly (TWSA) to a 10 km resolution over Iran using remote sensing images. The Growing Neural Gas (GNG)

**Table 4**
Comparison of the SVM kernels (Rahmadani and Lee, 2020; Scholkopf et al., 1997; Panja et al., 2019; Patle and Chouhan, 2013).

| Kernel | Description | Formula | Advantages | Disadvantages |
|---|---|---|---|---|
| RBF (Gaussian) | The RBF kernel maps input vectors into an infinite-dimensional space and is effective for capturing complex non-linear relationships. | $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ | - Handles non-linear relationships well<br>- Effective in high-dimensional spaces<br>- Few parameters to tune | - Requires careful tuning of $\gamma$<br>- May overfit with high $\gamma$<br>- Computationally demanding for large datasets |
| Polynomial kernel | Captures interactions of features by computing polynomial combinations of inputs. Useful when prior knowledge suggests polynomial relationships. | $K(x, x') = (a \cdot x \cdot x' + b)^d$ | - Models non-linear patterns<br>- Adjustable via degree $d$<br>- Performs well when data is polynomially separable | - Computationally expensive for large $d$<br>- Prone to overfitting<br>- Requires parameter tuning ($a$, $b$, $d$) |
| Sigmoid kernel | Inspired by neural networks, this kernel simulates the behavior of an activation function. | $K(x, x') = \tanh(a \cdot x \cdot x' + b)$ | - Mimics neural network behavior<br>- Useful for binary classification<br>- Captures certain non-linearities | - Sensitive to parameter choices<br>- Less robust than RBF<br>- Not always positive semi-definite |
| Linear kernel | Assumes linear separability. Best suited for high-dimensional data with a clear linear margin. | $K(x, x') = x \cdot x'$ | - Fast and simple<br>- Fewer hyperparameters<br>- Works well with linearly separable data<br>- Less prone to overfitting | - Ineffective for non-linear problems<br>- Limited flexibility<br>- May underfit complex patterns |

**Notes:** $x$, $x'$ are input feature vectors; $x \cdot x'$ denotes their dot product; $\|x - x'\|^2$ is the squared Euclidean distance.
$\gamma$ is a kernel parameter controlling the spread in the RBF kernel; $a$ is a scale factor; $b$ is a bias term; $d$ is the degree of the polynomial.
tanh is the hyperbolic tangent function used in the sigmoid kernel. Proper tuning of these parameters is essential for optimal model performance.

**Table 5**
Comparison of feature selection methods (King and Jackson, 1999; Tang et al., 2014; Omiotek et al., 2019; Huang et al., 2017).

| Method | Description | Advantages | Disadvantages |
|---|---|---|---|
| PCA | PCA is a dimensionality reduction technique that transforms features into a set of linearly uncorrelated components, ordered by variance. | - Reduces dimensionality<br>- Highlights key features<br>- Simplifies data visualization | - May lose interpretability<br>- Assumes linear relationships<br>- Not ideal for non-linear data |
| Chaos theory | Chaos theory examines complex systems that appear disordered but are deterministic. In ML, it identifies chaotic patterns in data. | - Identifies hidden patterns<br>- Useful for complex systems<br>- Captures dynamic behaviors | - Requires complex calculations<br>- Interpretation can be difficult<br>- Sensitive to initial conditions |
| GRA | Evaluates the relationship between multiple criteria using grey relational grades to rank and select important features. | - Handles uncertainty and complex multi-criteria problems well<br>- Requires fewer data points compared to some methods<br>- Provides a clear ranking of features based on relational grades | - Can be complex to implement and interpret<br>- Sensitive to the selection of reference series<br>- May require normalization of data for accurate analysis |
| FA | A statistical method that identifies underlying factors that explain the relationships between variables, reducing dimensionality. | - Reduces dimensionality by identifying latent factors<br>- Helps in understanding the underlying structure of data<br>- Simplifies the feature space by focusing on significant factors | - Assumes linear relationships among variables<br>- Can be sensitive to outliers and data noise<br>- Interpretation of factors can be subjective and complex |
| Hellwig method | The Hellwig method evaluates the importance of features by considering their correlation with the target variable and redundancy among themselves. | - Considers both relevance and redundancy<br>- Facilitates selection of non-redundant features<br>- Provides a clear ranking of features | - May require large datasets<br>- Can be computationally intensive<br>- Relies on accurate correlation measures |

algorithm clustered TWSA data to identify similar pixels for model inputs and outputs. ConvLSTM utilized the ReLU activation function, with layer optimization performed through trial-and-error. The results indicated that ConvLSTM outperformed RF and FFNN, effectively downscaling GRACE data and producing groundwater storage maps for Iran. Ehteram et al. (2023) developed a hybrid SATCN-LSTM model integrating self-attention and temporal convolution with LSTM to improve GWL prediction accuracy. The model was tested using meteorological inputs and outperformed baseline SATCN, TCN, and standalone LSTM models. It achieved the lowest MAE (0.06) and RMSE (0.08), addressing limitations of vanishing gradients and redundant inputs in sequence forecasting. Heudorfer et al. (2023) used global LSTM models to predict GWL from 108 wells in Germany. Dynamic inputs included precipitation and temperature, while static features (e.g., land cover, hydrogeology) were selected via correlation and spatial relevance analysis. Feature importance was assessed using permutation feature importance (PFI). Models with static features performed no better than random ones in-sample. Out-of-sample, the dynamic-only model gave the best results, highlighting the stronger role of climatic inputs. Jing et al. (2023) applied six models (RF, XGBoost (XGB), GBR, LightGBM, Vanilla-LSTM, EnDe-LSTM) to simulate GWL in China's North China Plain. Feature engineering created GWSAFE and Human_activity from GRACE and GLDAS datasets. Feature importance was assessed using GINI and Permutation Feature Importance (PFI). Human_activity emerged as the dominant predictor. Deep learning models, especially EnDe-LSTM, outperformed tree-based models in capturing groundwater variation across aquifer types. Zheng et al. (2024) developed a VMD-iTransformer model to predict GWL in China's Kubuqi Desert using data from nine monitoring stations. Variational Mode Decomposition (VMD) was used as a preprocessing step to decompose non-stationary time series into intrinsic components. The VMD-iTransformer model outperformed both classic Transformer and LSTM models. Elzain et al. (2024) used CBR, XGB, LGBM and LSTM, GRU, Transformer to forecast water table rise (WTR) in Oman. SHAP was used for feature selection. Data preprocessing involved lag creation, differencing, and seasonal aggregation. Stacked models outperformed individual ones, with GRU leading in one-week ahead forecasts. Sun et al. (2023) assessed GWL prediction in Beijing's Yongding River fan using Long Short-Term Memory (LSTM) models and a physics-based (PB) model (PGMS). Input variables included water supply, precipitation, and runoff, with lag times identified via correlation analysis. Improved LSTM models consistently outperformed the PB model in short-term forecasts. The PB model remained better for long-term trends.

Refer to supplementary information for a detailed description of LSTM.

#### 3.6.2. Results

1. The Adam optimizer, introduced by Kingma and Ba (2014), was predominantly utilized for training the LSTM model. Renowned for its efficiency and adaptability, Adam combines the strengths of RMSprop and Momentum optimizers, resulting in faster and more reliable convergence compared to traditional algorithms. Its ability to dynamically adjust learning rates for each parameter during training proves particularly advantageous for complex models like LSTMs, where varying learning rates enhance convergence. The widespread adoption of Adam is evidenced by its over 100,000 citations within eight years of publication. In comparative studies, LSTM consistently outperformed ANN models, likely due to its robust architecture, which effectively mitigates issues like local minima that often impede convergence during training.
2. The ReLU activation function was commonly utilized as the activation function for the LSTM model. The ReLU activation function was created by Nair and Hinton (2010) and it is well known for its ability to handle the gradient vanishing problem. Using the ReLU function as the activation in a neural network, as opposed to the sigmoid function, results in partial derivatives of the loss function having values of either 0 or 1. This property effectively mitigates the issue of gradient vanishing, making ReLU an effective choice for preventing gradient-related problems.
3. Hyperparameters were mostly adjusted through trial and error. However, in some cases, like in the study by Ao et al. (2021), the grid search method was used. The grid search approach is a systematic method used in machine learning to find the optimal hyperparameters for a model. it involves defining a grid of possible values for each hyperparameter and then evaluating the model's performance for every possible combination of these values.
4. The selection of optimal input variables primarily relied on correlation analyses. However, in some select cases, methods such as Cross-wavelet analysis and the Granger causality (Kim et al., 2023), Multi Collinearity Test, and Variance Inflation Factor (VIF) (Manna and Anitha, 2023) were employed to improve model performance through enhanced feature selection. Also, it is worth mentioning that one of the primary advantages of the LSTM model compared to other AI models like FFNN is that, while important features of the input dataset typically need to be identified through mathematical measures such as Cross-correlation (CC) or MI, the LSTM model can automatically achieve this through its hidden layers (Sharghi et al., 2022). Table 6 provides an overview of these methods as well as some strengths and limitations.
5. In terms of performance, LSTM consistently outperformed ANN models, possibly due to its inherent capability to overcome the limitations associated with local minima, which can hinder convergence in the training process.

### 3.7. RF (Standalone and hybrid) models

#### 3.7.1. Bibliographic review

In the scope of GWL simulation and modeling with RF, Lendzioch et al. (2021) assessed the RF model's ability to predict peat bog GWL and soil moisture in the Rokytka Peat Bog using ultrahigh-resolution UAV maps. Predictor selection, hyperparameter tuning, and performance evaluation were conducted using a leave-location-out (LLO) spatial cross-validation strategy combined with forward feature selection (FFS) to mitigate overfitting and enhance predictions for untested locations. The study concluded that the RF model demonstrated strong predictive performance. Mosavi et al. (2021) utilized ensemble models, including GamBoost, AdaBoost, and Bagged CART, to simulate and predict GWLs in the Dezekord-Kamfiruz watershed, Iran. Variable selection was optimized using Recursive Feature Elimination (RFE) and multicollinearity assessment via the Variance Inflation Factor (VIF). The study concluded that Bagging methods, particularly Random Forest and Bagged CART, outperformed Boosting models like AdaBoost and GamBoost, with Random Forest achieving the highest performance.

Zhou et al. (2022) explored the use of GRACE satellite data, GLEAM, and GLDAS datasets, combined with meteorological variables, to predict GWL using RF, SVR, and ELM models. For the RF model, hyperparameters such as Ntree (100–500) and Mtry (1–3) were fine-tuned. The study concluded that RF demonstrated the best predictive performance, followed by SVR and ELM. Liu et al. (2022) developed multiple models, including RF, SVM, GRNN, DT, CNN, LSTM, and GRU, to simulate GWLs in the lower Tarim River basin. The SHAP method was employed to evaluate the impact of covariates on model performance. Results indicated that the RF model consistently outperformed the other models. Pham et al. (2022) tested the validity of the RF model in simulating and forecasting GWL fluctuations in two wells in northwest Bangladesh. The study compared the RF model with six other models: Random Tree (RT), Decision Stump, M5P, SVM, Locally Weighted Linear Regression (LWLR), and Reduced Error Pruning Tree (REP Tree). They concluded that the Bagging RF and Bagging RT models outperformed the others. Rafik et al. (2023) evaluated Random Forest (RF), SVM, and k-Nearest Neighbors (k-NN) for GWL prediction in the Saïs basin, Morocco. Input data included precipitation, soil moisture, runoff, and evapotranspiration from ERA5-Land, along with NDVI and land surface temperature (LST). A correlation matrix was used for feature selection. Among the models, RF showed the most satisfactory performance, making it the preferred choice for regional GWL prediction. Zowam and Milewski (2024) applied RF and SVR for statewide GWL anomaly prediction in Arizona, integrating geostatistical interpolation (EBK) to estimate GWL and using it as a predictor in the RF model. Final features were selected using permutation feature importance after training. The integrated RF+EBK model showed high accuracy in unconsolidated aquifers, emphasizing the role of geology and data quality. May-Lagunes et al. (2023) assessed several ML and deep learning models including ARIMA, XGBoost, RF, LSTM, and TFT for GWL prediction in California's Sacramento River Basin. Inputs included historical GWLs, well characteristics (e.g., depth, usage, location), SWE, ET, and seasonal features from Fourier decomposition. Feature selection involved experimenting with lagged variables and measuring importance in XGBoost, where well-specific features ranked highest. XGBoost outperformed all models, especially for 3-month forecasts, highlighting the value of well-level information over climate predictors. Hikouei et al. (2023) applied MLR, RF, and XGBoost to predict GWL in Indonesia's Mawas peat dome. Inputs included elevation, precipitation, ET, and distance from canals. Tree-based models provided feature importance, with elevation ranked highest. XGBoost outperformed RF and MLR, showing superior accuracy and lower residuals. Yi et al. (2024) predicted GWL near South Korea's Baekje weir using five ML models: RF, ANN, SVR, GB, and XGBoost. Feature selection was done via permutation-based importance. XGBoost consistently outperformed all others across evaluation metrics. Gupta et al. (2024) applied three Bagging-based models, Random Forest (RF), Bagging-REPTree, and Bagging-DSTree, to forecast groundwater levels in Punjab, India, using long-term GWL data from 14 wells. Feature importance was derived using the RF model. RF

**Table 6**
Analytical methods for feature selection in GWL using LSTM (Thompson et al., 2017; Maraun and Kurths, 2004; Akinwande et al., 2015; Torrence and Compo, 1998; Attanasio et al., 2013; Stokes and Purdon, 2017).

| Method | Description | Advantages | Disadvantages |
|---|---|---|---|
| Cross-wavelet analysis | Analyses the relationship between two time series in the time–frequency domain, providing insights into their co-movement at different frequencies and times. | Provides detailed time–frequency information<br>Useful for analyzing non-stationary time series<br>Can identify phase relationships | Computationally intensive<br>Requires careful interpretation<br>Sensitivity to noise and boundary effects |
| Granger causality | Determines whether one time series can predict another, indicating a directional influence between the variables. | Helps identify potential causal relationships | Assumes linear relationships<br>May not capture complex dynamics<br>Sensitive to the choice of lag length |
| Multi Collinearity Test | Assesses the degree of correlation among independent variables in a regression model, indicating potential multicollinearity issues. | Identifies problematic correlations<br>Helps in improving model stability<br>Simple and easy to apply | Does not quantify the impact on the model<br>May miss subtle collinearity issues<br>Assumes linear relationships |
| Variance Inflation Factor (VIF) | Quantifies the degree of multicollinearity by measuring how much the variance of an estimated regression coefficient increases if the predictors are correlated. | Provides a clear numerical measure of multicollinearity<br>Easy to interpret and use<br>Helps in model refinement | Only applicable to linear models<br>Does not indicate causality<br>May not detect non-linear relationships |

consistently outperformed the other models in both pre- and post-monsoon seasons across multiple evaluation metrics. Chi et al. (2024) used Tree Ensemble models (DT, RF, XGBoost) with rolling means and lagged features to predict GWL. Inputs included groundwater depth, temperature, rainfall, and drainage volumes. SHAP analysis highlighted lagged GWL and moving averages as most important.

Refer to supplementary information for a detailed description of RF.

#### 3.7.2. Results

1. In the extensive body of research we have examined, it is clear that both standalone RF models and their hybrid counterparts consistently outperform traditional models like ANN and SVM in various fields. RF's strength lies primarily in its exceptional predictive accuracy, which makes it a valuable tool across many applications.
2. In conjunction with correlation analysis, the SHAP (SHapley Additive exPlanations), the Recursive Feature Elimination (RFE) (Mosavi et al., 2021), and forward feature selection (FFS) (Lendzioch et al., 2021) methods were occasionally employed to select optimal input variables for the models. Table 7 provides an overview of these methods as well as some strengths and limitations.
3. A noteworthy observation is that many papers did not explicitly specify the hyperparameter values for 'ntree' and 'mtry'. The optimal values for 'ntree' and 'mtry' in RF models are recognized to be problem-dependent (Martínez-Muñoz and Suárez, 2010). While specific values may vary based on the nature of the problem, it is commonly acknowledged that RF often performs reasonably well with default hyperparameter values. For instance, the default value for 'ntree' is typically set at 500 trees, and 'mtry' is defaulted to one-third of the total number of variables (Mutanga et al., 2012). This suggests that researchers may rely on these default values without explicit specification. However, the significance of choosing appropriate hyperparameter values, particularly in the context of groundwater recharge studies, should not be overlooked.

## 4. Trends and gaps in feature selection for GWL modeling

Across the reviewed studies, a diverse array of feature selection techniques have been employed, ranging from classical statistical tools to modern hybrid and model-driven approaches. Broadly, these methods fall into three categories: filter-based, wrapper, and embedded approaches. Filter methods evaluate features based on statistical metrics independent of the model, wrapper methods iteratively assess feature subsets using the learning algorithm itself, and embedded methods perform selection as part of the model training process. Each comes with trade-offs in interpretability, computational demand, and robustness. A close inspection reveals that while performance-centric motivations drive most feature selection choices, methodological inconsistencies and reporting gaps persist, limiting comparability and reproducibility across studies.

### 4.1. Filter-based approaches

Filter methods are independent of the machine learning model and select features based on general characteristics of the data, such as statistical relevance or correlation. These techniques are computationally efficient and easy to implement. However, they may overlook feature interactions and dependencies that are only revealed during model training.

Filter techniques were extensively applied across early ANN, ANFIS, and SVM studies. ACF, PACF, cross-correlation, and Mutual Information (MI) were commonly used for lag selection. For instance, Nourani et al. (2015) and Bahaa et al. (2015) used MI to identify relevant time lags and improve model accuracy by capturing nonlinear dependencies. Yu et al. (2021) implemented Gray Relational Analysis (GRA) and Factor Analysis (FA) in the SVM frameworks, observing better precision than conventional methods. The Hellwig method was adopted in SVR (Kajewska-Szkudlarek et al., 2022) to assess feature weights using a synthetic capacity index. LSTM-based studies often employed correlation matrices to select relevant features (Sun et al., 2023), while PCA was used in hybrid ANN-ANFIS frameworks (Seifi et al., 2020) to reduce dimensionality. VIF and multicollinearity tests were also used (Manna and Anitha, 2023). Additional studies supporting filter-based selection include Wei et al. (2023a), who applied wavelet decomposition and phase space reconstruction (PSR) to preprocess inputs before feeding them into ANN models. Lyapunov exponent analysis supported the relevance of PSR in chaotic systems. In Ghafoor et al. (2024), ANN models using interpolated and upsampled GWL time series outperformed ARIMA, reinforcing ANN's capability to learn from preprocessed, irregular data. Correlation and standard preprocessing approaches (e.g., normalization, cubic spline interpolation) were also applied in studies by Fahim et al. (2024), Seidu et al. (2023), Feng et al. (2024). Zheng et al. (2024) used Variational Mode Decomposition (VMD) as a preprocessing tool to decompose non-stationary time series before feeding data into the iTransformer model.

### 4.2. Wrapper approaches

Wrapper methods evaluate subsets of features by training and testing a specific model. These approaches often yield better performance as they consider interaction effects and model dynamics. However, they

**Table 7**
Feature selection methods applied with RF models in GWL prediction studies (Strobl et al., 2007; Chen and Jeong, 2007; Kuhn et al., 2013b; Lundberg and Lee, 2017; Jović et al., 2015; Kamalov et al., 2024; Takefuji, 2025).

| Method | Description | Advantages | Disadvantages |
|---|---|---|---|
| SHAP (SHapley Additive exPlanations) | Game-theory-based approach that assigns each feature an importance value by quantifying its marginal contribution to predictions. | Provides consistent, interpretable attributions; captures nonlinear interactions. | Computationally expensive for large datasets; results can be sensitive to background data selection. |
| Recursive Feature Elimination (RFE) | Iteratively fits the model, removes the least important feature(s), and repeats until the optimal subset is identified. | Efficiently identifies strong predictors; helps reduce dimensionality. | Greedy elimination may discard interacting features too early; requires repeated model training. |
| Forward Feature Selection (FFS) | Begins with no features, adds predictors sequentially based on performance improvement until no significant gain is observed. | Computationally less intensive than exhaustive search; yields parsimonious models; avoids inclusion of irrelevant predictors. | Greedy nature may miss optimal subsets; struggles with correlated predictors; can overfit without proper validation. |
| RF Inherent Feature Importance | RF naturally provides importance measures based on mean decrease in impurity (MDI) or mean decrease in accuracy (MDA). | Embedded in the model; efficient and scalable; offers direct ranking of features. | Importance may be biased toward variables with more categories or higher variance; does not always reflect causal relevance. |

are computationally expensive and prone to overfitting, especially on small datasets.

Wrappers were increasingly adopted in more recent studies. Forward Feature Selection (FFS) was used by Lendzioch et al. (2021) to select UAV-based predictors for RF models under a leave-location-out strategy, showing strong performance in untested locations. Recursive Feature Elimination (RFE) was used in ensemble studies (Mosavi et al., 2021), leading to the selection of non-redundant inputs and improved accuracy for RF and Bagged CART. The Taguchi method was used to select the optimal combinations of characteristics and parameters for ANFIS in Seifi et al. (2020), producing the best performing hybrid (ANFIS-GOA). Metaheuristic wrappers such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) were used to tune the inputs and parameters of SVM and ANFIS (Kayhomayoon et al., 2022; Zhou et al., 2017), improving the generalization of the model. Akbari Majd et al. (2024) applied multiple preprocessing stages involving decomposition and metaheuristic optimization (GA, PSO, ACO) in ANN frameworks, with wrapper-based improvements delivering up to 76% accuracy gain. In Seifi et al. (2024), Boruta was used as a wrapper for feature importance before applying decomposition and GRU-RVM sequence models. Moradi et al. (2023) evaluated different sets of characteristics using internal validation performance, choosing only GWL for the final training after evaluating the correlation patterns.

### 4.3. Embedded approaches

Embedded methods integrate feature selection directly into the model training process. These include regularization techniques and importance measures derived from model parameters. Embedded methods are more efficient than wrappers and often yield robust and interpretable outcomes.

In studies based on RF and XGBoost, the importance of permutation characteristics and GINI was frequently applied after training to rank predictors (Yi et al., 2024; Jing et al., 2023). These techniques revealed the dominant role of well-level inputs and GRACE-derived groundwater storage indicators. SHAP values were used in the RF, SVM and LSTM studies (Elzain et al., 2024; Niu et al., 2023), providing interpretable information on variable contributions. MARS and RF were used as embedded selectors in ensemble pipelines (Roy et al., 2023), while permutation importance was used to refine features in ConvLSTM and GRU models trained on remote sensing inputs (Foroumandi et al., 2023). GNG clustering, while not a direct selection method, helped identify spatially coherent input structures prior to model training (Heudorfer et al., 2023). Chi et al. (2024) used SHAP to analyze lagged inputs in tree ensemble models, confirming the dominance of moving averages and historical GWLs. Abdi et al. (2024) used CNN and DNN models in which the relevance of input characteristics was indirectly derived from the learned filters. Their preprocessing relied on model-based interpolation techniques (e.g., M5P) to reconstruct missing inputs. Feng et al. (2024) further evaluated model-specific importance rankings using Pearson and Spearman correlations to interpret dominant drivers. Jing et al. (2023) used GINI and Permutation Feature Importance (PFI) to assess the relevance of engineered features like GWSAFE and Human_activity, with deep learning models (especially EnDe-LSTM) outperforming tree-based alternatives.

Despite methodological innovations, a recurring limitation across all studies is the lack of standardization in reporting. Feature selection procedures were often vaguely described or entirely omitted, complicating reproducibility and interpretation. Only a handful of works explicitly distinguished between filter, wrapper, and embedded selection techniques. To date, only one study by Saroughi et al. (2024) has come close to systematically evaluating the impact of input processing strategies on groundwater level modeling. Their work tested 126 data preprocessing methods across multiple AI models (SVR, ANN, LSTM, and POA-ANN), focusing on how statistical, wavelet, and decomposition-based transformations influenced predictive accuracy. However, even this comprehensive effort emphasized data pre-processing rather than a formal comparison of feature selection methods.

This signals a methodological gap where future research could systematically evaluate and benchmark feature selection techniques, especially under varying hydrogeological and data availability scenarios, to establish standardized best practices for ML-based GWL forecasting. Table 9 provides a summary of all the feature selection approaches reviewed.

## 5. General overview and discussion

This section will highlight key findings from the analysis of the 223 reviewed articles. These findings encompass various aspects, including optimization algorithms, the treatment of time steps, selection of input variables, validation metrics, etc.

### 5.1. Optimization algorithms

In recent decades, several bio-inspired optimization techniques have been created. According to Tang and Wu (2009), these algorithms can be broadly categorized into three primary types: swarm intelligence, bacterial foraging algorithms, and evolutionary algorithms. Swarm intelligence algorithms were the most popular optimization techniques used in groundwater studies, according to the articles reviewed. Table 10 provides some advantages and disadvantages of the optimization algorithms mentioned above. Refer to the Supplementary Information for a detailed description of the workflow of all the optimization algorithms discussed.

**Table 8**
Summary of the reviewed paper used in the bibliographic sections.

| Paper Ref | AI used | Prediction horizon | Input variables | Time steps | Data division (%) | Length of total data |
|---|---|---|---|---|---|---|
| Dash et al. (2010) | ANN | Short-term (1-week ahead) | GWL,R | Weekly | 70/30 | 1993–2002 (2340 samples) |
| Mohanty et al. (2010) | FFNN | Short-term (1,2,3,4-weeks ahead) | R, E, RS, water level in the drain, pumping rate, GWL | Weekly | 70/30 | Feb 2004–June 2007 (174 samples) |
| Chen et al. (2010) | RBFN | Short-term (1-month ahead) | GWL | Monthly | 79/21 | 1997–2003 (63 samples) |
| Chen et al. (2011) | BPN | Short-term (1-month ahead) | GWL | Monthly | 79/21 | Jul 1998–Nov 2004 (76 samples) |
| Trichakis et al. (2011) | ANN | Short-term (1-day ahead) | GWL, wet days, R, pumping rate | Daily | 80/20 | - (7109 samples) |
| Rakhshandehroo et al. (2012) | FFNN, RBNN, ELNN, GRNN | Short-term (1-month ahead) | GWL, Temp, Runoff, R | Monthly | 77/23 | 1993 to 2003 (4524 samples) |
| Adamowski and Chan (2011) | ANN, ARIMA | Short-term (1-month ahead) | GWL, Temp, R | Monthly | 80/20 | Nov 2002–Oct 2009 (84 samples) |
| Nourani et al. (2011) | ANN, Geostatistics | Short-term (1-month ahead) | GWL, Temp, R, discharge, lake level | Monthly | 80/20 | Apr 1994–Mar 2006 (144 samples) |
| Taormina et al. (2012) | FFNN, ARX | Short-term (1-h ahead) | GWL, Evapotranspiration, R | Hourly | 83/17 | Oct 2006–June 2008 (23 850 samples) |
| Sahoo and Jha (2013) | ANN, MLR | Short-term (1-month ahead) | GWL, Evapotranspiration, R, Temp, river stage, SDV | Monthly | 73/27 | 1999–2004 (72 samples) |
| Nourani et al. (2015) | FFNN, ARIMAX | Short-term (1-month ahead) | GWL, Runoff, R | Monthly | 72/28 | 1988–2012 (298 samples) |
| Guzman et al. (2015) | ANN, SVR | Short-term (1-day ahead) | GWL, R | Daily | 55/45 | Jun 1984–Sep 1994 (2679 samples) |
| Bahaa et al. (2015) | ANN, MLR, ENN | Short-term(1-day,week,month ahead) | Tailing Recharge, R, Air temp | Daily | 90/10 | May 2009–Oct 2011 (900 samples) |
| Chang et al. (2015) | ANN | Short-term (10-days ahead) | GWL, Temp, R | Daily | 56/44 | Jul 2009–Dec 2012 (653 samples) |
| Gholami et al. (2015) | ANN | Monthly (dendrochronology) | Tree-ring diameter, R | Monthly | 70/30 | 1970–2013 (44 samples) |
| Sun et al. (2016) | FFNN | Short-term (1,3,7 days ahead) | GWL, R | Daily | 50/50 | Jan 2012–Dec 2013 (731 samples) |
| Choubin and Malekian (2017) | ANN, ARIMA | Short-term (1 months ahead) | R, SF, Temp, Evaporation, GWL | Monthly | 83/17 | 1993–2010 (216 samples) |
| Wen et al. (2017) | ANN | Short-term (1,2,3 months ahead) | R, Temp, Evaporation, GWL, R | Monthly | 74/26 | Jun 2003–Dec 2010. (91 samples) |
| Sahoo et al. (2017) | ANN, MLR, MNLR | Seasonal-annual (Ĩ year ahead) | R, Temp, ENSO, NAO, SF, ID | Monthly | 70/30 | 1980–2012 (148 368 samples) |
| Ebrahimi and Rajaee (2017) | ANN | Short-term (1-month ahead) | R, Temp, GWL | Monthly | 70/30 | Apr 2002–Mar 2013 (132 samples) |
| Guzman et al. (2017) | NARX | Short-term (3-months ahead) | R, GWL | Daily | 70/30 | 1987–1994 (2922 samples) |
| Wunsch et al. (2018) | NARX | short-mid-term(1 week-6 months ahead) | R, Temp | Daily | 90/10 | 1948–2015 (13 676 samples) |
| Jeong and Park (2019) | NARX, LSTM, GRU | Short-term (1-day ahead, continuous) | R, Temp, Humidity, GWL, CSH, AAP | Daily | 50/50 | 2005–2014 (7284 samples) |
| Lee et al. (2019) | FFNN | Short-term (1-h ahead) | SWL, GA, GHPU | Hourly | 60/40 | Feb 2016– Apr 2017 (8712 samples) |
| Kouziokas et al. (2018) | FFNN | Short-term (1-day ahead) | Humidity, Temp, R, GWL | Daily | 70/30 | Jan 2014–Dec 2014 (365 samples) |
| Ghose et al. (2018) | RNN | Short-mid-term (monthly steps) | Humidity, Temp, R, Runoff, Evapotranspiration | Monthly | 70/30 | 1988–2007 (120 samples) |

(*continued on next page*)

### 5.2. Validation metrics

Machine learning models are prone to limitations that can hinder their predictive performance, one of the most notable being overfitting. Overfitting occurs when a model learns patterns and noise specific to the training dataset, resulting in high training accuracy but poor performance on unseen test data. Detecting overfitting therefore requires evaluating model performance on both training and test datasets, and comparing the results. The role of validation metrics is to quantify prediction accuracy and reliability, enabling meaningful comparison of results across datasets and studies.

**Table 8** (*continued*).

| | | | | | | |
|---|---|---|---|---|---|---|
| Van Thieu et al. (2023) | MLP | Short-term (1-month ahead) | R, Temp, Tidal height, GWL | Monthly | – | 1989–2012 (552 samples) |
| Aderemi et al. (2023) | NARX | Short-term (No. specified) | R, Temp | Monthly | 70/30 | 1996–2021 (6240 samples) |
| Panahi et al. (2023) | RBFNN, ANFIS, SVM, MLP | Long-term (Years ahead) | R, Max Temp, Min Temp, Evaporation | Daily | – | 2022–2064 (10 950 samples) |
| Jalalkamali et al. (2011) | ANFIS, ANN | Short-term (1-month ahead) | R, Temp, GWL | Monthly | 80/20 | 1988–2009 (264 samples) |
| Shirmohammadi et al. (2013) | ANFIS, ARIMA, ARMA, ARX, ARMAX, SARIMA | Short-term (1,2,3-months ahead) | Discharge, R, Evaporation | Monthly | | 1992–2007 (180 samples) |
| Emamgholizadeh et al. (2014) | ANFIS, ANN | Long-term (2-Years ahead) | Recharge, Pumping Rates, IRF | Monthly | 80/20 | 2002–2011 (108 samples) |
| Maiti and Tiwari (2014) | ANFIS, ANN, BNN | Short-term (1-month ahead) | Temp, R | Monthly | 75/25 | Sep 1972–Oct 2001 (350 samples) |
| Mirzavand et al. (2015) | ANFIS, SVR | Long-term (No. Specified) | Streamflow, Evaporation, SD, AD, R | Monthly | 70/30 | Jan 1990–Jan 2010 (240 samples) |
| Gong et al. (2016) | ANFIS, ANN, SVM | Short-term (1,2,3-months ahead) | GWL, Temp, Lake levels, AD, R | Monthly | 80/20 | 1998–2009 (144 samples) |
| Seifi et al. (2020) | ANFIS, ANN | Short-term (1-month ahead) | GWL | Monthly | 80/20 | Jan 2000–Sep 2012 (140 samples) |
| Kayhomayoon et al. (2022) | ANFIS | Short-term (1-month ahead) | GWL, R, Temp, GW | Monthly | 70/30 | 2001–2017 (204 samples) |
| Behzad et al. (2010) | SVM, ANN | Daily, Weekly, Monthly | Pump Rates, R, Temp, GWL | Daily, Weekly, Biweekly, Monthly, Bimonthly | – | 2002–2002 (63,93,109,116,122 samples) |
| Yoon et al. (2011) | SVM, ANN | Short-term (multiple-hours ahead) | Tide level, R, GWL | Hourly | – | Jun 2004–Nov 2006 (3213 samples) |
| Suryanarayana et al. (2014) | SVR, ANN, ARIMA | Short-term (1-month ahead) | GW Depth, Max Temp, Mean Temp, R | Monthly | – | May 2001–Feb 2012 (130 samples) |
| Yoon et al. (2016) | SVR, ANN | Long-term (1-day ahead) | GWL, R | Daily | 70/30 | 2003–2008 (10 229 samples) |
| Zhou et al. (2017) | SVM, ANN | Short-term (1-month ahead) | GWL, R | Monthly | 75/25 | Jan 1974–Dec 2010 (444 samples) |
| Nie et al. (2017) | SVM, RBF-ANN, MLR | Short-term (1-month ahead) | R, Evaporation, Temp | Monthly | 83/17 | Jan 2003 to Dec 2014 (144 samples) |
| Huang et al. (2017) | SVM, BPNN | Short-term (1-day,week,month ahead) | GWL | daily, weekly, monthly | 50/50 | 2006–2014 (171, 90, 24 samples) |
| Mukherjee and Ramachandran (2018) | SVM, ANN, LRM | Short-term (No. Specified) | ▵TWS, R, Temp, Humidity | monthly | 80/20 | Jan 2005–Dec 2013 (35, 67 samples) |
| Hosseini et al. (2016) | ANN | Short-term (1-month ahead) | R, Average Discharge, Temp, Evaporation | Monthly | – | Oct 2000 8̃0/20 Sep 2009 (108 samples) |
| Yadav et al. (2020) | SVM, ANN | Short-term (1,2-months ahead) | GWL, R, Temp, Population, GR, SOI, NOI, NINO3 | Monthly | 70/30 | 2010–2017 (96 samples) |
| Liu et al. (2022) | RF, SVM, GRNN, DT, CNN, LSTM, GRU | Long-term (Month-years-ahead) | Temperature, Humidity | Monthly | 75/25 | Jan 2000–Dec 2020 (1̃8 648 samples) |
| Guzman et al. (2019) | SVR, NARX | Short-term (days-ahead) | GWL R, Evapotranspiration | Daily | 70/30 | Jun 1985–Sep 1994 (2592 samples) |
| Chen et al. (2020) | MLP, RBF, MODFLOW, SVM | Long-term (1 month-ahead) | SF, GWL | Daily | 80/20 | Jan 1986–Dec 2010 (300 samples) |

(*continued on next page*)

In groundwater level (GWL) modeling, a wide variety of evaluation metrics have been employed. Fig. 7 shows the frequency of performance metrics reported in the reviewed literature. Among these, RMSE is the most frequently adopted, appearing in nearly 160 instances, followed by $R^2$, correlation coefficient (Corr), and NSE. Metrics such as MAE, MSE, and MASE are also widely used, reflecting a strong preference for error-based evaluations. In contrast, less common metrics like ME, RE, and NRMSE appear less frequently but still contribute to the assessment landscape. The “Others” category includes less standardized or study-specific measures.

**Table 8** (*continued*).

| | | | | | | |
|---|---|---|---|---|---|---|
| Kajewska-Szkudlarek et al. (2022) | SVR, MLP | Long-term (Monthly) | R, Temp, GWL | Monthly | 75/25 | 1975–2014 (414, 411, 476 samples) |
| Zhang et al. (2018) | LSTM, FFNN | Long-term (1-month ahead) | R, Temp, Evaporation, Water Division, Time | Monthly | 85/15 | 2000–2013 (168 samples) |
| Solgi et al. (2021) | LSTM-NN | Short&Long-term (days,months-ahead) | GWL | Monthly | 80/20 | Nov 1932–Jul 2020 (31 239, 1043 samples) |
| Haq et al. (2021) | LSTM-NN, ARIMA | Long-term (65 months ahead) | GRACE | Monthly | 85/15 | Jan 2003–Jun 2020 (185 samples) |
| Wu et al. (2021a) | LSTM | Short&Long-term (months-ahead) | River Stage, GWL | Monthly | 70/30 | Nov 2003–Dec 2019 (194, 10 176, samples) |
| Ao et al. (2021) | LSTM, KNEA, GRU | Long-term (1-month ahead) | Temp, GWL, R, GSR, Irrigation Quantity | Monthly | 70/30 | 1990–2015 (312 samples) |
| Gaffoor et al. (2022) | LSTM, GBDT | Long-term (1-month ahead) | Temp, GWL Anomaly, R, runoff, Evapotranspiration | Monthly | 70/30, 80/20, 65/36 | 2009–2019 (42, 88 samples) |
| Kim et al. (2023) | LSTM | Short-term (2-weeks head) | Sun Hours, GWL Anomaly, R, runoff, Min Temp, Mean Wind Speed, Evaporation, R | Daily | 90/10 | 2012–2021 (3653 samples) |
| Vu et al. (2023) | BiLSTM | Long-term (7,4,30-days ahead) | River Level, River Flux, Sea Level, R, Temp, Seinen River Data | Daily, Hourly | 65/35 | 1971–2021 (5479 samples) |
| Manna and Anitha (2023) | DEBi-LSTM | Long-term (Multiple days ahead) | GW Recharge, Natural GW Discharge, R | Daily | 60/40 | 2000 to 2021 (12 000 samples) |
| Patra et al. (2023) | LSTM | Short-term (Daily) | GWL | Daily | – | 2013 83/17 Jun 2021 (139 653 samples) |
| Foroumandi et al. (2023) | ConvLSTM, FFNN, RF | – | LST, NDVI, R, DEM, ET, SM, SWE | Monthly | 70/30 | Apr 2002–Dec 2016 (297 samples) |
| Lendzioch et al. (2021) | RF | – | GWL, NDVI, SM | Monthly | – | Aug 2018–Nov 2019 (630 samples) |
| Gonzalez and Arsanjani (2021) | RF, ANN, SVM | Long-term (years-ahead) | Max Temp, Min Temp, R, Mean Temperature, Topography | Monthly | 80/20 | 1990–2018 (over 10 000 samples) |
| Zhou et al. (2022) | RF, SVR, ELM | Short-term (1,2,3-months ahead) | R, Temp, Actual Evapotranspiration, GLEAM, GLDAS | Monthly | 80/20 | Apr 2002–Jun 2017(183 samples) |
| Kalu et al. (2022) | DBN | Short-term (5-months ahead) | GWL, ENSO, NAO, AMO, IOD, PDO | Monthly | 70/30 | Nov 2007–Oct 2012 (1620 samples) |
| Hussein et al. (2020) | MLR, MLP, RF, XGB | Short-term (1-month ahead) | GRACE | Monthly | 90/10 | Mar 2002–May 2019 (161 samples) |
| Liu et al. (2021) | SVM | Short-term (3-months ahead) | GRACE, Temp, R, Solar Radiation, IST | Monthly | 50/50 | Jan 2007–Dec 2016 (5520 samples) |
| Yu et al. (2021) | SVM, BPNN, RBFNN | Short-term (Months ahead) | Meteorological, Socio-economic factors | Monthly | 86/14 | 2004–2010 (168 samples) |
| Derbela and Nouiri (2020) | ANN | Medium-term (months-ahead) | R, GWL, Evapotranspiration | Monthly | 80/20 | 2000–2018 (1944 samples) |
| Chenjia et al. (2024) | LSTM, GRU, MLP, 1DCNN, TR | Short-term (12,24,36-days ahead) | GWL, Extraction, ET | Daily | 70/30 | 2017–2022 (2190 samples) |
| Akbari Majd et al. (2024) | ANN | Short-term (Months ahead) | Climate Variables | Monthly | 70/30 | 2001–2019 (7068 samples) |
| Mohapatra et al. (2021) | DNN, ANFIS, SVM | Long-term (seasons ahead) | GWL, R, Temp, | Monthly | 70/30 | 1996–2016 (835 044 samples) |

(*continued on next page*)

RMSE and $R^2$ are especially prominent choices because they provide complementary insights: RMSE quantifies the average magnitude of prediction error in the same units as the observed data, while $R^2$ measures the proportion of variance in the observations explained by the model. The choice of metric should be informed by the study's objectives, the characteristics of the dataset, and the model's intended application, ensuring that the evaluation process captures both accuracy and explanatory power.

### 5.3. *Covariates used*

ML models offer a unique advantage by learning patterns in data, whether simple or complex, to predict specific outcomes. In the context of this systematic review, we examine the use of different types of data for groundwater modeling, including climate variables, hydrogeological parameters, and aquifer characteristics, albeit the latter being

**Table 8** (*continued*).

| | | | | | | |
|---|---|---|---|---|---|---|
| Müller et al. (2021) | LSTM, MLP, RNN, CNN | Short-term (1-day ahead) | Streamflow, Temp, R | Daily | 67/33 | 2010–2018 (2191 samples) |
| Sharghi et al. (2022) | FFNN, ANFIS, LSTM, GMDH | Long-term (1,2,3 - months ahead) | GWL, R, Runoff | Monthly | 70/30 | 1989–2018 (17 640 samples) |
| Yin et al. (2021) | ANN, SVM, RSR, BMA | Short-term (1-month ahead) | ASR, GP, SWD | Monthly | 80/20 | Oct 1973–Sept 2015 (2016 samples) |
| Maheswaran and Khosa (2013) | ANN, LR, DAR | Long-term (multiple months ahead) | GWL | Monthly | 90/10 | 1975–2002 (648 samples) |
| Dehghani et al. (2022) | SVR | Long-term (Months & Years ahead) | R, Temp, GWL, Water Withdrawal | Monthly | 70/30 | 2000–2020 and 2021–2040 (960 samples) |
| Pham et al. (2022) | RT, Decision Stump, M5P, SVM, LWLR, REP, Tree | Long-term (months ahead) | GWL, Mean Temp, R, Humidity | Monthly | 80/20 | Jan 1981–Dec 2017 (888 samples) |
| Collados-Lara et al. (2023) | NARX, NAR, ELNN | Short-term (1–6 months ahead) | R, Min Temp, Max Temp | Monthly | 70/30 | Jan 2000–Jan 2020 (12 852 samples) |
| Chang et al. (2016) | NARX | Short-term (1-month ahead) | GWL, SF, R | Monthly | 80/20 | 2000–2013 (168 samples) |
| Banadkooki et al. (2020) | RBFNN, MLP | Short-term (months-ahead) | R, Temp | Monthly | 70/30 | 2000–2012 (156 samples) |
| Guo et al. (2021) | ConvLSTM | Short-term (5-days intervals) | GWL, R, GSD, HPD | Monthly | 70/30 | Jan 2012–Dec 2012 (3456 samples) |
| Rafik et al. (2023) | RF, SVM | Short-term (Months-ahead) | TWS, R, ET, Q, SM, NDVI, and LST | Monthly | – | Apr 2002–2022 (264 samples) |
| Elmotawakkil et al. (2024) | GBR, SVR, RF, and DT | Short-term (Months-ahead) | ET, R, NDVI, and LST | Daily | 70/30 | Feb 2000–Feb 2023 (42 000 samples) |
| Bonkoungou et al. (2024) | NeuralProphet, LSTM, XGBoost | Short-term (multiple days ahead) | GWLs, R, and EVI | Daily | 85/15 | 2010–2021 (26 298 samples) |
| Singh et al. (2024) | AutoML, RF, Boosting EL, BDT, GAM, GRNN, LR, ANN, SVR, RBNN, KR, LSTM | Long-term (quarterly seasons) | GWLs, EVAP, Temp, R, RH, Soil type | Monthly | 70/30 | 1997–2018 (8310 samples) |
| Fahim et al. (2024) | MLR, Tree models, SVM, GPR, and ANN | Spatial prediction | GLDAS GWS data, ED, Temp, R, PD, ID, GRACE | Monthly | 80/20 | 2003–2019 (183 samples) |
| Osman et al. (2024) | ANN, SVR, XGBoost, and LSTM | Short-term (1-day ahead) | R, Temp, EVAP, GWLs | Daily | 70/30 | Jan 2030–Dec 2039 (1390 samples) |
| Wei et al. (2023a) | ANN | Short-term (m × 10-days ahead) | GWLs | Daily | 80/20 | 1991–2015 (576 samples) |
| Seifi et al. (2024) | RVM, ANN, MLP | Short-term (1-month ahead) | R, Average Temp, RH, IRF | Daily | 80/20 | 1995–2015 (252 samples) |
| Ali et al. (2024) | TFT, LSTM | Short-term (7,30,60 days ahead) | R, GWL | Daily | 70/30 | 2001–2023 (8400 samples) |
| LaBianca et al. (2024) | CatBoost, GBDT, PB | Spatial prediction | DEM, LandUse, TS, GWLs | Monthly | 67/33 | 2006–2022 (280 samples) |
| Moradi et al. (2023) | ANN, ELM, ORELM, and GMDH | Short-term (1-month ahead) | GWL, R, Temp, and EVAP | Monthly | 80/20 | 2010–2016 (198 samples) |
| Heudorfer et al. (2023) | LSTM | Short-term (weeks ahead) | R, Temp, RH, Tsin | Weekly | 50/50 | Jan 2000–Dec 2015 (205 343 samples) |
| Roy et al. (2023) | ANFIS, Bagged RF, Boosted RF, GPR, LSTM MARS, SVR | Short-term (1–3 weeks ahead) | GWLs | Weekly | 70/30 | Feb 1984–Sep 2018 (7228 samples) |

(*continued on next page*)

non-mandatory compared to conventional models. The choice of these covariates is critical, significantly affecting model performance.

Groundwater levels are closely tied to climate conditions, and the ever-growing concerns of climate change, marked by events like droughts, floods, and shifting precipitation patterns, emphasize the urgency of groundwater management. Fig. 8 shows that historical GWL data is the most employed covariate for ML models predicting groundwater levels.

**Table 8** (*continued*).

| | | | | | | |
|---|---|---|---|---|---|---|
| Chi et al. (2024) | WT-Model | Short-term (7-days ahead) | Temp and R | Daily | 80/20 | Jan 2009–Jun 2020 (5223, 609 samples) |
| Abdi et al. (2024) | CNN, DNN | Short-term (1-month ahead) | GWS | Monthly | 70/30 | 2018–2021 (1632 samples) |
| Zowam and Milewski (2024) | SVR, RF | Long-term (1-month ahead) | R, SM, ET, LST, VI, CN, SHC, and GSA | Monthly | 60/40 and 85/15 | Jan 2010–Dec 2019 (7080 samples) |
| Zhu et al. (2024) | GBDT | Short-mid-term (days,month-ahead) | GWL, R, Q, CWD | Monthly | 70/30 | Jan 1997 to Dec 2011 (163 155 samples) |
| Feng et al. (2024) | CNN, RNN, SVM, DT, RF, and GAN | – | GRE, R, and RFR | Monthly | 70/30 | 2018–2022 (2136 samples) |
| May-Lagunes et al. (2023) | XGBoost, RF, TFT, GDR, LR | Short-term (3-months ahead) | SWE, ET, WSE | Monthly | – | 2010–2020 (3960 samples) |
| Jing et al. (2023) | RF, XGBoost, GBR, LightGBM, LSTM | – | R, ET, RD, GWL | Daily | – | 2003–2014 (1584 samples) |
| Zheng et al. (2024) | iTransformer, LSTM | Short-term (4-months ahead) | Agricultural irrigation, climatic component, basin factor, and human factor | Hourly | – | Jul 2021–Sep 2022 (24 678 samples) |
| Hikouei et al. (2023) | RF, XGBoost, MLR | – | GWL ET ET, R, Z, DFC | Monthly | 80/20 | 2010–2012 (11 641 samples) |
| Seidu et al. (2023) | ANN, BPNN, RBFNN, GMDH, GRNN | – | R, EVAP, Temp | Monthly | 50/50, 60/40, 70/30, 80/20, 90/10 | 2013–2018 (676 samples) |
| Yi et al. (2024) | RF, ANN, SVR, GB, XGBoost | Short-term (1-day ahead) | DWL, R | Daily | 80/20 | 2011–2021 (17 350 samples) |
| Chidepudi et al. (2023) | LSTM, BiLSTM, GRU | Short-term (Daily) | TEMP, R | Daily | 80/20 | 1970–2020 (600) |
| Sriram et al. (2023) | RF, ML, MD, KNN | – | – | – | – | - (4044 samples) |
| Sun et al. (2023) | LSTM, PB | Short-term (3-months ahead) | R, WS, GWL, Q | Daily | 80/20 | Jan 2018–Sep 30 (68 958 samples) |
| Elzain et al. (2024) | CBR, XGB, LGBM, LSTM, GRU | Short-term (1,2,3-weeks ahead) | WTR | Daily | 90/10 | Dec 2017–Jan 2019 (19 465 samples) |
| Chen et al. (2023) | CBR, XGB, LGBM, LSTM, GRU | Short-term (1–10 days ahead) | GWL, R | Daily | 70/30 | 2002–2021 (7305 samples) |
| Sarkar et al. (2024) | KNN, SVM-RBF, PR, XGB, RF | – | GRACE | Daily | 83/13 | Jan 2022–Oct 2022 (177 samples) |
| Ehteram et al. (2023) | SATCN, LSTM | Short-term (1-month ahead) | Temp, R, Elevation, WS, RH, GWL | Monthly | 75/25 | 1995–2010 (192 samples) |
| Niu et al. (2023) | SVM | Short-term (1-month ahead) | R, wind speed WS, Temp, POP,GDP, EIA | Monthly | 60/40 | Jan 1991–Dec 2019 (240 samples) |
| Kayhomayoon et al. (2023) | SVR, Least-Square SVR (LSSVR), MODFLOW | Long-term (Years ahead) | GWL, Recharge, Withdrawal | Monthly | 75/25 | Oct 2010–Sep 2013 (150 samples) |
| Bai and Tahmasebi (2023) | GWN, GWN-adaptive, LSTM, GRU | Long-term (Multiple weeks ahead) | Temp, R, VP, SR, day | Weekly (GWL), Daily (Climate) | 60/40 | 2010–2020 (23 452 samples) |
| Nan et al. (2023) | Attention-GRU, LSTM, RNN, CNN | Short-mid-term (6-months ahead) | GWL, R, Temp, ELEV, SM | Monthly | 75/25 | 2017–2022 (24 120 samples) |
| Li et al. (2023) | MLR, MARS, ANN, RFR, GBR | – | ELEV, soil type, Climate Data, NDVI | Seasonal | 84/16 | Oct 2005–Sep 2007 (75 samples) |
| Nand et al. (2024) | MLP-GA, MODFLOW | Long-term (Annual/Seasons ahead) | ETC, R, GR, GD | Annual | 67/33 | 2000–2015 (880 samples) |

(*continued on next page*)

**Table 8** (*continued*).

| | | | | | | |
|---|---|---|---|---|---|---|
| Bhadani et al. (2024) | F-IWO-GWL, RF, BoostEL, BDT, LSTM, SVR, GAM, ANN, RBNN, GRNN, hybrids | Short-mid-term (Annual/Seasons ahead) | R, Temp, ET, RH, GWL | Daily | 70/30 | 1997–2018 (8310 samples) |
| Sun et al. (2024) | Transformer, MLP, LSTM, CNN | Short-term (10, 20 days ahead) | R, Flow, GWL | Daily | 60/40 | 2000–2019 (725 samples) |
| Zhang et al. (2023) | TCN, LSTM | Short-term (1,3,7,15-days ahead) | Tidal-level, R, GWL | Hourly | 80/20 | Oct 2011–Oct 2012 (28 836 samples) |
| Xie and Zhang (2024) | STA-GRU, LSTM, GRU, CNN+GRU, CNN+BiLSTM+ Attention | Short-term (1-month ahead) | Climate, hydrogeological, Topographic | Monthly | 70/30 | 2017–2117 (7272 samples) |
| Zhou et al. (2024) | LSTM | Short-term (1-month ahead) | GWL | Monthly | 81/19 | 2018–2022 (20 340 samples) |
| Wei et al. (2023b) | M5, RF, GRBFN, MLP, Ensemble methods | Spatial (No time-series) | Remote sensing Data | – | 80/20 | 2018 (436 samples) |
| Wang et al. (2024) | LSTM, RR-LSTM, Geo-RR-LSTM, ARIMA, ARIMAX, MLR | Long-term (1-year ahead) | R, geographical features | Monthly | 78/22 | 2009–2013 (3816 samples) |
| Su et al. (2024) | MLR, SVR, LSTM | Long-term (20-months ahead) | GWL, Extraction | Daily | 89/11 | 2007–2021 (964 260 samples) |
| Wu et al. (2023) | SVM, LSTM, MLP, GRU | Short-term (1-h ahead) | GWL | Hourly | 70/30 | 2018–2020 (32 880 samples) |
| Eghrari et al. (2023) | LSTM, GRU | Long-term (Monthly) | Climate Data | Monthly | 80/20 | Nov 2002–Jun 2022 (1652 samples) |
| Fronzi et al. (2024) | Prophet, ARIMA, MARS, ETS | Short-term (2-weeks ahead) | GWL, Hydrological, Exogenous, Atmospheric variables | Daily | 87/13 | Mar 2023–Apr 2023 (485 samples) |
| Gupta et al. (2024) | RF, Bagging-REPTree, Bagging-DSTree | Seasonal | GWL | Seasonal | 70/30 | 1997–2018 (616 samples) |
| Ghazi et al. (2021) | ANN, LSSVM, NARX | Long-term (Years ahead) | R, Temp, Time Delay | Monthly | 70/30 | 1966–2019 (6840 samples) |
| Idrizovic et al. (2020) | ANN | Long-term (Years ahead) | R, Temp, PET, GWL | Daily and monthly | 70/30 | 1988–2016 (342 samples) |
| Karthikeyan et al. (2013) | FFNN, RNN | Short-term (1-week ahead) | R, Temp, EVAPO, GWL | weekly | 70/30 | May 2004–May 2006 (109 samples) |
| Yan and Ma (2016) | ARIMA, RBFN | Short-term (1-month ahead) | GWL | monthly | 83/17 | Jan 1998–Dec 2010 (144 samples) |
| Secci et al. (2023) | NARX, LSTM, CNN | Long-term (years ahead) | R, Temp | monthly | 90/10 | Mar 2005–Dec 2020 (190 samples) |
| Fallah-Mehdipour et al. (2013) | ANFIS, GP | Medium-term (months) | EVAPO, R, GWL | monthly | 86/14 | 7-year (84-month) (252 samples) |

**Abbreviations:** R, Precipitation; SM, Soil moisture; E, efficiency coefficient; LWLR; locally weighted linear regression, REP Tree; reduce error pruning tree; TWI, Topographic wetness index; TPI, Topographic position index; TRI, Topographic roughness index, Dd; Drainage density; Dff, Distance from fault; NDVI, Normalized Difference Vegetation Index; DEM: Digital Elevation Map; GSR; Global Solar Radiation, IST; Infrared Surface Temperature, Streamflow; SF, Seasonal Dummy Variables, SDV; Streamflow Discharge, SD; Aquifer Discharge, AD; Linear Regression, LR; Dynamic Auto-Regressive, DAR; Nash sutcliffe criteria, NSC; Geological Structure Data, GSD; Hydrogeological Parameter Data; HPD, Terrestrial Water Storage; TWS, evapotranspiration; ET, Soil Moisture; SM, Gradient Boosting Regression; GBR, Enhanced Vegetation Index; EVI, Runoff; Q, Fuzzy Inference Systems; FIS, Teaching–Learning Based Optimization; TLBO, Ant Colony Optimization; ACO, Harmony Search; HS, Evaporation; EVAP, Relative Humidity; RH, Self-Normalizing Neural Network; SNN, Standardized Bathymetry Data; SBA, Population Data; PD, ID; irrigation data, elevation data; ED, phase space reconstruction; PSR, Relevance Vector Machine; RVM, multivariate variational mode decomposition; MVMD, Boruta feature selection algorithm; BFSA; Irrigation Flow; IRF, Temporal Fusion Transformer; TFT, Bayesian optimization; BO, Terrain slope; TS, Group Method of Data Handling; GMDH, Annual sinusoidal curve fitted to temperature; Tsin, drainage_volumes; V, drainage_volumes; V, groundwater surfaces; GWS, Empirical Bayesian kriging; EBK, Land Surface Temperature; LST, Generative adversarial network; GAN, Groundwater extraction rate; GRE, River flow rate; RFR, crop water demands; CWD, Temporal Fusion Transformer; TFT, Gradient Descent Regressor; GDR, Snow water equivalent; SWE, Water Surface Elevation; WSE, River Discharge; RD, Elevation; Z, and Distance from canal; DFC, Daily weir level; DWL, Transformer; TR,LightGBM; LGBM,CatBoost Regressor; CBR, Water Table Rise; WTR, Polynomial Regression; PR, population; POP, Gross domestic product; GDP, and Effective irrigated area; EIA, One-Dimensional Convolutional Neural Network; 1DCNN, Vapor Pressure; VP, Solar radiation; SR, Crop evapotranspiration; ETc, deep percolation; GR, applied irrigation water; GD, Rainfall intensity; RI, pore water pressure; PWP, permeability coefficients PCs, lateral flow coefficient; LFC, initial water level height; IWLH, self-attention; SA, convolutional network; SATCN.

Out of 223 papers, 187 incorporate groundwater level data as an input, with 77 using it as the sole input without additional factors. Precipitation data is also frequently used, appearing in 161 instances. Other hydrological data, such as temperature, river discharge, evapotranspiration, and surface water levels, have been employed as inputs. Some papers have explored additional variables like irrigation patterns,

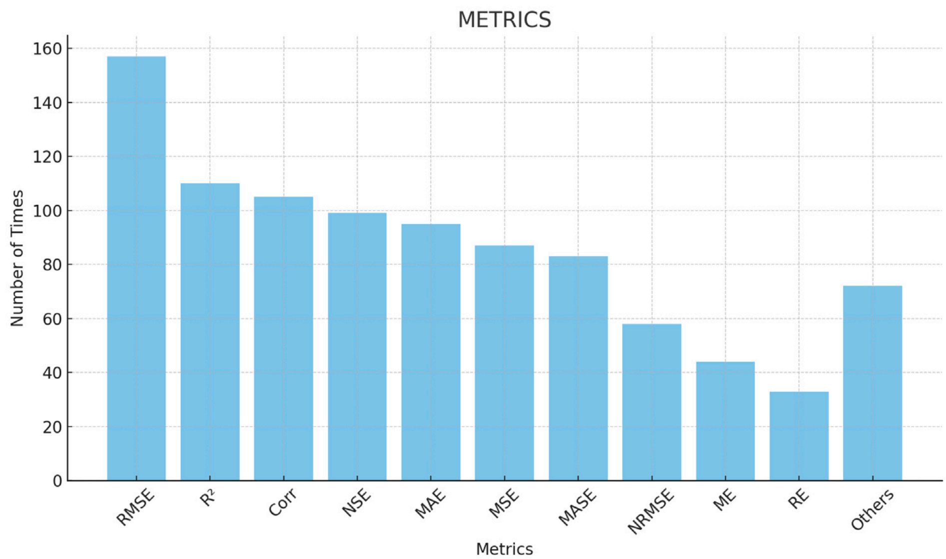


**Fig. 7.** Various metrics used in validation reviewed ML models' performance.

**Table 9**
Feature selection methods identified in groundwater level (GWL) modeling studies, organized by general category. This taxonomy reflects common practice but is not absolute, as some methods overlap across categories.

| Feature selection method | Type |
|---|---|
| Autocorrelation Function (ACF)/Partial ACF (PACF) | Filter |
| Mutual Information (MI) | Filter |
| Cross-correlation analysis | Filter |
| Correlation matrix/Pearson/Spearman correlations | Filter |
| Principal Component Analysis (PCA) | Filter |
| Grey Relational Analysis (GRA) | Filter |
| Factor Analysis (FA) | Filter |
| Hellwig method | Filter |
| Variance Inflation Factor (VIF) | Filter |
| Multicollinearity test | Filter |
| Lyapunov exponent + Phase Space Reconstruction (PSR) | Filter |
| Wavelet decomposition/Variational Mode Decomposition (VMD) | Filter |
| Recursive Feature Elimination (RFE) | Wrapper |
| Forward Feature Selection (FFS) | Wrapper |
| Taguchi design method | Wrapper |
| Genetic Algorithm (GA) | Wrapper |
| Particle Swarm Optimization (PSO) | Wrapper |
| Ant Colony Optimization (ACO) | Wrapper |
| Boruta algorithm | Wrapper |
| Internal validation performance-based selection | Wrapper |
| SHAP (SHapley Additive exPlanations) | Embedded |
| GINI importance (from decision trees) | Embedded |
| Permutation Feature Importance (PFI) | Embedded |
| Model-derived importance (RF, XGBoost, MARS, etc.) | Embedded |
| Growing Neural Gas (GNG) clustering | Embedded |
| Model-based interpolation (e.g., M5P) | Embedded |

Note: Some methods fall at the boundary of categories. For example, Boruta leverages Random Forest importance measures but is generally treated as a wrapper due to its iterative retraining process. Permutation Feature Importance (PFI) is technically post-hoc but grouped under embedded methods here because it is tied directly to fitted models. Wavelet and variational mode decomposition (VMD) are primarily preprocessing/feature extraction steps but are treated as filter techniques in this context.

population figures, seasonal factors, and more, though to a lesser extent. These variables may present challenges during the input selection process.

### 5.4. Programming languages utilized

In the review, we realized that most of the research papers used MATLAB, PYTHON, and R to build the various ML algorithms used. Researchers predominantly utilizing these programs for developing machine learning algorithms in their studies can be attributed to the fact that these programming languages offer extensive libraries and frameworks specifically tailored for machine learning, simplifying algorithm development and implementation. Python, in particular, boasts a vast machine-learning ecosystem, including popular libraries like Scikit-Learn and TensorFlow, making it highly versatile and ideal for a wide variety of ML tasks.

R is favored for its exceptional statistical capabilities and visualization tools, providing researchers with robust data analysis and model interpretation capabilities alongside machine learning functionalities. These languages are open-source, facilitating collaboration and accessibility for researchers globally, while also significantly reducing research costs. Their active and supportive communities continuously contribute to the development and improvement of machine learning tools and resources, ensuring researchers have access to the latest advancements in the field. Information regarding these software programs is available online, and we do not delve into their specifics here. Nevertheless, it is noteworthy that MATLAB is commonly favored in the development of AI models, although several papers have not explicitly mentioned the software used.

### 5.5. Time steps and forecast horizons

Fig. 9 shows the time steps used in the reviewed GWL modeling studies. Most studies (187 papers) employed monthly data, followed by daily (19) and hourly (8) intervals. Weekly, biweekly, and bimonthly resolutions appeared only rarely. The dominance of monthly data can be partly attributed to its wider availability from monitoring networks and its suitability for capturing the delayed recharge response of precipitation, which must travel through the vadose zone before reaching the water table. In inland aquifers, groundwater levels often vary slowly, making coarse temporal resolution sufficient. In contrast, coastal aquifers (Yoon et al., 2011; Taormina et al., 2012) and areas near large reservoirs (Rajaee et al., 2019) require finer time steps to represent rapid tidal or reservoir-driven fluctuations.

A notable trend shown in Table 8 is the strong emphasis on short-term forecasts (3rd column), typically ranging from days to a few months, in the majority of studies. This focus reflects the strengths of machine learning models, which are effective at capturing recent patterns and correlations directly from data. However, the predictive accuracy of ML models tends to decline as the forecasting lead time increases (Shirmohammadi et al., 2013; Sun et al., 2016; Yu et al., 2018; Wu et al., 2021b; Momeneh and Nourani, 2022; Roy et al., 2023).

**Table 10**
Comparison of optimization algorithms (Gandomi and Alavi, 2012; Mehrabian and Lucas, 2006; Saremi et al., 2017; Mafarja et al., 2019; Wang et al., 2018; Asgari et al., 2016; Chandirasekaran and Jayabarathi, 2019; Puente et al., 2009).

| Algorithm | Type of optimization | Advantages | Disadvantages |
|---|---|---|---|
| GA | Evolutionary | -Good for global optimization<br>-Flexible and adaptable<br>-Handles complex spaces | -Requires careful parameter tuning<br>-Can be computationally intensive<br>-May converge to local optima |
| CSO | Swarm intelligence | -Balances exploration and exploitation<br>-Good for dynamic problems<br>-Simple implementation | -Sensitivity to parameters<br>-Requires parameter tuning<br>-Can be slow in convergence |
| PSO | Swarm intelligence | -Simple implementation<br>-Few parameters to adjust<br>-Efficient global search | -May get stuck in local minima<br>-Sensitivity to initial settings<br>-Can require many iterations |
| GOA | Swarm intelligence | -Balances exploration and exploitation<br>-Effective for continuous problems<br>-Adaptable to various problems | -Computationally expensive<br>-Sensitive to parameter settings<br>-May require many iterations |
| WOA | Bio-inspired | -Good for dynamic environments<br>-Mimics natural adaptability<br>-Simple concept to understand | -Can be computationally demanding<br>-Requires careful tuning<br>-Sensitivity to initial population |
| KHA | Swarm intelligence | Effective in multi-modal problems<br>Adaptive to problem complexity<br>Good convergence properties | Computationally intensive<br>Sensitive to parameters<br>Requires many iterations |
| ACO | Swarm intelligence | -Good for combinatorial problems<br>-Utilizes collective intelligence<br>-Finds good paths through search space | -Can be slow to converge<br>-Parameter sensitive<br>-May require large computational resources |
| BWO | Bio-inspired | -Early convergence due to cannibalism<br>-Maintains diversity<br>-Simple implementation | -Risk of losing good solutions<br>-Sensitive to cannibalism and mutation rates<br>-Requires careful tuning |

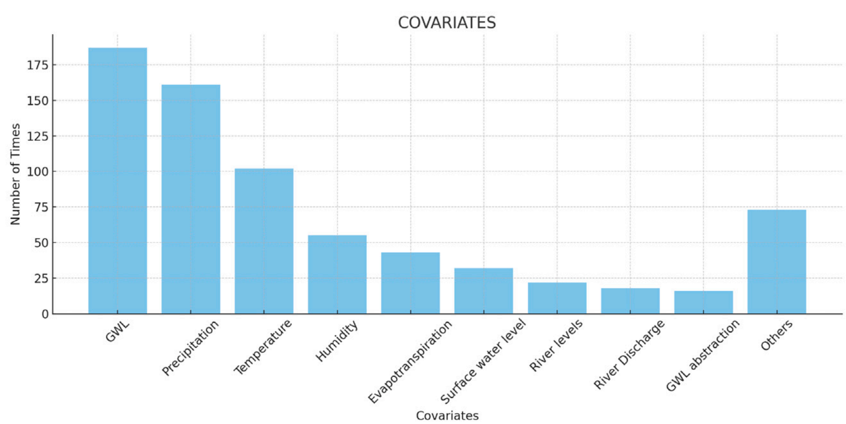


**Fig. 8.** Various covariates employed in reviewed articles.

Over longer horizons, these models may experience error accumulation and reduced generalization, especially when future conditions differ from those present in the training data.

Time step choice is closely linked to the intended forecast horizon. Short-term predictions often employ daily or hourly steps to capture high-frequency variations (Wu et al., 2023; Zhang et al., 2023; Sun et al., 2024), while long-term studies generally adopt monthly intervals to reflect slower groundwater responses (Ghazi et al., 2021; Eghrari et al., 2023; Zowam and Milewski, 2024). Selecting an appropriate combination of time step and forecast horizon should be guided by the dynamics of the aquifer system and the overall purpose of the prediction, balancing the short-term skill of ML methods with the need for robustness in longer-term projections.

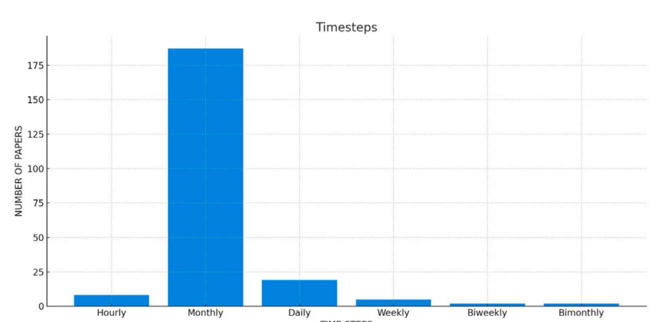


**Fig. 9.** Different time steps used in GWL modeling.

## 6. Implications and future directions

This review identifies several key methodological patterns and research gaps that can inform future GWL modeling efforts and guide the development of more robust, interpretable, and operational models.

First, the lack of comparative studies on feature selection methods represents a major limitation. Although various techniques such as mutual information, cross-correlation, importance of permutation, SHAP, and elimination of recursive features have been used in studies, very few have systematically evaluated their impact within a consistent modeling framework. Given the strong influence of input features on model performance, future research should prioritize controlled comparisons of feature selection strategies, especially in cases involving high-dimensional or multi-source inputs.

Second, the consistent superiority of hybrid models over standalone approaches highlights a promising direction. Many hybrid frameworks, such as LSTM-CNN or RF-GA, combined different model components or algorithms to address specific limitations. For instance, decomposition methods like EMD and VMD (Wu et al., 2021b) helped isolate signal components with clearer temporal patterns, making them easier to model. Optimization algorithms such as GA and PSO were often used to fine-tune hyperparameters, reducing manual trial-and-error. Ensemble techniques also improved prediction by combining outputs from multiple learners (Yin et al., 2021), reducing variance and enhancing robustness. These strategies allowed hybrid and ensemble models to better capture the complex, nonlinear, and multi-scale dynamics of groundwater systems. Future studies should explore their broader integration, particularly in conjunction with hybrid model architectures.

Third, recent developments in deep learning point to new opportunities. Transformer-based architectures, though used in only a few studies, demonstrated strong predictive capabilities and the ability to handle long sequences and complex dependencies. Their limited use in groundwater applications suggests a valuable area for further exploration, especially for multi-step forecasting or spatially distributed predictions.

Fourth, this review finds that remotely sensed data, particularly from GRACE satellites, offer great promise for groundwater modeling in data-scarce regions. GRACE-derived storage changes, while only used in a few studies, enhanced prediction accuracy and captured large-scale dynamics that in-situ measurements could not. There is a strong case for integrating remote sensing into machine learning workflows, especially in regions with sparse monitoring networks.

Additionally, long-term prediction horizons (beyond one year) remain a significant gap, with very few studies attempting extended forecasts due to data limitations, non-stationarity, and issues like model drift. This is particularly relevant for water planning and early warning systems. Addressing this will require robust temporal validation strategies and possibly the integration of seasonal climate forecasts.

Also, despite limited datasets, most ML-based GWL studies achieved strong short-term predictive accuracy. A majority of the reviewed models were developed using relatively small datasets, often fewer than 500 samples, yet still demonstrated high performance, particularly for short-term forecasts. This confirms the flexibility of ML models in data-scarce settings, especially when supported by decomposition, optimization, and appropriate input selection. However, the predictive accuracy of ML models tends to decline as the forecasting lead time increases, as shown by Shirmohammadi et al. (2013), Sun et al. (2016), Yu et al. (2018), Wu et al. (2021b), Momeneh and Nourani (2022) and Roy et al. (2023).

A notable mention in the reviewed literature is the consistent commendation of Long Short-Term Memory (LSTM) networks for their strong performance in long-term groundwater level forecasting (Kajewska-Szkudlarek et al., 2022; Zhang et al., 2018; Solgi et al., 2021; Haq et al., 2021; Wu et al., 2021a; Ao et al., 2021). These studies highlighted LSTM's ability to capture complex temporal dependencies and delayed hydrological responses, even under extended lead times. Ao et al. (2021), in particular, compared LSTM with Kernel-based Nonlinear Extreme Learning Algorithm (KNEA) and Gated Recurrent Unit (GRU) models across multiple districts, and found that LSTM not only achieved a good accuracy but also demonstrated better generalization across spatially distinct sites. This is likely due to LSTM's internal memory and gated structure, which allow it to retain relevant information over long sequences, making it especially suitable for modeling persistent and gradual groundwater dynamics.

While few studies directly compared ML with conceptual models, Chen et al. (2020) found that physically based models like MODFLOW offered better generalization under varying conditions because of the inclusion of physical mechanisms, although not higher short-term accuracy. Based on these findings, future studies could explore combining both approaches. For example, a physically based model could simulate the overall groundwater system, while a machine learning model could improve predictions by learning from the errors or gaps in the physical model. Sun et al. (2023) demonstrated the value of this approach by integrating LSTM with a PB model, achieving prediction accuracy improvements for over 67% of wells when PB model performance was moderate or better, and over 77% when PB performance was poor. Similarly, Kayhomayoon et al. (2023) combined MODFLOW with support vector regression variants, showing that the MODFLOW–LSSVR configuration achieved near-perfect accuracy (NSE = 0.998) in forecasting groundwater levels under climate change scenarios. These examples illustrate how physically consistent PB outputs can provide a robust baseline, while ML components refine predictions by capturing nonlinear and site-specific dynamics. This kind of setup could help improve forecast accuracy across both short- and long-term timeframes.

Finally, the near absence of socio-economic and anthropogenic factors in GWL modeling remains a critical oversight. Land use change, irrigation practices, and population pressures play a significant role in groundwater dynamics, yet are rarely modeled. Future work should consider coupling physical and socio-economic data to capture human–water interactions more effectively.

## 7. Limitations

Although we employed a broad and comprehensive set of keywords to capture a wide range of relevant studies, systematic reviews are inherently challenged by the diversity of terminology used in the literature. For example, some studies may refer to "groundwater level" as "water table" or use "forecasting" instead of "prediction". Similarly, certain papers highlight specific algorithms without explicitly mentioning "machine learning" or "AI". While our expanded keyword strategy mitigated much of this risk, it remains possible that a small number of relevant studies were omitted.

Another limitation is the tendency of the reviewed literature to predominantly report positive results of machine learning applications. Few studies explicitly discussed the failures or shortcomings of the models they employed. This publication bias may have led to an over-representation of successful applications, thereby limiting the scope for a balanced evaluation of machine learning performance.

Finally, our review focused exclusively on peer-reviewed journal articles. This decision was intended to ensure methodological rigor and comparability, but it also excluded potentially valuable insights from grey literature, such as conference papers, theses, or technical reports. While including these sources might have broadened the perspective, it would also have increased heterogeneity and made systematic analysis more challenging.

## 8. Conclusion

ML models have shown significant promise in hydrology and groundwater modeling. This review examines 223 research articles, published from 2010 to 2024, focusing on ML applications in groundwater level (GWL) modeling across diverse geographical settings. These models excel at identifying complex patterns in groundwater datasets, enabling accurate GWL simulations and predictions. Key findings emphasize the critical roles of variable selection, hyperparameter tuning, model architecture, and data preprocessing in achieving optimal model performance. Selecting relevant variables and employing effective preprocessing and optimization techniques can significantly enhance model accuracy and reliability. This review offers practical guidance for researchers adopting ML models for GWL studies, highlighting best practices in variable selection, architecture design, and optimization strategies to improve the accuracy and interpretability of ML-based groundwater predictions

1. During the review, feature selection methods such as correlation analysis and SHapley Additive exPlanations (SHAP) were identified, but no single optimal method emerged. The choice depends on data characteristics, model needs, and research objectives, encouraging researchers to test multiple techniques for relevance. Linear correlation, widely used, assumes proportional relationships between variables. However, groundwater systems involve complex, non-linear interactions like precipitation and aquifer properties, which linear methods often overlook, affecting model performance (Huang et al., 2017). Advanced methods, including mutual information, recursive feature elimination, and model-based approaches, are recommended for identifying features in non-linear contexts.
2. Most studies relied on trial-and-error for hyperparameter tuning, but combining machine learning models with optimization algorithms consistently improved performance. particle swarm optimization (PSO), known for strong global search capabilities, was particularly effective for SVM (Lin et al., 2008; Fei et al., 2009; Huang et al., 2017). Other algorithms, such as Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Grasshopper Optimization Algorithm (GOA), offered robustness and adaptability. However, research on newer techniques like Weed Optimization Algorithm (WOA) and Black Widow Optimization (BWO) remains limited in GW Level modeling studies using ML. Further exploration of these methods is recommended to enhance model performance and optimization.
3. Our review of hybrid models, particularly those integrating wavelet transforms with machine learning, highlights their value for improving groundwater level simulations (e.g., Adamowski and Chan, 2011; Nourani et al., 2015; Ebrahimi and Rajaee, 2017; Wei et al., 2023a; Saroughi et al., 2024). Daubechies wavelets, especially db2 and db4, were widely used. Their popularity may be due to their short support and strong time localization, which help capture sudden fluctuations in groundwater signals. However, because they rely on downsampling, they may distort temporal alignment. To address this, some studies used the à trous wavelet transform, which maintains the original signal length and better preserves structure in seasonal or noisy data. The à trous method has been described as well suited for forecasting applications (Mallat, 1989; Bahaa et al., 2015).
4. It is not possible to recommend a single best ML model for simulating groundwater levels, but the review shows that hybrid models consistently performed better than individual models. This advantage often results from improved handling of signal complexity, more effective parameter tuning, and reduced overfitting. Some studies (Yin et al., 2021) also applied ensemble techniques like Bayesian model averaging, where combining predictions from multiple models produced more stable and accurate results than any standalone model. By testing and combining different models, it is possible to achieve optimal performance in groundwater level simulations.
5. Our review identified Self-Organizing Maps (SOM), Growing Neural Gas (GNG), Fuzzy C-Means (FCM), and K-means as common clustering techniques, particularly in studies utilizing GRACE-derived Terrestrial Water Storage Anomaly (TWSA) data for downscaling to finer resolutions. These algorithms effectively clustered TWSA data to identify similar pixels for inputs and outputs, enhancing GWL understanding in areas with limited observations. SOM was also applied to optimize the number of hidden layer neurons, improving standalone model accuracy. However, most studies relied on trial and error for model structure, highlighting a gap in systematic research.
6. Among the 223 articles reviewed, the data splits for training and testing varied from 50%–50%, 56%–44% (Chang et al., 2015; Seidu et al., 2023; Heudorfer et al., 2023) to 90%–10% (Maheswaran and Khosa, 2013; Bahaa et al., 2015; Secci et al., 2023), with an 80/20 split being the most common. Most studies used monthly datasets spanning over 10 years, with sample sizes ranging from 35 (Mukherjee and Ramachandran, 2018), 48 (Guo et al., 2021), 23,850 (Taormina et al., 2012), 445,104 (Sahoo et al., 2017) 835,044 (Mohapatra et al., 2021) to 964,260 (Su et al., 2024). Larger datasets, particularly for training, generally improved model performance.
7. We recommend further exploration of modeled meteorological variables from IPCC climate change scenarios, such as Shared Socioeconomic Pathways (SSPs) and Representative Concentration Pathways (RCPs), as inputs for GWL models. Assessing how well these models simulate future conditions under climate change is crucial but currently underexplored. Only a hand full of studies explored this in detail (e.g. Javadinejad et al. (2020), Secci et al. (2023) and Osman et al. (2024)). Expanding this research could deepen our understanding of climate impacts on groundwater resources and improve future resource management strategies.
8. Finally, future work should explore combining physically based models with machine learning to benefit from the process representation of conceptual models and the pattern-learning strengths of ML. Recent work by Sun et al. (2023) demonstrated that integrating LSTM with physically based models can substantially improve prediction accuracy, especially when the physical model's performance is moderate or poor. Given the strong long-term forecasting performance of LSTM under data-scarce conditions (Gaffoor et al., 2022), it is a strong candidate for inclusion in such hybrid frameworks. This integration could enhance forecast robustness across varying time horizons, and model development should also move beyond trial-and-error by adopting more systematic optimization techniques.

## List of abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| ANFIS | Adaptive Neuro-Fuzzy Inference System |
| ACF | Autocorrelation Function |
| PACF | Partial Autocorrelation Function |
| MI | Mutual Information |
| SHAP | SHapley Additive exPlanations |
| RFE | Recursive Feature Elimination |
| PSO | Particle Swarm Optimization |
| GA | Genetic Algorithm |
| ACO | Ant Colony Optimization |
| CNN | Convolutional Neural Network |
| LSTM | Long Short-Term Memory |
| RF | Random Forest |
| SVM | Support Vector Machine |
| MLP | Multilayer Perceptron |
| GWL | Groundwater Level |
| $R^2$ | Coefficient of Determination |
| RMSE | Root Mean Square Error |
| MAE | Mean Absolute Error |
| MSE | Mean Squared Error |
| ME | Mean Error |
| RE | Relative Error |
| GCM | Global Climate Model |
| GRACE | Gravity Recovery and Climate Experiment |
| DA | Data Assimilation |
| WOA | Weed Optimization Algorithm |

| | |
|---|---|
| BWO | Black Widow Optimization |
| LM | Levenberg–Marquardt |
| BR | Bayesian Regularization |
| ARX | AutoRegressive with eXogenous inputs |
| FCM | Fuzzy C-Means |
| SOM | Self-Organizing Maps |
| GNG | Growing Neural Gas |
| SSPs | Shared Socioeconomic Pathways (SSPs) |
| RCPs | Representative Concentration Pathways |
| GOA | Grasshopper Optimization Algorithm |
| CSO | Cat Swarm Optimization |
| KHA | Krill Herd Algorithm |

## CRediT authorship contribution statement

**Gilbert Jesse:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cyril D. Boateng:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Jeffrey N.A. Aryee:** Writing – review & editing, Validation, Supervision. **Marian A. Osei:** Writing – review & editing, Validation, Supervision. **David D. Wemegah:** Writing – review & editing, Validation, Supervision. **Solomon S.R. Gidigasu:** Writing – review & editing, Validation, Supervision. **Akyana Britwum:** Writing – review & editing, Validation. **Samuel K. Afful:** Writing – review & editing, Validation, Data curation. **Haoulata Touré:** Writing – review & editing, Validation. **Vera Mensah:** Writing – review & editing, Validation. **Prinsca Owusu-Afriyie:** Writing – review & editing, Validation, Data curation.

## Computer code availability

No code or software has been developed for this research.

## Funding

Support for implementation of project activities was made possible by the Research Grant (109705-001/002) by the Responsible Artificial Intelligence Network for Climate Action in Africa (RAINCA) consortium made up of WASCAL, RUFORUM and AKADEMIYA 2063 provided by IDRC.

## Declaration of competing 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.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.acags.2025.100303.

## Data availability

No data was used for the research described in the article.

## References

Abdi, E., Ali, M., Santos, C.A.G., Olusola, A., Ghorbani, M.A., 2024. Enhancing groundwater level prediction accuracy using interpolation techniques in deep learning models. Groundw. Sustain. Dev. 26, 101213.
Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407 (1–4), 28–40.
Aderemi, B.A., Olwal, T.O., Ndambuki, J.M., Rwanga, S.S., 2023. Groundwater levels forecasting using machine learning models: A case study of the groundwater region 10 at Karst Belt, South Africa. Syst. Soft Comput. 5, 200049.
Afrifa, S., Zhang, T., Appiahene, P., Varadarajan, V., 2022. Mathematical and machine learning models for groundwater level changes: a systematic review and bibliographic analysis. Futur. Internet 14 (9), 259.
Ahmadi, A., Olyaei, M., Heydari, Z., Emami, M., Zeynolabedin, A., Ghomlaghi, A., Daccache, A., Fogg, G.E., Sadegh, M., 2022. Groundwater level modeling with machine learning: a systematic review and meta-analysis. Water 14 (6), 949.
Akbari Majd, A., Rasoulzadeh, A., Hasanpour Kashani, M., Kisi, O., et al., 2024. Enhancing the accuracy of metaheuristic neural networks in predicting underground water levels using meteorological data and remote sensing: A case study of Ardabil Plain, Iran.
Akinwande, M.O., Dikko, H.G., Samson, A., et al., 2015. Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. Open J. Stat. 5 (07), 754.
Ali, A.J., Ahmed, A.A., Abbod, M.F., 2024. Groundwater level predictions in the Thames Basin, London over extended horizons using transformers and advanced machine learning models. J. Clean. Prod. 484, 144300.
Ao, C., Zeng, W., Wu, L., Qian, L., Srivastava, A.K., Gaiser, T., 2021. Time-delayed machine learning models for estimating groundwater depth in the Hetao Irrigation District, China. Agricult. Water. Manag. 255, 107032.
Arabameri, A., Roy, J., Saha, S., Blaschke, T., Ghorbanzadeh, O., Tien Bui, D., 2019. Application of probabilistic and machine learning models for groundwater potentiality mapping in Damghan sedimentary plain, Iran. Remote. Sens. 11 (24), 3015.
Asgari, H.-R., Bozorg Haddad, O., Pazoki, M., Loáiciga, H.A., 2016. Weed optimization algorithm for optimal reservoir operation. J. Irrig. Drain. Eng. 142 (2), 04015055.
Attanasio, A., Pasini, A., Triacca, U., 2013. Granger causality analyses for climatic attribution. Atmospheric Clim. Sci. 2013.
Bahaa, K., Stefan, B., Jan, A., Ozga-Zieliński, B., Amanda, D., 2015. Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural network models. Hydrogeol. J. 23 (1).
Bai, T., Tahmasebi, P., 2023. Graph neural network for groundwater level forecasting. J. Hydrol. 616, 128792.
Banadkooki, F.B., Ehteram, M., Ahmed, A.N., Teo, F.Y., Fai, C.M., Afan, H.A., Sapitang, M., El-Shafie, A., 2020. Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. Nat. Resour. Res. 29, 3233–3252.
Banerjee, C., Mukherjee, T., Pasiliao, Jr., E., 2019. An empirical study on generalizations of the ReLU activation function. In: Proceedings of the 2019 ACM Southeast Conference. pp. 164–167.
Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.-X., Veyrat-Charvillon, N., 2011. Mutual information analysis: a comprehensive study. J. Cryptology 24 (2), 269–291.
Behzad, M., Asghari, K., Coppola, Jr., E.A., 2010. Comparative study of SVMs and ANNs in aquifer water level prediction. J. Comput. Civ. Eng. 24 (5), 408–413.
Bhadani, V., Singh, A., Kumar, V., Gaurav, K., 2024. Nature-inspired optimal tuning of input membership functions of fuzzy inference system for groundwater level prediction. Environ. Model. Softw. 175, 105995.
Bonkoungou, A.A., Zio, S., Sabane, A., Kafando, R., Kabore, A.K., Bissyande, T.F., 2024. A comparison of AI methods for groundwater level prediction in burkina faso. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, pp. 3–16.
Boo, K.B.W., El-Shafie, A., Othman, F., Khan, M.M.H., Birima, A.H., Ahmed, A.N., 2024. Groundwater level forecasting with machine learning models: A review. Water Res. 252, 121249.
Bourke, P., 1996. Cross correlation. Cross Correlation'', Auto Correlation—2D Pattern Identification 596.
Broomhead, D.S., Lowe, D., 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks. R. Signals Radar Establ. Malvern (UK) 25 (3), 1–8.
Brunner, P., Simmons, C.T., 2012. HydroGeoSphere: a fully integrated, physically based hydrological model. Ground Water 50 (2), 170–176.
Butler, Jr., J., Stotler, R., Whittemore, D., Reboulet, E., 2013. Interpretation of water level changes in the High Plains aquifer in western Kansas. Groundwater 51 (2), 180–190.
Chakraborty, S., Maity, P.K., Das, S., 2020. Investigation, simulation, identification and prediction of groundwater levels in coastal areas of Purba Midnapur, India, using MODFLOW. Environ. Dev. Sustain. 22, 3805–3837.
Chandirasekaran, D., Jayabarathi, T., 2019. Cat swarm algorithm in wireless sensor networks for optimized cluster head selection: a real time approach. Clust. Comput. 22, 11351–11361.

Chang, F.-J., Chang, L.-C., Huang, C.-W., Kao, I.-F., 2016. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. J. Hydrol. 541, 965–976.
Chang, J., Wang, G., Mao, T., 2015. Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. J. Hydrol. 529, 1211–1220.
Chen, L.-H., Chen, C.-T., Lin, D.-W., 2011. Application of integrated back-propagation network and self-organizing map for groundwater level forecasting. J. Water Resour. Plan. Manag. 137 (4), 352–365.
Chen, L.-H., Chen, C.-T., Pan, Y.-G., 2010. Groundwater level prediction using SOM-RBFN multisite model. J. Hydrol. Eng. 15 (8), 624–631.
Chen, C., He, W., Zhou, H., Xue, Y., Zhu, M., 2020. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. Sci. Rep. 10 (1), 3904.
Chen, X.-w., Jeong, J.C., 2007. Enhanced recursive feature elimination. In: Sixth International Conference on Machine Learning and Applications. ICMLA 2007, IEEE, pp. 429–435.
Chen, H.-Y., Vojinovic, Z., Lo, W., Lee, J.-W., 2023. Groundwater level prediction with deep learning methods. Water 15 (17), 3118.
Chenjia, Z., Xu, T., Zhang, Y., Ma, D., 2024. Deep learning models for groundwater level prediction based on delay penalty. Water Supply 24 (2), 555–567.
Chi, D.T.K., Thiem, D.D., Quynh, T.T.N., Nguyen, T.Q., 2024. Enhancing prediction accuracy and data handling for environmental applications in innovative modeling of underground water level fluctuations based on tree ensembles technique.
Chidepudi, S.K.R., Massei, N., Jardani, A., Henriot, A., Allier, D., Baulon, L., 2023. A wavelet-assisted deep learning approach for simulating groundwater levels affected by low-frequency variability. Sci. Total Environ. 865, 161035.
Choubin, B., Malekian, A., 2017. Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. Environ. Earth Sci. 76, 1–10.
Collados-Lara, A.J., Pulido-Velazquez, D., Ruiz, L., Pegalajar, M., Pardo-Igúzquiza, E., Baena-Ruiz, L., 2023. A parsimonious methodological framework for short-term forecasting of groundwater levels. Sci. Total Environ. 881, 163328.
Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. J. Hydrol. 309 (1–4), 229–240.
Dasgupta, R., Banerjee, G., Hidayetullah, S.M., Saha, N., Das, S., Mazumdar, A., 2024. A comparative analysis of statistical, MCDM and machine learning based modification strategies to reduce subjective errors of DRASTIC models. Environ. Earth Sci. 83 (7), 211.
Dash, N.B., Panda, S.N., Remesan, R., Sahoo, N., 2010. Hybrid neural modeling for groundwater level prediction. Neural Comput. Appl. 19, 1251–1263.
Dehghani, R., Poudeh, H.T., Izadi, Z., 2022. The effect of climate change on groundwater level and its prediction using modern meta-heuristic model. Groundw. Sustain. Dev. 16, 100702.
Derbela, M., Nouiri, I., 2020. Intelligent approach to predict future groundwater level based on artificial neural networks (ANN). Euro-Mediterranean J. Environ. Integr. 5, 1–11.
Ding, B., Qian, H., Zhou, J., 2018. Activation functions and their characteristics in deep neural networks. In: 2018 Chinese Control and Decision Conference. CCDC, IEEE, pp. 1836–1841.
Dokeroglu, T., Deniz, A., Kiziloz, H.E., 2022. A comprehensive survey on recent metaheuristics for feature selection. Neurocomputing 494, 269–296.
Ebrahimi, H., Rajaee, T., 2017. Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. Glob. Planet. Change 148, 181–191.
Eghrari, Z., Delavar, M., Zare, M., Mousavi, M., Nazari, B., Ghaffarian, S., 2023. Groundwater level prediction using deep recurrent neural networks and uncertainty assessment. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. 10, 493–500.
Ehteram, M., et al., 2023. An advanced deep learning model for predicting groundwater level.
El Ibrahimi, A., Baali, A., Couscous, A., Hamdani, N., et al., 2017. Comparative study of the three models (ANN-PMC),(DWT-ANN-PMC) and (MLR) for prediction of the groundwater level of the surface water table in the Saïss Plain (North of Morocco). Int. J. Intell. Eng. Syst. 10 (5).
Elmotawakkil, A., Sadiki, A., Enneya, N., 2024. Predicting groundwater level based on remote sensing and machine learning: a case study in the Rabat–Kénitra region. J. Hydroinformatics 26 (10), 2639–2667.
Elzain, H.E., Abdalla, O., Al-Maktoumi, A., Kacimov, A., Eltayeb, M., 2024. A novel approach to forecast water table rise in arid regions using stacked ensemble machine learning and deep artificial intelligence models. J. Hydrol. 640, 131668.
Emamgholizadeh, S., Moslemi, K., Karami, G., 2014. Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). Water Resour. Manag. 28, 5433–5446.
Fahim, A.K.F., Kamal, A.M., Shahid, S., 2024. Modeling spatial groundwater level patterns of Bangladesh using physio-climatic variables and machine learning algorithms. Groundw. Sustain. Dev. 25, 101142.
Fallah-Mehdipour, E., Haddad, O.B., Mariño, M., 2013. Prediction and simulation of monthly groundwater levels by genetic programming. J. Hydro-Environ. Res. 7 (4), 253–260.
Famiglietti, J.S., 2014. The global groundwater crisis. Nat. Clim. Chang. 4 (11), 945–948.
Fei, S.-w., Wang, M.-J., Miao, Y.-b., Tu, J., Liu, C.-l., 2009. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. Energy Convers. Manage. 50 (6), 1604–1609.
Feng, F., Ghorbani, H., Radwan, A.E., 2024. Predicting groundwater level using traditional and deep machine learning algorithms. Front. Environ. Sci. 12, 1291327.
Foroumandi, E., Nourani, V., Huang, J.J., Moradkhani, H., 2023. Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: A deep learning approach. J. Hydrol. 616, 128838.
Fronzi, D., Narang, G., Galdelli, A., Pepi, A., Mancini, A., Tazioli, A., 2024. Towards groundwater-level prediction using prophet forecasting method by exploiting a high-resolution hydrogeological monitoring system. Water 16 (1), 152.
Gaffoor, Z., Pietersen, K., Jovanovic, N., Bagula, A., Kanyerere, T., Ajayi, O., Wanangwa, G., 2022. A comparison of ensemble and deep learning algorithms to model groundwater levels in a data-scarce aquifer of Southern Africa. Hydrology 9 (7), 125.
Gandomi, A.H., Alavi, A.H., 2012. Krill herd: a new bio-inspired optimization algorithm. Commun. Nonlinear Sci. Numer. Simul. 17 (12), 4831–4845.
Ghafoor, H., Umer, R., Muhammad, J., Rauf, Z., Khan, P., et al., 2024. Predicting groundwater levels at Colorado state of USA using ARIMA and ANN models. In: 2024 2nd International Conference on Foundation and Large Language Models. FLLM, IEEE, pp. 188–193.
Ghazi, B., Jeihouni, E., Kalantari, Z., 2021. Predicting groundwater level fluctuations under climate change scenarios for Tasuj Plain, Iran. Arab. J. Geosci. 14 (2), 115.
Gholami, V., Chau, K.W., Fadaee, F., Torkaman, J., Ghaffari, A., 2015. Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers. J. Hydrol. 529, 1060–1069.
Ghose, D., Das, U., Roy, P., 2018. Modeling response of runoff and evapotranspiration for predicting water table depth in arid region using dynamic recurrent neural network. Groundw. Sustain. Dev. 6, 263–269.
Gong, Y., Zhang, Y., Lan, S., Wang, H., 2016. A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida. Water Resour. Manag. 30, 375–391.
Gonzalez, R.Q., Arsanjani, J.J., 2021. Prediction of groundwater level variations in a changing climate: A Danish case study. ISPRS Int. J. Geo-Inf. 10 (11), 792.
Guo, F., Yang, J., Li, H., Li, G., Zhang, Z., 2021. A convLSTM conjunction model for groundwater level forecasting in a karst aquifer considering connectivity characteristics. Water 13 (19), 2759.
Gupta, S.K., Sahoo, S., Sahoo, B.B., Srivastava, P.K., Pateriya, B., Santosh, D., 2024. Prediction of groundwater level changes based on machine learning technique in highly groundwater irrigated alluvial aquifers of south-central Punjab, India. Phys. Chem. Earth, Parts A/B/C 135, 103603.
Guzman, S.M., Paz, J.O., Tagert, M.L.M., 2017. The use of NARX neural networks to forecast daily groundwater levels. Water Resour. Manag. 31, 1591–1603.
Guzman, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A., 2015. Artificial neural networks and support vector machines: Contrast study for groundwater level prediction. In: 2015 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.
Guzman, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A.E., 2019. Evaluation of seasonally classified inputs for the prediction of daily groundwater levels: NARX networks vs support vector machines. Environ. Model. Assess. 24, 223–234.
Haq, M.A., Jilani, A.K., Prabu, P., 2021. Deep learning based modeling of groundwater storage change. CMC-Comput. Mater. Contin. 70, 4599–4617.
Heudorfer, B., Liesch, T., Broda, S., 2023. On the challenges of global entity-aware deep learning models for groundwater level prediction. Hydrol. Earth Syst. Sci. Discuss. 2023, 1–28.
Hikouei, I.S., Eshleman, K.N., Saharjo, B.H., Graham, L.L., Applegate, G., Cochrane, M.A., 2023. Using machine learning algorithms to predict groundwater levels in Indonesian tropical peatlands. Sci. Total Environ. 857, 159701.
Hosseini, Z., Gharechelou, S., Nakhaei, M., Gharechelou, S., 2016. Optimal design of BP algorithm by ACO R model for groundwater-level forecasting: A case study on Shabestar Plain, Iran. Arab. J. Geosci. 9, 1–16.
Huang, F., Huang, J., Jiang, S.-H., Zhou, C., 2017. Prediction of groundwater levels using evidence of chaos and support vector machine. J. Hydroinformatics 19 (4), 586–606.
Hussein, E.A., Thron, C., Ghaziasgar, M., Bagula, A., Vaccari, M., 2020. Groundwater prediction using machine-learning tools. Algorithms 13 (11), 300.
Idrizovic, D., Pocuca, V., Mandic, M.V., Djurovic, N., Matovic, G., Gregoric, E., 2020. Impact of climate change on water resource availability in a mountainous catchment: a case study of the Toplica River catchment, Serbia. J. Hydrol. 587, 124992.
Iranzad, R., Liu, X., 2024. A review of random forest-based feature selection methods for data science education and applications. Int. J. Data Sci. Anal. 1–15.
Jalalkamali, A., Sedghi, H., Manshouri, M., 2011. Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman Plain, Iran. J. Hydroinformatics 13 (4), 867–876.
Jamel, T.M., Khammas, B.M., 2012. Implementation of a sigmoid activation function for neural network using FPGA. In: 13th Scientific Conference of Al-Ma'Moon University College. vol. 13, pp. 1589–1591.

Javadinejad, S., Dara, R., Jafary, F., 2020. How groundwater level can predict under the effect of climate change by using artificial neural networks of NARX. Resour. Environ. Inf. Eng. 2 (1), 90–99.
Jeong, J., Park, E., 2019. Comparative applications of data-driven models representing water table fluctuations. J. Hydrol. 572, 261–273.
Jing, H., He, X., Tian, Y., Lancia, M., Cao, G., Crivellari, A., Guo, Z., Zheng, C., 2023. Comparison and interpretation of data-driven models for simulating site-specific human-impacted groundwater dynamics in the north China Plain. J. Hydrol. 616, 128751.
John, B., Das, S., 2020. Identification of risk zone area of declining piezometric level in the urbanized regions around the City of Kolkata based on ground investigation and GIS techniques. Groundw. Sustain. Dev. 11, 100354.
John, B., Das, S., Das, R., 2023. Natural groundwater level fluctuations of Kolkata City based on seasonal field data and population growth using geo-spatial application and characterised statistical techniques. Environ. Dev. Sustain. 25 (7), 6503–6528.
Jović, A., Brkić, K., Bogunović, N., 2015. A review of feature selection methods with applications. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics. MIPRO, IEEE, pp. 1200–1205.
Kajewska-Szkudlarek, J., Kubicz, J., Kajewski, I., 2022. Correlation approach in predictor selection for groundwater level forecasting in areas threatened by water deficits. J. Hydroinformatics 24 (1), 143–159.
Kalaiselvi, T., Padmapriya, S., Somasundaram, K., Praveenkumar, S., 2022. E-Tanh: a novel activation function for image processing neural network models. Neural Comput. Appl. 34 (19), 16563–16575.
Kalu, I., Ndehedehe, C.E., Okwuashi, O., Eyoh, A.E., Ferreira, V.G., 2022. A new modelling framework to assess changes in groundwater level. J. Hydrol.: Reg. Stud. 43, 101185.
Kamalov, F., Elnaffar, S., Cherukuri, A., Jonnalagadda, A., 2024. Forward feature selection: empirical analysis. J. Intell. Syst. Internet Things 11 (1), 44–54.
Kanji, S., Das, S., 2025. Assessing groundwater potentialities and replenishment feasibility using machine learning and MCDM models considering hydro-geological aspects and water quality constituents. Environ. Earth Sci. 84 (1), 16.
Karthikeyan, L., Kumar, D.N., Graillot, D., Gaur, S., 2013. Prediction of ground water levels in the uplands of a tropical coastal riparian wetland using artificial neural networks. Water Resour. Manag. 27, 871–883.
Kayhomayoon, Z., Babaeian, F., Ghordoyee Milan, S., Arya Azar, N., Berndtsson, R., 2022. A combination of metaheuristic optimization algorithms and machine learning methods improves the prediction of groundwater level. Water 14 (5), 751.
Kayhomayoon, Z., Jamnani, M.R., Rashidi, S., Milan, S.G., Azar, N.A., Berndtsson, R., 2023. Soft computing assessment of current and future groundwater resources under CMIP6 scenarios in northwestern Iran. Agricult. Water. Manag. 285, 108369.
Kim, D., Jang, C., Choi, J., Kwak, J., 2023. A case study: Groundwater level forecasting of the gyorae area in actual practice on Jeju island using deep-learning technique. Water 15 (5), 972.
King, J.R., Jackson, D.A., 1999. Variable selection in large environmental data sets using principal components analysis. Environmetrics: Off. J. Int. Environmetrics Soc. 10 (1), 67–77.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kouziokas, G.N., Chatzigeorgiou, A., Perakis, K., 2018. Multilayer feed forward models in groundwater level forecasting using meteorological data in public management. Water Resour. Manag. 32 (15), 5041–5052.
Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Phys. Rev. E—Statist. Nonlinear Soft Matter Phys. 69 (6), 066138.
Kuhn, M., Johnson, K., Kuhn, M., Johnson, K., 2013a. Data pre-processing. Appl. Predict. Model. 27–59.
Kuhn, M., Johnson, K., et al., 2013b. Applied Predictive Modeling, vol. 26, Springer.
LaBianca, A., Koch, J., Jensen, K.H., Sonnenborg, T.O., Kidmose, J., 2024. Machine learning for predicting shallow groundwater levels in urban areas. J. Hydrol. 632, 130902.
Latham, P.E., Roudi, Y., 2009. Mutual information. Scholarpedia 4 (1), 1658.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, S., Lee, K.-K., Yoon, H., 2019. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeol. J. 27 (2).
Lendzioch, T., Langhammer, J., Vlček, L., Minařík, R., 2021. Mapping the groundwater level and soil moisture of a montane peat bog using uav monitoring and machine learning. Remote. Sens. 13 (5), 907.
Li, W., Finsa, M.M., Laskey, K.B., Houser, P., Douglas-Bate, R., 2023. Groundwater level prediction with machine learning to support sustainable irrigation in water scarcity regions. Water 15 (19), 3473.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health-care interventions: explanation and elaboration. BMJ 339, http://dx.doi.org/10.1136/bmj.b2700, URL: https://www.bmj.com/content/339/bmj.b2700. arXiv:https://www.bmj.com/content/339/bmj.b2700.full.pdf.
Lin, S.-W., Ying, K.-C., Chen, S.-C., Lee, Z.-J., 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst. Appl. 35 (4), 1817–1824.
Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., Hu, B.X., 2022. Simulation of regional groundwater levels in arid regions using interpretable machine learning models. Sci. Total Environ. 831, 154902.
Liu, D., Mishra, A.K., Yu, Z., Lü, H., Li, Y., 2021. Support vector machine and data assimilation framework for groundwater level forecasting using GRACE satellite data. J. Hydrol. 603, 126929.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Mafarja, M., Aljarah, I., Faris, H., Hammouri, A.I., Ala'M, A.-Z., Mirjalili, S., 2019. Binary grasshopper optimisation algorithm approaches for feature selection problems. Expert Syst. Appl. 117, 267–286.
Maheswaran, R., Khosa, R., 2013. Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics. Comput. Geosci. 52, 422–436.
Maiti, S., Tiwari, R., 2014. A comparative study of artificial neural networks, Bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction. Environ. Earth Sci. 71, 3147–3160.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 674–693.
Manna, T., Anitha, A., 2023. Deep ensemble-based approach using randomized low-rank approximation for sustainable groundwater level prediction. Appl. Sci. 13 (5), 3210.
Maraun, D., Kurths, J., 2004. Cross wavelet analysis: significance testing and pitfalls. Nonlinear Process. Geophys. 11 (4), 505–514.
Martínez-Muñoz, G., Suárez, A., 2010. Out-of-bag estimation of the optimal sample size in bagging. Pattern Recognit. 43 (1), 143–152.
May-Lagunes, G., Chau, V., Ellestad, E., Greengard, L., D'Odorico, P., Vahabi, P., Todeschini, A., Girotto, M., 2023. Forecasting groundwater levels using machine learning methods: The case of California's central valley. J. Hydrol. X 21, 100161.
Mehrabian, A.R., Lucas, C., 2006. A novel numerical optimization algorithm inspired from weed colonization. Ecol. Inform. 1 (4), 355–366.
Mendeley, 2022. Mendeley. https://www.mendeley.com. (Accessed 17 July 2023).
Milan, S.G., Kayhomayoon, Z., Azar, N.A., Berndtsson, R., Ramezani, M.R., Moghaddam, H.K., 2023. Using machine learning to determine acceptable levels of groundwater consumption in Iran. Sustain. Prod. Consum. 35, 388–400.
Mirzavand, M., Khoshnevisan, B., Shamshirband, S., Kisi, O., Ahmad, R., Akib, S., et al., 2015. Evaluating groundwater level fluctuation by support vector regression and neuro-fuzzy methods: a comparative study. Nat. Hazards 1 (1), 1–15.
Mohanty, S., Jha, M.K., Kumar, A., Sudheer, K., 2010. Artificial neural network modeling for groundwater level forecasting in a river island of eastern India. Water Resour. Manag. 24, 1845–1865.
Mohapatra, J.B., Jha, P., Jha, M.K., Biswal, S., 2021. Efficacy of machine learning techniques in predicting groundwater fluctuations in agro-ecological zones of India. Sci. Total Environ. 785, 147319.
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 (6), 914–944.
Moradi, A., Akhtari, A.A., Azari, A., 2023. Prediction of groundwater level fluctuation using methods based on machine learning and numerical model. J. Appl. Res. Water Wastewater 10 (1), 20–28.
Moravej, M., Amani, P., Hosseini-Moghari, S.-M., 2020. Groundwater level simulation and forecasting using interior search algorithm-least square support vector regression (ISA-LSSVR). Groundw. Sustain. Dev. 11, 100447.
Mosavi, A., Sajedi Hosseini, F., Choubin, B., Goodarzi, M., Dineva, A.A., Rafiei Sardooi, E., 2021. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. Water Resour. Manag. 35, 23–37.
Motagh, M., Shamshiri, R., Haghighi, M.H., Wetzel, H.-U., Akbari, B., Nahavandchi, H., Roessner, S., Arabi, S., 2017. Quantifying groundwater exploitation induced subsidence in the Rafsanjan Plain, southeastern Iran, using InSAR time-series and in situ measurements. Eng. Geol. 218, 134–151.
Mukherjee, A., Ramachandran, P., 2018. Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: Analysis of comparative performances of SVR, ANN and LRM. J. Hydrol. 558, 647–658.
Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., Agarwal, D., 2021. Surrogate optimization of deep neural networks for groundwater predictions. J. Global Optim. 81, 203–231.
Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. Int. J. Appl. Earth Obs. Geoinf. 18, 399–406.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. ICML-10, pp. 807–814.
Nan, T., Cao, W., Wang, Z., Gao, Y., Zhao, L., Sun, X., Na, J., 2023. Evaluation of shallow groundwater dynamics after water supplement in north China Plain based on attention-GRU model. J. Hydrol. 625, 130085.
Nand, V., Narjary, B., Singh, V.K., Kumar, N., Islam, A., Kumar, S., 2024. Reliability of artificial intelligence-based models compared to numerical model for predicting groundwater level under changing climate. J. Agric. Eng. 61 (spl), 339–356.
Narayan, S., 1997. The generalized sigmoid activation function: Competitive supervised learning. Inform. Sci. 99 (1–2), 69–82.

Nie, S., Bian, J., Wan, H., Sun, X., Zhang, B., 2017. Simulation and uncertainty analysis for groundwater levels using radial basis function neural network and support vector machine models. J. Water Supply: Res. Technol.—AQUA 66 (1), 15–24.

Niu, X., Lu, C., Zhang, Y., Zhang, Y., Wu, C., Saidy, E., Liu, B., Shu, L., 2023. Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values. Sci. Total Environ. 904, 166662.

Nourani, V., Alami, M.T., Vousoughi, F.D., 2015. Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. J. Hydrol. 524, 255–269.

Nourani, V., Ejlali, R.G., Alami, M.T., 2011. Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study. Environ. Eng. Sci. 28 (3), 217–228.

Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378.

Omiotek, Z., Stepanchenko, O., Wójcik, W., Legieć, W., Szatkowska, M., 2019. The use of the Hellwig's method for feature selection in the detection of myeloma bone destruction based on radiographic images. Biocybern. Biomed. Eng. 39 (2), 328–338.

Osman, A.I.A., Latif, S.D., Boo, K.B.W., Ahmed, A.N., Huang, Y.F., El-Shafie, A., 2024. Advanced machine learning algorithm to predict the implication of climate change on groundwater level for protecting aquifer from depletion. Groundw. Sustain. Dev. 25, 101152.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.

Panahi, G., Hassanzadeh Eskafi, M., Faridhosseini, A., Khodashenas, S.R., Rohani, A., 2023. Prediction of groundwater level fluctuations under climate change based on machine learning algorithms in the Mashhad aquifer, Iran. J. Water Clim. Chang. 14 (3), 1039–1059.

Panja, S., Chatterjee, A., Yasmin, G., 2019. Kernel functions of svm: A comparison and optimal solution. In: Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2. Springer, pp. 88–97.

Patle, A., Chouhan, D.S., 2013. SVM kernel functions for classification. In: 2013 International Conference on Advances in Technology and Engineering. ICATE, IEEE, pp. 1–9.

Patra, S.R., Chu, H.-J., et al., 2023. Regional groundwater sequential forecasting using global and local LSTM models. J. Hydrol.: Reg. Stud. 47, 101442.

Pham, Q.B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A.R.M.T., Talukdar, S., Nguyen, X.C., Ahmed, A.N., Anh, D.T., 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. Neural Comput. Appl. 34 (13), 10751–10773.

Piesse, M., 2020. Global water supply and demand trends point towards rising water insecurity.

Pudjihartono, N., Fadason, T., Kempa-Liehr, A.W., O'Sullivan, J.M., 2022. A review of feature selection methods for machine learning-based disease risk prediction. Front. Bioinform. 2, 927312.

Puente, J., Gomez, A., Fernández, I., Priore, P., 2009. Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms. Comput. Ind. Eng. 56 (4), 1232–1242.

Rafik, A., Ait Brahim, Y., Amazirh, A., Ouarani, M., Bargam, B., Ouatiki, H., Bouslihim, Y., Bouchaou, L., Chehbouni, A., 2023. Groundwater level forecasting in a data-scarce region through remote sensing data downscaling, hydrological modeling, and machine learning: A case study from Morocco. J. Hydrol.: Reg. Stud. 50, 101569.

Rahmadani, F., Lee, H., 2020. ODE-based epidemic network simulation of viral Hepatitis A and kernel support vector machine based vaccination effect analysis. Journal of Korean Institute of Intelligent Systems 30 (2), 106–112.

Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. J. Hydrol. 572, 336–351.

Rakhshandehroo, G.R., Vaghefi, M., Aghbolaghi, M.A., 2012. Forecasting groundwater level in Shiraz Plain using artificial neural networks. Arab. J. Sci. Eng. 37, 1871–1883.

Ramachandran, P., Zoph, B., Le, Q.V., 2017. Searching for activation functions. arXiv preprint arXiv:1710.05941.

Ramsey, F.L., 1974. Characterization of the partial autocorrelation function. Ann. Statist. 1296–1301.

Rasamoelina, A.D., Adjailia, F., Sinčák, P., 2020. A review of activation function for artificial neural network. In: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics. SAMI, IEEE, pp. 281–286.

Roy, D.K., Munmun, T.H., Paul, C.R., Haque, M.P., Al-Ansari, N., Mattar, M.A., 2023. Improving forecasting accuracy of multi-scale groundwater level fluctuations using a heterogeneous ensemble of machine learning algorithms. Water 15 (20), 3624.

Sahoo, S., Jha, M.K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol. J. 21 (8), 1865.

Sahoo, S., Russo, T., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour. Res. 53 (5), 3878–3895.

Saremi, S., Mirjalili, S., Lewis, A., 2017. Grasshopper optimisation algorithm: theory and application. Adv. Eng. Softw. 105, 30–47.

Sarkar, H., Goriwale, S.S., Ghosh, J.K., Ojha, C.S.P., Ghosh, S.K., 2024. Potential of machine learning algorithms in groundwater level prediction using temporal gravity data. Groundw. Sustain. Dev. 25, 101114.

Saroughi, M., Mirzania, E., Achite, M., Katipoğlu, O.M., Al-Ansari, N., Vishwakarma, D.K., Chung, I.-M., Alreshidi, M.A., Yadav, K.K., 2024. Evaluate effect of 126 pre-processing methods on various artificial intelligence models accuracy versus normal mode to predict groundwater level (case study: Hamedan–Bahar Plain, Iran). Heliyon 10 (7).

Scholkopf, B., Sung, K.-K., Burges, C.J., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V., 1997. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans. Signal Process. 45 (11), 2758–2765.

Secci, D., Tanda, M.G., D'Oria, M., Todaro, V., 2023. Artificial intelligence models to evaluate the impact of climate change on groundwater resources. J. Hydrol. 627, 130359.

Seidu, J., Ewusi, A., Kuma, J.S.Y., Ziggah, Y.Y., Voigt, H.-J., 2023. Impact of data partitioning in groundwater level prediction using artificial neural network for multiple wells. Int. J. River Basin Manag. 21 (4), 639–650.

Seifi, A., Ehteram, M., Singh, V.P., Mosavi, A., 2020. Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. Sustainability 12 (10), 4023.

Seifi, A., Pourebrahim, S., Ehteram, M., Shabanian, H., 2024. A robust multi-model framework for groundwater level prediction: The BFSA-MVMD-GRU-RVM model. Results Eng. 24, 103250.

Sharafati, A., Asadollah, S.B.H.S., Neshat, A., 2020. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. J. Hydrol. 591, 125468.

Sharghi, E., Nourani, V., Zhang, Y., Ghaneei, P., 2022. Conjunction of cluster ensemble-model ensemble techniques for spatiotemporal assessment of groundwater depletion in semi-arid plains. J. Hydrol. 610, 127984.

Shirmohammadi, B., Vafakhah, M., Moosavi, V., Moghaddamnia, A., 2013. Application of several data-driven techniques for predicting groundwater level. Water Resour. Manag. 27, 419–432.

Singh, A., Patel, S., Bhadani, V., Kumar, V., Gaurav, K., 2024. AutoML-GWL: automated machine learning model for the prediction of groundwater level. Eng. Appl. Artif. Intell. 127, 107405.

Solgi, R., Loaiciga, H.A., Kram, M., 2021. Long short-term memory neural network (LSTM-NN) for aquifer level time series forecasting using in-situ piezometric observations. J. Hydrol. 601, 126800.

Solomatine, D.P., 2006. Data-driven modeling and computational intelligence methods in hydrology. Encycl. Hydrol. Sci..

Sriram, R., et al., 2023. Groundwater level prediction: A novel study on machine learning based approach with regression models for sustainable resource management. In: 2023 IEEE International Conference on Cloud Computing in Emerging Markets. CCEM, IEEE, pp. 137–142.

Stokes, P.A., Purdon, P.L., 2017. A study of problems encountered in granger causality analysis from a neuroscience perspective. Proc. Natl. Acad. Sci. 114 (34), E7063–E7072.

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8 (1), 25.

Su, Y.-S., Hu, Y.-C., Wu, Y.-C., Lo, C.-T., 2024. Evaluating the impact of pumping on groundwater level prediction in the chuoshui river alluvial fan using artificial intelligence techniques.

Sun, W., Chang, L.-C., Chang, F.-J., 2024. Deep dive into predictive excellence: Transformer's impact on groundwater level prediction. J. Hydrol. 636, 131250.

Sun, K., Hu, L., Sun, J., Cao, X., 2023. Enhancing groundwater level prediction accuracy at a daily scale through combined machine learning and physics-based modeling. J. Hydrol.: Reg. Stud. 50, 101577.

Sun, Y., Wendi, D., Kim, D.E., Liong, S.-Y., 2016. Application of artificial neural networks in groundwater table forecasting–a case study in a Singapore swamp forest. Hydrol. Earth Syst. Sci. 20 (4), 1405–1412.

Suryanarayana, C., Sudheer, C., Mahammood, V., Panigrahi, B.K., 2014. An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing 145, 324–335.

Takefuji, Y., 2025. Addressing feature importance biases in machine learning models for early diagnosis of type 1 Gaucher disease. J. Clin. Epidemiol. 178.

Talpur, N., Salleh, M.N.M., Hussain, K., 2017. An investigation of membership functions on performance of ANFIS for solving classification problems. In: IOP Conference Series: Materials Science and Engineering. vol. 226, IOP Publishing, 012103, 1.

Tang, J., Alelyani, S., Liu, H., 2014. Feature selection for classification: A review. Data Classif.: Algorithms Appl. 37.

Tang, W., Wu, Q., 2009. Biologically inspired optimization: a review. Trans. Inst. Meas. Control 31 (6), 495–515.

Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., et al., 2022. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing 489, 271–308.

Taormina, R., Chau, K.-w., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Eng. Appl. Artif. Intell. 25 (8), 1670–1676.
Thompson, C.G., Kim, R.S., Aloe, A.M., Becker, B.J., 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl. Soc. Psychol. 39 (2), 81–90.
Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79 (1), 61–78.
Trichakis, I.C., Nikolos, I.K., Karatzas, G., 2011. Artificial neural network (ANN) based modeling for karstic groundwater level simulation. Water Resour. Manag. 25, 1143–1152.
Tyagi, K., Rane, C., Manry, M., 2022. Supervised learning. In: Artificial Intelligence and Machine Learning for EDGE Computing. Elsevier, pp. 3–22.
Van Thieu, N., Barma, S.D., Van Lam, T., Kisi, O., Mahesha, A., 2023. Groundwater level modeling using augmented artificial ecosystem optimization. J. Hydrol. 617, 129034.
Veyrat-Charvillon, N., Standaert, F.-X., 2009. Mutual information analysis: how, when and why? In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, pp. 429–443.
Vu, M.T., Jardani, A., Massei, N., Deloffre, J., Fournier, M., Laignel, B., 2023. Long-run forecasting surface and groundwater dynamics from intermittent observation data: An evaluation for 50 years. Sci. Total Environ. 880, 163338.
Wang, Y., Guo, F., Chen, S., Zhang, H., Zhang, Z., Li, A., 2024. Quantitative study of rainfall lag effects and integration of machine learning methods for groundwater level prediction modelling. Hydrol. Process. 38 (5), e15171.
Wang, D., Tan, D., Liu, L., 2018. Particle swarm optimization algorithm: an overview. Soft Comput. 22 (2), 387–408.
Wei, A., Li, X., Yan, L., Wang, Z., Yu, X., 2023a. Machine learning models combined with wavelet transform and phase space reconstruction for groundwater level forecasting. Comput. Geosci. 177, 105386.
Wei, Y., Wang, F., Hong, B., Yang, S., 2023b. Revealing spatial variability of groundwater level in typical ecosystems of the Tarim Basin through ensemble algorithms and limited observations. J. Hydrol. 620, 129399.
Wen, X., Feng, Q., Deo, R.C., Wu, M., Si, J., 2017. Wavelet analysis–artificial neural network conjunction models for multi-scale monthly groundwater level predicting in an arid inland river basin, northwestern China. Hydrol. Res. 48 (6), 1710–1729.
Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environ. Model. Softw. 54, 108–127.
Wu, M., Feng, Q., Wen, X., Yin, Z., Yang, L., Sheng, D., 2021b. Deterministic analysis and uncertainty analysis of ensemble forecasting model based on variational mode decomposition for estimation of monthly groundwater level. Water 13 (2), 139.
Wu, Z., Lu, C., Sun, Q., Lu, W., He, X., Qin, T., Yan, L., Wu, C., 2023. Predicting groundwater level based on machine learning: A case study of the Hebei Plain. Water 15 (4), 823.
Wu, C., Zhang, X., Wang, W., Lu, C., Zhang, Y., Qin, W., Tick, G.R., Liu, B., Shu, L., 2021a. Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model. Sci. Total Environ. 783, 146948.
Wunsch, A., Liesch, T., Broda, S., 2018. Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). J. Hydrol. 567, 743–758.
Xie, X., Zhang, X., 2024. Development of a deep surrogate model with spatiotemporal characteristics mining capabilities for the prediction of groundwater level in coastal areas. J. Environ. Manag. 370, 122724.
Yadav, B., Gupta, P.K., Patidar, N., Himanshu, S.K., 2020. Ensemble modelling framework for groundwater level prediction in urban areas of India. Sci. Total Environ. 712, 135539.
Yan, Q., Ma, C., 2016. Application of integrated ARIMA and RBF network for groundwater level forecasting. Environ. Earth Sci. 75, 1–13.
Yi, S., Kondolf, G.M., Sandoval Solis, S., Dale, L., 2024. Groundwater level forecasting using machine learning: a case study of the Baekje Weir in Four Major Rivers Project, South Korea. Water Resour. Res. 60 (5), e2022WR032779.
Yin, J., Medellín-Azuara, J., Escriva-Bou, A., Liu, Z., 2021. Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change. Sci. Total Environ. 769, 144715.
Yoo, J.-C., Han, T.H., 2009. Fast normalized cross-correlation. Circuits Systems Signal Process. 28, 819–843.
Yoon, H., Hyun, Y., Ha, K., Lee, K.-K., Kim, G.-B., 2016. A method to improve the stability and accuracy of ANN-and SVM-based time series models for long-term groundwater level predictions. Comput. Geosci. 90, 144–155.
Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol. 396 (1–2), 128–138.
Yu, S., Qiu, L., Xu, X., Yang, Y.-S., 2021. Machine learning-based algorithm for predicting the groundwater level in Minqin Oasis region of China. In: 2021 7th International Conference on Hydraulic and Civil Engineering & Smart Water Conservancy and Intelligent Disaster Reduction Forum. ICHCE & SWIDR, IEEE, pp. 89–95.
Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J., Wu, M., 2018. Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, northwest China. Water Resour. Manag. 32 (1), 301–323.
Zhang, X., Dong, F., Chen, G., Dai, Z., 2023. Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks. Hydrol. Earth Syst. Sci. 27 (1), 83–96.
Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018. Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929.
Zheng, H., Hou, H., Qin, Z., 2024. Research on a non-stationary groundwater level prediction model based on VMD-itransformer and its application in sustainable water resource management of ecological reserves. Sustainability 16 (21), 9185.
Zhou, T., Wang, F., Yang, Z., 2017. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. Water 9 (10), 781.
Zhou, T., Wen, X., Feng, Q., Yu, H., Xi, H., 2022. Bayesian model averaging ensemble approach for multi-time-ahead groundwater level prediction combining the GRACE, GLEAM, and GLDAS data in arid areas. Remote. Sens. 15 (1), 188.
Zhou, Y., Zhang, Q., Bai, G., Zhao, H., Shuai, G., Cui, Y., Shao, J., 2024. Groundwater dynamics clustering and prediction based on grey relational analysis and LSTM model: a case study in Beijing Plain, China. J. Hydrol.: Reg. Stud. 56, 102011.
Zhu, F., Han, M., Sun, Y., Zeng, Y., Zhao, L., Zhu, O., Hou, T., Zhong, P.-a., 2024. A machine learning framework for multi-step-ahead prediction of groundwater levels in agricultural regions with high reliance on groundwater irrigation. Environ. Model. Softw. 180, 106146.
Zowam, F.J., Milewski, A.M., 2024. Groundwater level prediction using machine learning and geostatistical interpolation models. Water 16 (19), 2771.