

Insight into groundwater level prediction with feature effectiveness: comparison of machine learning and numerical methods

Mohammad Reza Goodarzi ^{a,b,*}, Hamidreza Bana Bafrouei^c and Majid Vazirian^b

^a Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

^b Department of Civil Engineering, Yazd University, Yazd, Iran

^c Department of Civil, Water and Environmental Engineering, University of Shahid Beheshti, Tehran, Iran

*Corresponding author. E-mail: Goodarzimr@yazd.ac.ir

 MRG, 0000-0002-6296-2315

ABSTRACT

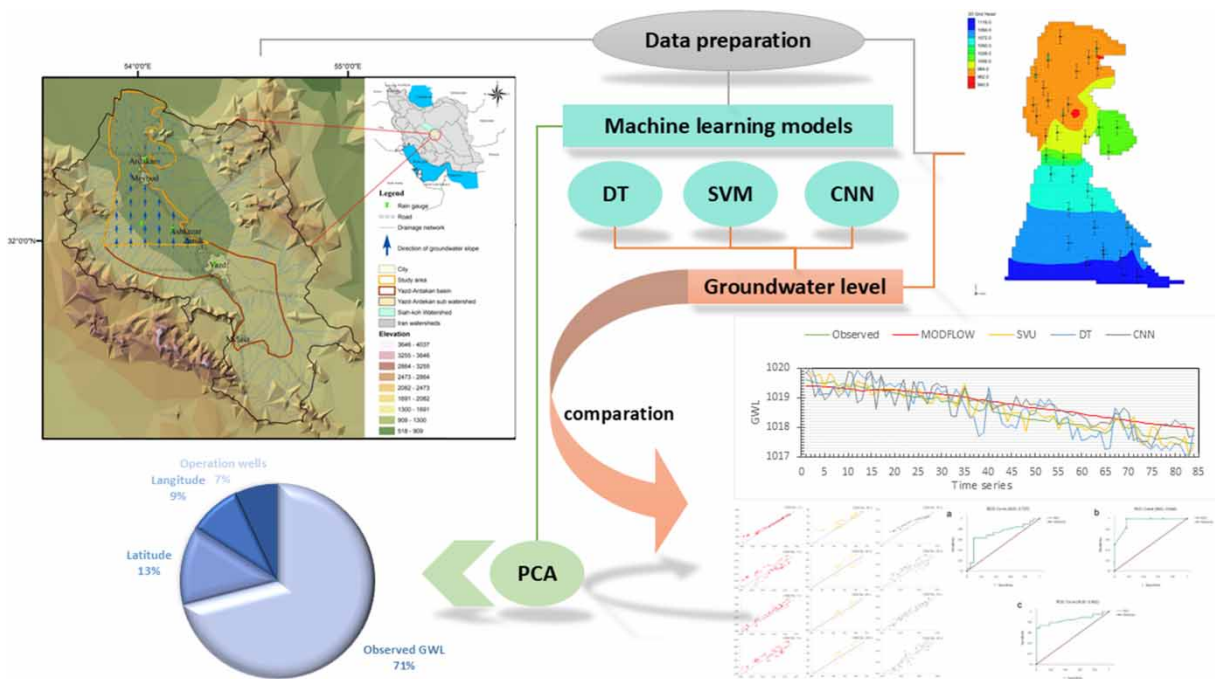
Reliable simulation of groundwater fluctuation is imperative for implementing sustainable groundwater management so that not only local people will be covered, but also environmental threats are going to be restrained in the future. This research aims to scrutinize the accuracy of numerical model (NM) and machine learning (ML) methods for groundwater level prediction (GWLP) for Yazd-Ardakan Plain from 2012 to 2019 in monthly steps. Additionally, principal component analysis (PCA) is applied to investigate the impact of each ML input feature on GWLP. The study area's aquifer data were analyzed and prepared to develop the conceptual model of MODFLOW and train ML algorithms for GWLP. Considering observation wells (OBWs), operation wells (OPWs), and their latitude and longitude as input features in convolutional neural networks (CNN), support vector machine (SVM), and decision tree (DT) algorithms, GWLP was performed. The results demonstrate that although MODFLOW considers the unique features of the aquifer, the most accurate GWLP was achieved by SVM, with root mean square errors (RMSE), correlation coefficient (r), and area under the receiver operating characteristics (ROC) curve (AUC) values of 0.12, 0.90, and 0.94, respectively. Furthermore, PCA presented that the observed groundwater level (GWL) was the most effective feature with 71%.

Key words: data-driven models, deep learning, GMS, groundwater management, numerical modeling, PCA

HIGHLIGHTS

- In this study, MODFLOW, a numerical model, is compared with machine learning methods for GWLP.
- PCA was employed in this study to assess the effectiveness of each input feature used in the algorithms and each one-year time period.
- This study also uses the important features of latitude and longitude, which significantly affect GWLP results when machine learning is applied.
- The results demonstrate the better performance of the MODFLOW model in terms of generalization ability (GeA) and the superiority of machine learning methods in terms of accuracy for GWLP.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Groundwater resources supply the majority proportion of the water in arid and semi-arid regions. In these areas, only 5–10% of the precipitation contributes to the recharging of the aquifer (Bouwer 2002). Due to climate change, the reduction in precipitation in these areas has made the natural recharging of the aquifer potentially insufficient (Sherif *et al.* 2022). Iran is a country with an arid and semi-arid climate, where the imbalance between natural recharge and groundwater extraction has caused a sharp drop in the groundwater level (GWL) (Gholami 2022). As a result of the indiscriminate use of these resources, groundwater level prediction (GWLP) is performed to manage and utilize them for drinking, industrial, and agricultural purposes (Rahman *et al.* 2020).

MODFLOW as a numerical model (NM) and a graphic simulator based on boundary physics (Tigabu *et al.* 2020), along with its flexibility in integrating with qualitative models (Lenin Sundar *et al.* 2022) and hydrological models (Yifru *et al.* 2022), has been an efficient tool over the last three decades for evaluating and solving many challenging problems of water resources management (Beyene *et al.* 2023). By comparing the effects of artificial recharge of aquifers using various techniques, such as infiltration basins (Chenini & Ben Mammou 2010) and injection wells (Bana Bafroei & Alimohammadi 2023), MODFLOW simulation can be utilized to determine which technique is most effective in advancing groundwater resources management objectives. MODFLOW simulator–optimizer models are employed to address groundwater challenges, such as reducing pumping costs from aquifers (Ayvaz & Elçi 2013), identifying pollution sources (Chakraborty & Prakash 2023), locating OPWs locations (Yin *et al.* 2020), addressing environmental concerns, and protecting the aquifer (Maghsoudi *et al.* 2023). MODFLOW simulator–optimizer models' results have proven to be highly beneficial for optimal aquifer management.

Hydrological problems across different scales have been predicted and solved in recent decades using artificial intelligence and machine learning (ML) methods. With numerous hidden layers, the convolutional neural network (CNN) is an advanced version of the artificial neural network and is classified as one of the deep learning algorithms that can be adapted for time series forecasting. This adaptability allows CNN to effectively stimulate patterns in GWL (Afzaal *et al.* 2020). CNN has been used less frequently than other techniques in GWLP (Afrifa *et al.* 2022). According to a study by Afzaal *et al.* (2020), CNN outperforms other algorithms both in terms of speed and precision. When utilizing large datasets for GWLP, CNN demonstrates superior performance (Wunsch *et al.* 2021), making them a valuable tool in scenarios where extensive historical data is available.

ML models for GWLP enhance accuracy (Amaranto *et al.* 2019), and one of the most frequently employed is support vector machine (SVM). Amaranto *et al.* (2018) applied five ML models to provide GWLP models for an aquifer located in Nebraska, USA, and SVM demonstrated accurate predictive abilities like other ML models, particularly in non-linear scenarios. Khan Mohammad & Coulibaly (2006) for Lake Erie in North America and Yoon *et al.* (2016) in South Korea evaluated GWL using several algorithms and compared their performances. In both studies, SVM demonstrated superior accuracy in GWLP. Additionally, Choi *et al.* (2020) predicted the water level of a vast wetland in South Korea using various techniques including SVM and decision tree (DT), and found that incorporating multiple features was both essential and highly effective. However, the DT is used less frequently in GWLP studies, and researchers are encouraged to investigate it further (Afrifa *et al.* 2022).

A few studies have compared NM and ML techniques for GWLP in order to discover which is the most effective. These comparisons used groundwater unit hydrograph (Mohammed *et al.* 2023), precipitation (Amiri *et al.* 2023), OPWs (Almuhaylan *et al.* 2020), GWL (Malekzadeh *et al.* 2019), recharging and evaporation rates (Moghaddam *et al.* 2019), temperature, and surface flow (Mohanty *et al.* 2013). Furthermore, in all the mentioned studies, ML was found to be a superior method for GWLP. However, Ebrahimi *et al.* (2023) noticed that although both models are robust, MODFLOW is the better model. Chen *et al.* (2020) also conducted another comparison between MODFLOW and ML using RMSE, R^2 , and generalization ability (GeA) indicators, and the results highlighted MODFLOW's strength in GeA and ML's superiority in prediction accuracy. In another comparison, Almuhaylan *et al.* (2020) emphasized the need to compare and utilize deep learning or CNN techniques. Interestingly though, PCA has not been applied in any of the comparisons mentioned in the literature. PCA is commonly applied in hydrology (Cai *et al.* 2021), frequently used to decrease data dimensionality and improve the accuracy and efficiency of GWLP models by extracting the key characteristics from datasets when paired with ML models for GWLP by extracting core features that affect prediction accuracy in hydrologic datasets (Seifi *et al.* 2020).

Very little GWLP using NM and ML has been conducted in the study area. The area, which experiences extreme water stress and scarcity, is one of Iran's highest-density industrial and agricultural regions. The purpose of this research is to compare the GWLP methods by employing MODFLOW, SVM, CNN, and DT algorithms. Furthermore, PCA is utilized to determine the importance of each input feature used in algorithms and each 1-year time period. This study also utilizes the impressive features of latitude and longitude, which have a substantial effect on GWLP results when using ML.

2. METHODOLOGY

2.1. Study area

The Yazd-Ardakan Plain is one of the 11 sub-basins of the Siah-Koh-Zarin watershed. It is categorized as one of Iran's arid regions, with an average annual evaporation rate of 2,400 mm, an average rainfall of 137 mm, and a mean temperature of 18.9 °C, with an average relative humidity of 35.3% (Goodarzi *et al.* 2022). Lying between latitudes 32° and 33° North and longitudes 52° and 53° East, this region covers 11,630 km² (Ministry of Energy 2012). The Yazd-Ardakan Plain is located in the northern part of the Yazd Province. It extends southward to Shirkoh, which is the plain's highest point at 4,037 m above sea level, and is enclosed by the Siah-Koh desert to the north (Goodarzi *et al.* 2023). According to estimates, 617 million m³ of water are extracted annually from these resources (Ministry of Energy 2012). The northern part of the Yazd-Ardakan Plain, which covers 1,625 km², was chosen as a study area due to the increased accuracy of NM (Figure 1). The groundwater and land slopes are oriented from the southern plain toward the north. There are no rivers, lakes, or any reliable water resources and the region's only source of water supply is its groundwater resources which have been intensively exploited through unauthorized OPWs recently, resulting in significant water shortages since the study area is one of the densest industrial areas in Iran and also has a high level of agricultural activity (Banadkooki *et al.* 2022). So, forecasting the GWL and implementing management strategies are vital to preventing further depletion of the aquifer.

The study area's aquifer is an unconfined one, with 36 OBWs and 405 OPWs (Figure 2) that collectively extract 463,708 m³ of water from it per day and without lakes and rivers (Ministry of Energy 2012). Unconfined aquifers are those that have reasonably high intrinsic permeability from the earth's surface to the bedrock. As a result, water storage continues from the bottom to the top, and these layers may receive recharge from precipitation, snowmelt, or other sources. Although the area's surface digital elevation model (DEM) was available, bedrock information for the aquifer was not. Thus, the DEM of the aquifer's bedrock was developed utilizing the area's few geological logs analyses and the depth of the OPWs. The values of hydraulic conductivity play an important role in transmitting water through the aquifer (Goodarzi *et al.* 2024c) and were calculated by using the aquifer's transmissivity reporting from the Ministry of Energy and the thickness of the

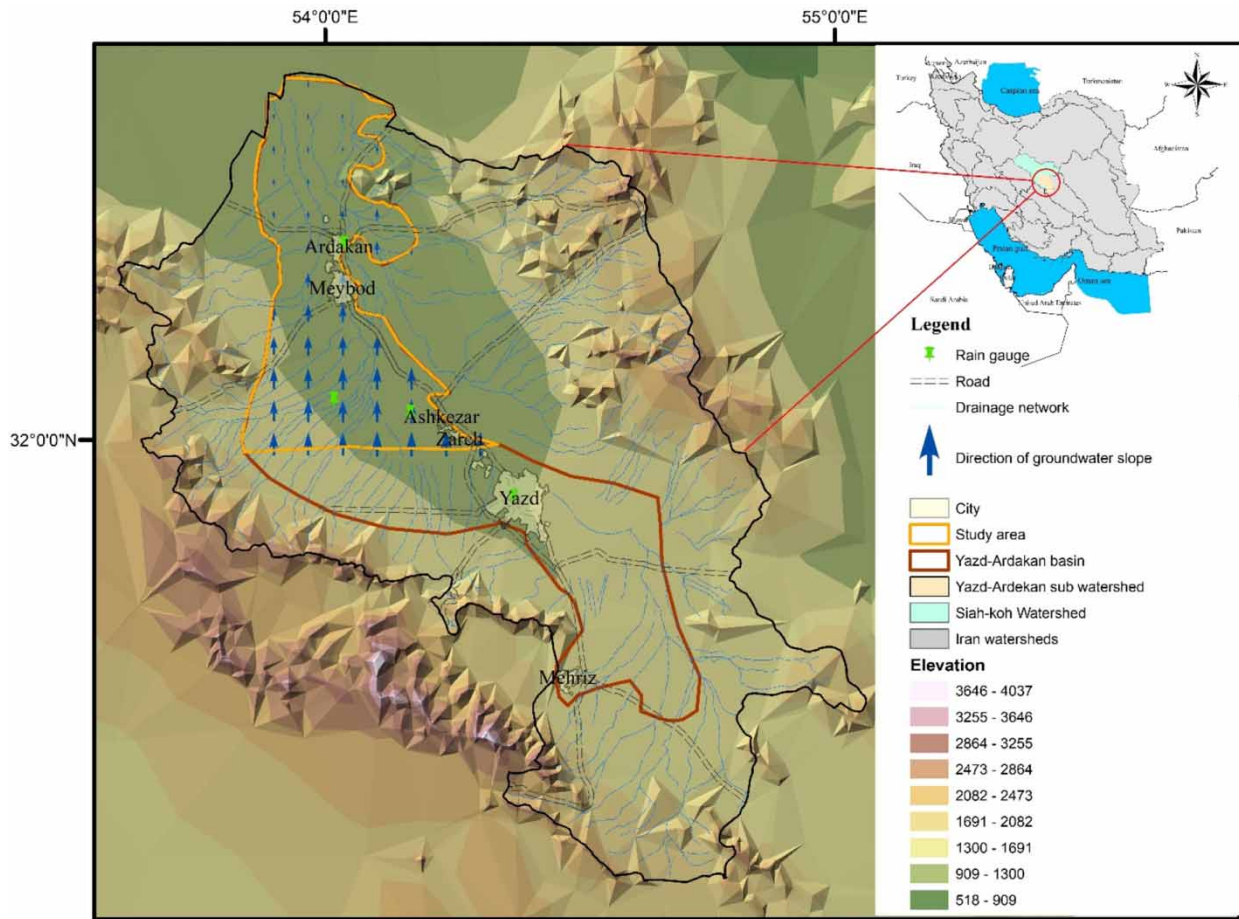


Figure 1 | A comprehensive map of the study area.

aquifer to introduce the initial values of hydraulic conductivity utilized by MODFLOW Equation (1):

$$T = Kb \quad (1)$$

where T is transmissivity (m^2/day), K is hydraulic conductivity m^2/day , and b is the aquifer thickness (m). The initial recharging values of the aquifer were determined based on the amount of rainfall that infiltrated the aquifer, the return of OPWs, and sewage wells. For application in the NM, all data were prepared employing GIS software.

2.2. Numerical modeling

MODFLOW, a 3D finite difference groundwater flow model (Harbaugh *et al.* 2000), is visualized through groundwater modeling system (GMS) graphics software and is used to simulate groundwater flow in the study area. Modeling the flow of groundwater in heterogeneous aquifer pores and under transient conditions, MODFLOW converts the physics concept of the issue into mathematical form. This movement is explained by the partial differential Equation (2):

$$\frac{\partial}{\partial x} \left(K_x \left(\frac{\partial h}{\partial x} \right) \right) + \frac{\partial}{\partial y} \left(K_y \left(\frac{\partial h}{\partial y} \right) \right) + \frac{\partial}{\partial z} \left(K_z \left(\frac{\partial h}{\partial z} \right) \right) = S_s \left(\frac{\partial h}{\partial t} \right) \pm W \quad (2)$$

where the hydraulic conductivity values along the coordinate axes of x , y , and z are indicated by K_x , K_y , and K_z (L/T). The variables h , W , t , and S_s represent the groundwater head (L), sources or sink volume (T^{-1}), time (T), and the aquifer's specific storage (L^{-1}).

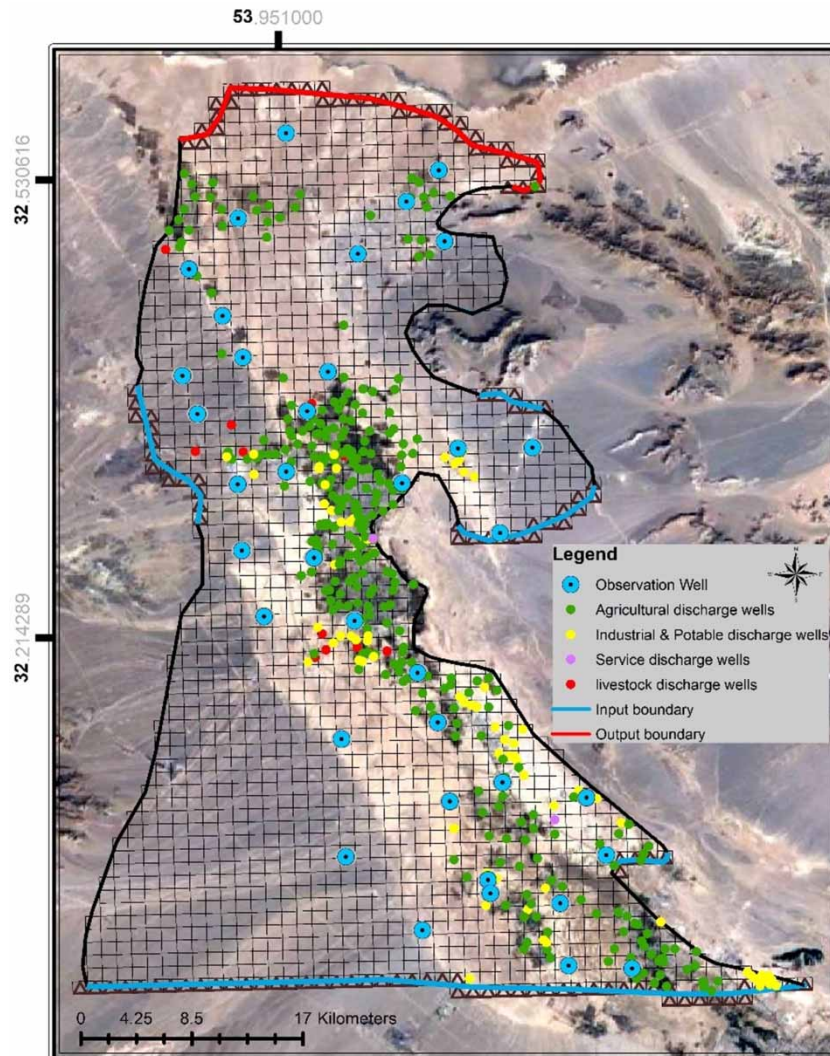


Figure 2 | An illustration of the aquifer's input and output boundaries, the model grid, the OBWs, and the OPWs.

MODFLOW has the ability to simulate confined and unconfined aquifers, transient and steady flows, rivers, drains, wells, springs, evaporation, and infiltration by rainfall and irrigation. In order to identify and comprehend the mathematical model of the initial conditions, parameters, and physics of the surrounding environment and to simplify it, the first step in groundwater modeling is the creation of a conceptual model. The hydrologist's knowledge of hydrology, groundwater flow dynamics, and hydrogeological conditions is essential for developing a conceptual model. Aquifer data, including field data, OPWs, the initial distribution of hydrological parameters, the DEM of the aquifer's bedrock and surface, and recharging, were utilized and implemented in GMS to develop a conceptual model. The solitary outflow boundary is found in the aquifer's northern region, and the majority of its borders are inflow boundaries (Ministry of Energy 2012). The model's grid was created with dimensions of $1,000 \times 1,000$ m. As shown in Figure 2, the model contains 1,725 active cells and 2,271 inactive cells. Using the parameter estimation (PEST) program, the model was calibrated due to GWL calculation, by taking into account the aquifer's hydraulic conductivities and calculating the difference between the recharging and discharge factors. Through the OBWs, the estimated and actual GWL of the aquifer are compared. Notably, the parameters introduced to PEST were calibrated within a suitable range that matched the aquifer's attributes. The model incorporates an interval representing the permitted range of fluctuation between the computed and observed GWLs; if that difference surpasses the established limit, a new calibration occurs. The interval that has been introduced to the model is 1 m. Following calibration stages, values of calibrated parameters were employed to validate the model in order to gain more assurance that the model will

be able to simulate GWL by applying calculated parameters in the calibration steps and different time periods, groundwater withdrawal, and groundwater recharge of calibration stage in acceptable RMSE.

2.3. ML models

For the ML simulation, the input features, including OBWs, OPWs, and the latitude and longitude of these parameters, were used for GWLP with CNN, SVM, and DT algorithms. Python 3.2 and K-fold cross-validation (with 5 folds) were applied during the simulation. To ensure reliable evaluation and minimize bias, the dataset was split into two subsets: 80% for training and the remaining 20% were utilized for testing the model. Data preprocessing, including normalization of all input variables, was performed to transform the data into a format that is more easily and effectively processed. Additionally, PCA was applied to reduce the dimensionality of the data and improve model efficiency and accuracy.

2.3.1. Convolutional neural network

[Lecun et al. \(1998\)](#) introduced CNN, a deep learning algorithm that is similar to the connections among neurons in the human brain and nearly mimics human vision. Unlike traditional ML algorithms, which require manual preprocessing for feature extraction, CNN automatically performs this task, reducing the need for such preprocessing. However, like other deep learning algorithms, CNN needs an enormous volume of data as well as computational power. Upon receiving input data, CNN assigns weights to different elements, distinguishing them based on their importance and regional characteristics ([Xu et al. 2020](#)).

The convolution layer, pooling, and fully connected are the three components of CNN. The CNN multi-layer structure is part of the convolution layer. The convolution layer can be expressed mathematically as the product of two matrices that have been multiplied correspondingly. In CNNs, each of the filtered pictures goes through a pooling step after the convolution layer. Smaller and more superfluous parts of information vanish during this process, which accelerates and improves the accuracy of learning the data processing algorithm by passing simpler and more useful information to the following layers. Every neuron in the fully connected layer is linked to every other neuron in the pooling layer, but not to any other neurons within the layer ([Xu et al. 2020](#)). This layer boosts the prediction accuracy by interpreting and analyzing the data.

CNNs are effective at capturing spatial and temporal patterns in GWL ([Wunsch et al. 2021](#)). The model learns hierarchical features from the input data, including OBWs, OPWs, and geographic coordinates, without the need for manual feature extraction. By treating latitude and longitude as spatial variables, CNNs can capture the spatial dependencies between OBWs and GWL. This makes CNN particularly suitable for modeling complex regional and temporal variations in GWL based on the input features.

2.3.2. Support vector machine

SVM, introduced by [Vapnik \(1995\)](#), is a powerful and precise ML algorithm used for both regression and classification ([Goodarzi et al. 2024a](#)). SVM is widely utilized because of its accurate detection abilities in a variety of fields. The primary function of SVM is to define a hyperplane in a multidimensional space that separates data points with maximum margin, fitting the regression to minimize errors in the predicted output ([Li et al. 2016](#)). The hyperplane's dimensions are determined by the number of features in the data ([Goodarzi et al. 2024b](#)). As the number of features increases, visualizing the hyperplane becomes more complex. SVM uses various kernel functions to search for optimal hyperplanes in high-dimensional spaces. Its ability to handle regression tasks with fewer features, combined with its complex structure and advanced techniques, makes it particularly suitable for estimating GWL.

In an SVM classifier, we try to find a line in the shape of $w \cdot x - b = 0$ so that the data of two or more classes can be properly separated from each other. The test dataset D contains n points, as Equation (3) follows ([Goodarzi 2024](#)):

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, +1\}\}_i^n \quad (3)$$

where x_i is observed, y_i is predicted, and y takes one of the -1 or 1 values. The goal is to find the hyperplane that separates data with the greatest distance from the edge points that separate the points $y = -1$ as well as $y = 1$. So, the hyperplane can be

written in the following form:

$$w \cdot x - b = 1$$

$$w \cdot x - b = -1$$

2.3.3. Decision tree

The DT is one of the best ML algorithms for regression and classification based on predetermined answers (Quinlan 1986). It operates based on a set of rules that guide the decision-making process. The root of the tree represents the initial decision point, and the branches represent possible outcomes from this decision. The process continues until it reaches the leaf nodes, which represent the final predictions. The complexity of a DT is directly related to its interpretability, and it can be controlled through pruning, which reduces the number of nodes, leaves, and attributes involved in the tree.

For GWLP, DT effectively models non-linear relationships by capturing the spatial and temporal dependencies between GWL and features such as OBWs, OPWs, and geographic coordinates. DT's ability to handle both categorical and continuous data allows it to model diverse groundwater conditions effectively.

2.3.4. Principal component analysis

High-dimensional feature spaces complicate models, increase computation times, and make interpretation more difficult. PCA, proposed by Pearson (1901), is a dimensionality reduction technique that addresses these issues. PCA has the ability to generate a plan to minimize the number of features while maintaining the highest level of information by determining the correlation and patterns in the features as well as the maximum variance among them (Zhang *et al.* 2021). PCA enables the model to employ features that have the greatest impact on its result rather than analyzing every feature.

The input features of the algorithms are the pumping values of aquifer OPWs, OBWs, and their longitude and latitude. In this study, PCA has been used to assess the impact and effect of each ML algorithm input feature over an interval of 1 year. The model's accuracy will be preserved, and computation times will also get quicker.

2.4. Model's performance criteria

To evaluate each model and compare it to others, criteria for assessing the models' performance are essential. The models employed in this study were assessed using root mean square error (RMSE), correlation coefficient (R^2), receiver operating characteristic (ROC), curve analysis, and area under curve (AUC). Regarding RMSE, it is one of the most popular and reliable measures for assessing model accuracy and error, as illustrated by Equation (4). The error values reach a power of two before averaging in RMSE, making it impossible to neutralize the impact of the difference between positive and negative averages when calculating error. Equation (5) presents R^2 as a measure of the linear relationship between the independent and dependent variables.

$$\text{RMSE} = \left(\frac{1}{N} \sum_{i=1}^N (h_c - h_o)^2 \right)^{0.5} \quad (4)$$

$$R^2 = \frac{\left[\sum_{i=1}^N (h_o - \bar{h}_o)(h_c - \bar{h}_c) \right]^2}{\sum_{i=1}^N (h_c - \bar{h}_c)^2 \sum_{i=1}^N (h_o - \bar{h}_o)^2} \quad (5)$$

where h_o is the observed water level, h_c is equal to the water level calculated by the model, and N represents the number of observations.

A crucial consideration when developing data-driven models (e.g., SVM and ANN) is GeA, as Equation (6) follows (Yoon *et al.* 2011):

$$\text{GeA} = \frac{\text{testing stage RMSE}}{\text{training stage RMSE}} \quad (6)$$

A GeA value of 1 represents that the model performs the simulation optimally.

ROC is one of the most powerful statistical and graphical techniques for calculating how accurate predictions and data-driven models are. The ROC curve displays false positive values on the Y-axis and true positive values on the X-axis for different thresholds (Chen *et al.* 2022). The sensitivity of the model increases as the false positive value rises. The AUC indicates the model's reliability and ability for prediction. AUC values between 0.9 and 1 are outstanding, 0.8 and 0.9 are very good, 0.7 and 0.8 are good, 0.6 and 0.7 are satisfactory, and 0.6 and 0.5 are acceptable since the more accurate model has a greater AUC. ROC-AUC provides valuable insights in regression-based GWLP by evaluating the models' ability to discriminate between GWL thresholds. By utilizing ROC-AUC, we analyze the discriminatory ability and overall reliability of the models. This approach has been applied to a variety of groundwater-related problems, including risk assessment of nitrate groundwater contamination (Liang *et al.* 2024), groundwater potentiality modeling (Mallick *et al.* 2022), and groundwater spring potential (Seifu *et al.* 2023).

3. RESULT

MODFLOW as an NM was utilized to simulate the GWL by collecting and categorizing the data, and developing the conceptual model for introducing data to the finite difference network. The first stage consisted of applying hydraulic conductivity and recharge parameters to calibrate a steady state based on 36 OBWs for August 2012. The calibration of the steady-state resulted in an RMSE value of 0.27. Figure 3 illustrates the GWL calculated by MODFLOW for August 2012.

Following that, hydraulic conductivity anisotropy, horizontal hydraulic conductivity, and specific yield were introduced at the pilot points for 2012–2018 as the next phase in the calibration of the transient flow. Afterward, the model was validated in 2018–2019. As shown in Figure 4, the RMSE values for the transient and validation periods were 0.82 and 0.84, respectively. The R^2 values for the transient and validation periods were 0.76 and 0.68, respectively (Figure 5).

Turning to the results of the ML models, monthly GWLP for the years 2012–2019 was performed. For the test period, the RMSE values for the CNN, SVM, and DT algorithms were calculated to be 0.26, 0.12, and 0.17, respectively. For the training period, these values were determined to be 0.25, 0.12, and 0.16 (Figure 4). The R^2 values for the CNN, SVM, and DT algorithms were 0.64, 0.90, and 0.80 during the test period, and 0.66, 0.91, and 0.81 for the training period (Figure 5).

For a clearer comparison, Figure 6 shows the RMSE values of the test/validation period, and Figure 7 presents the R^2 values for the train/calibration simulation models for each of the 36 OBWs individually for the MODFLOW and CNN, SVM, and DT algorithms. The SVM algorithm simulation, which has the lowest RMSE, produces head differences computed by the model with values less than 0.2 for all OBSs. Additionally, SVM achieves the highest R^2 value and performs properly for all OBSs. Based on the RMSE values obtained from the NM and ML algorithms, the GeA value for MODFLOW is 1.02, while the values for the CNN, SVM, and DT algorithms are 1.06, 1.03, and 1.04, respectively. The values indicate that all ML models have a GeA slightly greater than one, demonstrating effective performance. Despite the excellence of the ML results, the MODFLOW model is shown to be the superior model in evaluating GeA.

In addition, since there are many OBWs, three of them – numbers 7, 94, and 89 – located in the north, center, and south of the modeling area, respectively, were selected to compare the observed GWL with the GWL calculated by the ML models (Figures 8 and 9). The SVM model provides the most accurate GWL calculation, with RMSE values of 0.08 and 0.09 for OBW numbers 7 and 89, respectively. For OBW number 94, the DT and SVM algorithms, with an RMSE value of 0.16, outperformed the other models. However, for all three OBWs, the MODFLOW and CNN algorithms also had RMSE values less than 0.3 and 0.6, respectively, indicating that both models are reasonably accurate in predicting GWL. Notably, the R^2 values for the MODFLOW, CNN, SVM, and DT models are as follows: for OBW number 7, they are 0.99, 0.69, 0.95, and 0.84, respectively; for OBW number 89, they are 0.87, 0.69, 0.88, and 0.81; and for OBW number 94, they are 0.91, 0.69, 0.91, and 0.82, respectively. The results demonstrate an excellent correlation between observed and computed GWL data. Figure 8 illustrates a graph showing the observed GWL of OBWs numbers 7, 94, and 89 along with the calculated GWL. As can be seen in Figure 8, the observed GWL trends are smooth because real groundwater changes gradually due to aquifer storage and properties. MODFLOW, on the one hand, tends to capture smoother trends because it's based on physical groundwater flow equations and hydrological properties. On the other hand, ML models may show abrupt fluctuations. This difference occurs because these models, despite their predictive power, can react sensitively to smaller data changes rather than reflecting the natural, steady pace of groundwater fluctuation, which can result in a less stable prediction pattern compared to real observations.

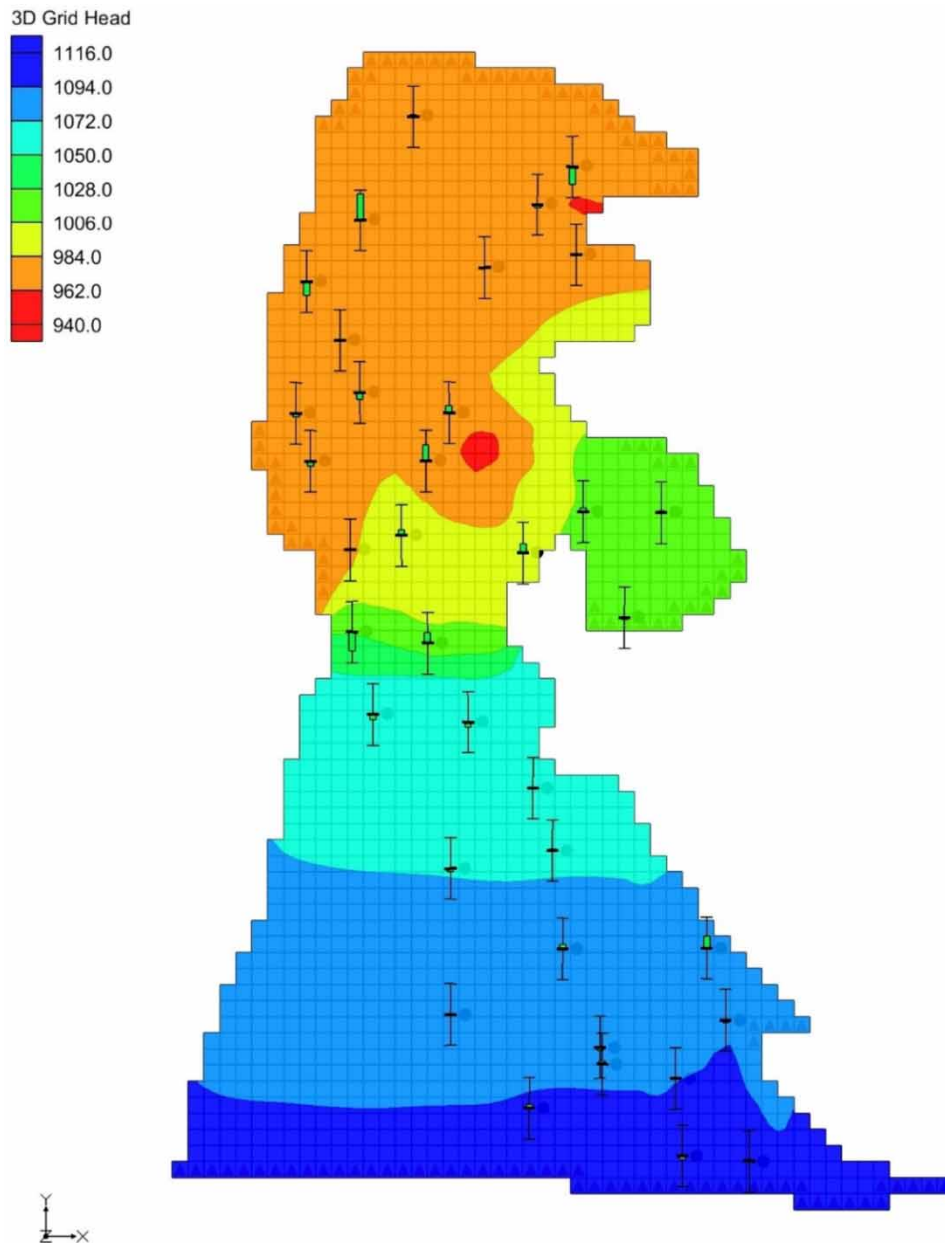


Figure 3 | The model's calculation of the GWL for August 2012.

The performance of the algorithms during the training period was further investigated using the AUC values. A higher AUC value indicates better model performance GWLP, approaching a value of one. The CNN, SVM, and DT algorithms have AUCs of 0.73, 0.94, and 0.73, respectively (Figure 10). With an AUC of 0.94, the SVM algorithm performs best in GWLP, indicating its robust capability to model non-linear relationships effectively and superior ability distinguishing between GWL thresholds. Following it is the DT algorithm, with an AUC of 0.85, also demonstrating effective handling of non-linear relationships despite being slightly less precise than SVM, while the CNN algorithm, with an AUC of 0.73, may reflect a less-than-ideal alignment of its architecture with the dataset's features. However, it is also effective but less so than the SVM and DT algorithms.

The RMSE values indicate that the GWLP performed perfectly despite the uncertainty regarding the input parameters of the ML models and MODFLOW. One reason for the accuracy of ML predictions is the inclusion of the latitude and longitude

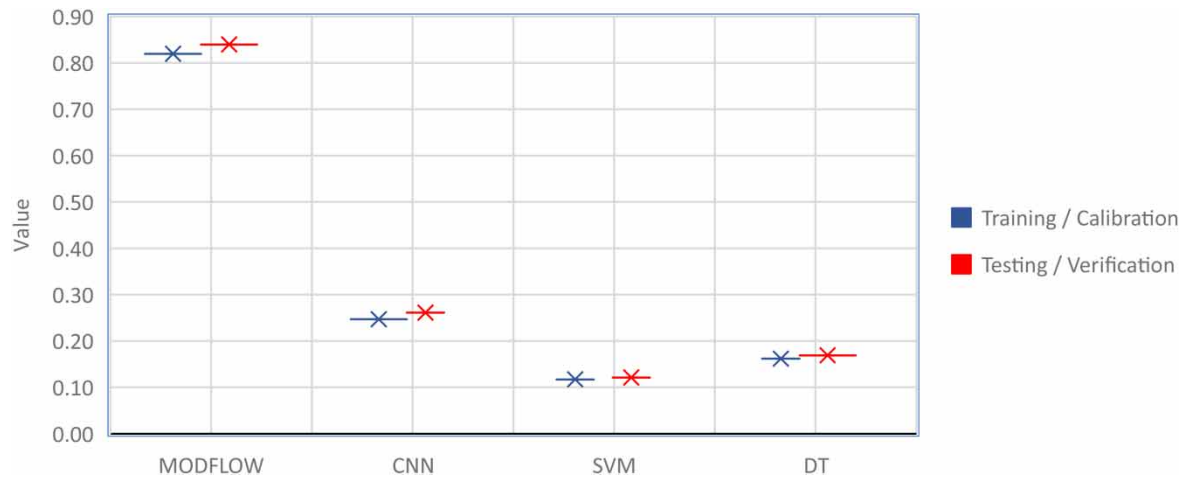


Figure 4 | RMSE values for the GWLP models during the training/calibration and testing/verification periods.

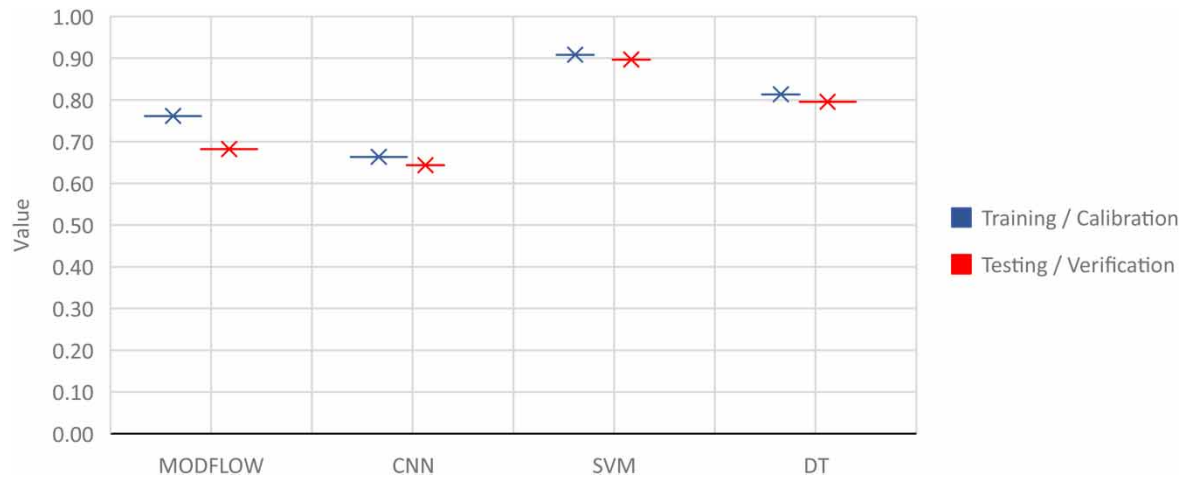


Figure 5 | R^2 values for the GWLP models during the training/calibration and testing/verification periods.

as input features. Table 1 presents the MODFLOW, CNN, SVM, and DT performance in GWLP using the criteria from 36 OBWs. The SVM algorithm provided highly accurate predictions. The RMSE, R^2 , and AUC values for the SVM algorithm were 0.12, 0.91, and 0.94, respectively. For 11 OBWs out of 36, the computed RMSE value was less than 0.099, and for 18 OBWs, it was less than 0.99. Moreover, for 24 OBWs, the R^2 value between the observed GWL using the SVM model was greater than 0.9; for 11 OBWs, it was greater than 0.87; and for one OBW, it was 0.81. The DT algorithm, the next most accurate model after SVM in this study, has been utilized in only a small percentage of GWLP studies. The RMSE, R^2 , and AUC values for the DT algorithm were 0.16, 0.81, and 0.85, respectively. Additionally, the CNN algorithm achieved good results, with an RMSE of 0.25, R^2 of 0.66, and an AUC of 0.73, demonstrating its effectiveness in predicting GWL within the study area.

PCA, which is an ML method for data analysis, was used in this study to reduce the dimensionality of the input features and select the most effective ones for simulating the GWL. In contrast to physics-based NM, ML does not require the use of all features for the applicable algorithms to model data accurately. When PCA was applied, the observed GWL feature accounted for 71% of the influence on the simulation. Moreover, the latitude and longitude features collectively contributed 22% to the simulation's accuracy (Figure 11). Finally, the OPWs feature contributed 7%. The fifth time period was found to be the most effective for the CNN and SVM models, whereas the sixth time period was the most effective for the DT model.

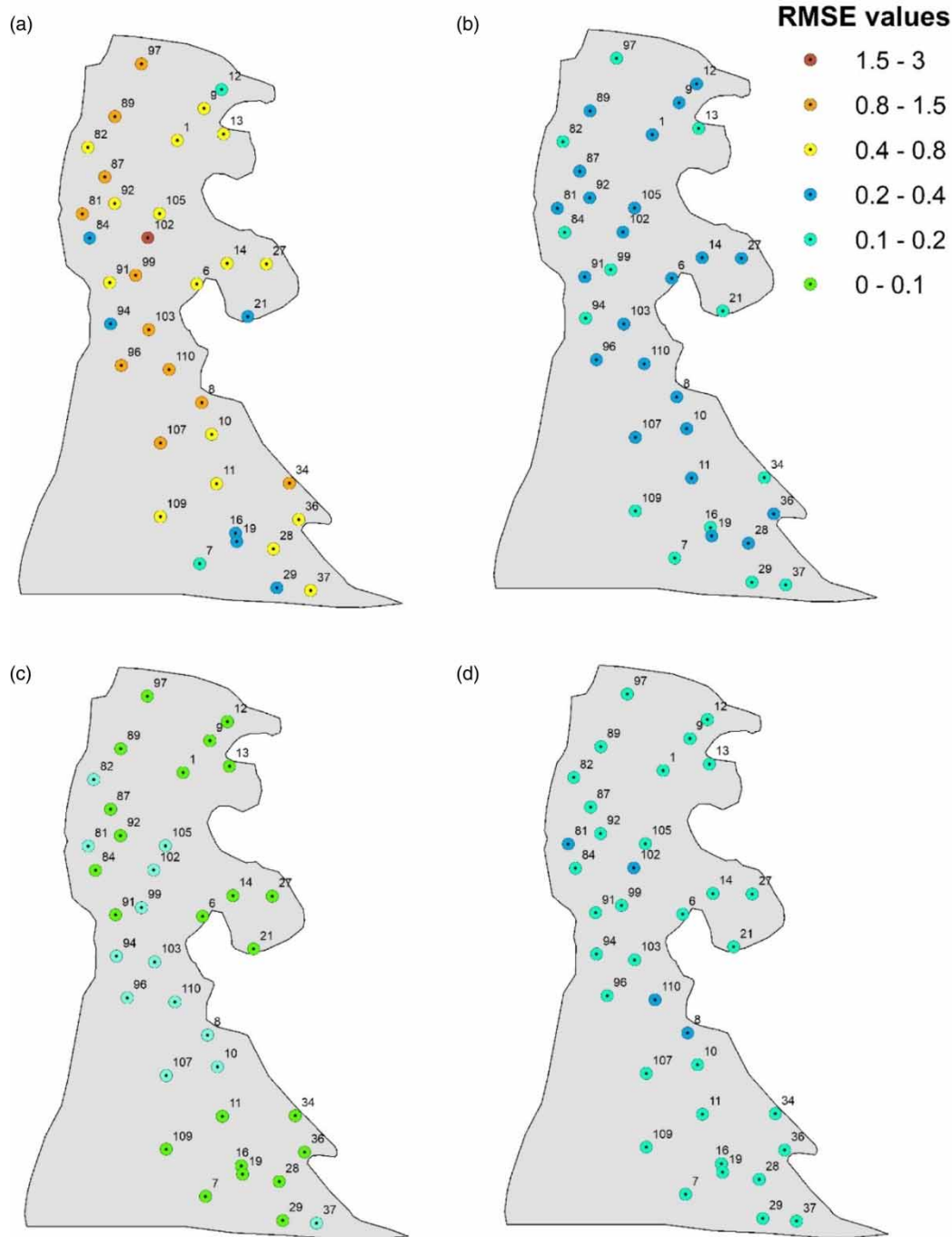


Figure 6 | RMSE maps comparing observed and predicted GWL for the test/validation period for (a) MODFLOW, (b) CNN, (c) SVM, and (d) DT.

4. DISCUSSION

Numerous studies have emphasized the importance of using NM and ML algorithms for GWLP. CNN, a deep learning algorithm; DT, a tree-based algorithm; SVM, an ML algorithm; and MODFLOW, an NM, were employed in this study for GWLP. 80% of the data from the study area were applied for the training, and 20% were used for testing. The results demonstrate the superiority of ML techniques over the NM by performing statistical indicators for comparison methods. Among ML methods, SVM exhibited the best performance in GWLP according to the model's RMSE and R^2 , aligning with [Chen et al. \(2020\)](#).

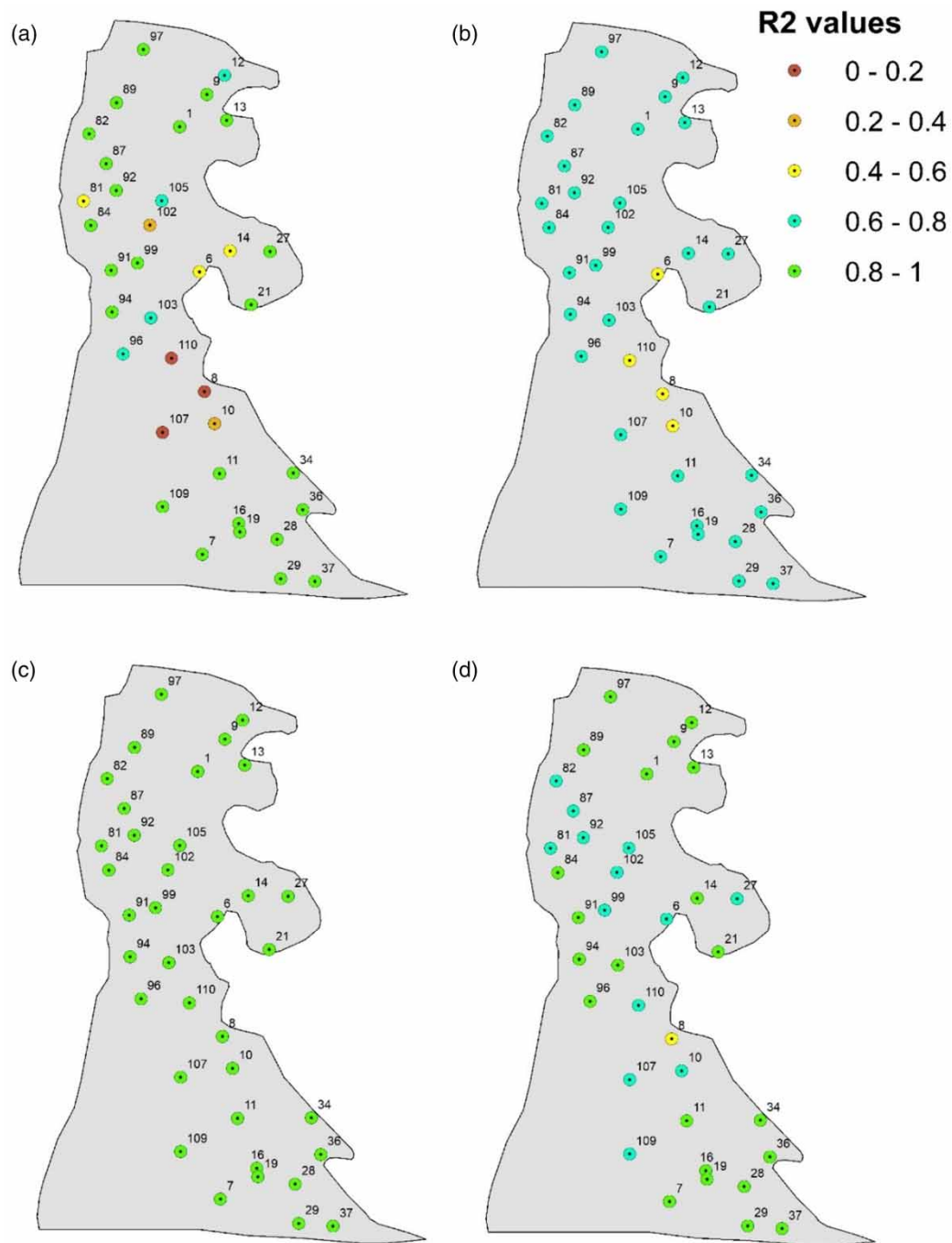


Figure 7 | R^2 maps comparing observed and predicted GWL for the train/calibration period for (a) MODFLOW, (b) CNN, (c) SVM, and (d) DT.

Additionally, in this study, AUC values and the ROC curve were also utilized to further assess and compare ML methods. The highest AUC score shows that SVM performs remarkably well in distinguishing between different GWL classes, proving its capability to handle non-linear relationships effectively. Furthermore, the DT capability to handle non-linear relationships was effective, though its value was slightly less than SVM. Nevertheless, the relatively lower AUC score indicates that CNN struggles with the task compared to SVM and DT. In contrast to CNN's powerful abilities for image and spatial data, their performance might be limited by the structure of the groundwater data as they are numerical, or time series in

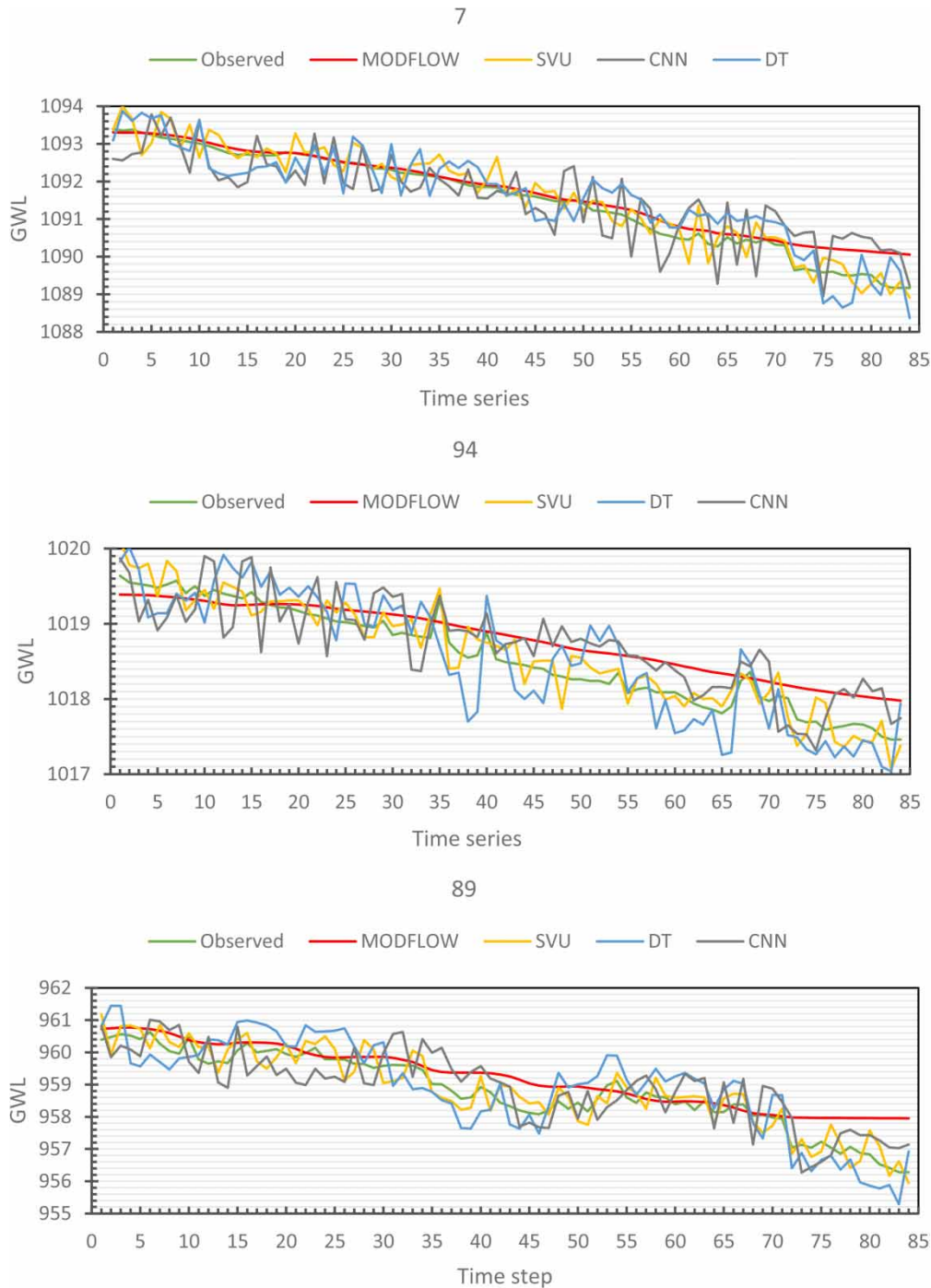


Figure 8 | Comparison of time series observed GWL (m) and obtained GWL (m) from ML models and MODFLOW in verification or testing period for OBW numbers 7, 94, and 89.

nature, which may not present the kind of spatial or hierarchical patterns that CNNs are designed to exploit. However, in this study, like the results of a study by Wunsch *et al.* (2021), CNN achieved notable accuracy for GWLP. Especially, when it trained on rich input features such as latitude and longitude, which our PCA analysis identified as effective features because CNNs significantly benefit from extensive datasets.

MODFLOW outperformed ML methods in the GeA index, similar to the findings of Chen *et al.* (2020). The fact is that MODFLOW considers the unique properties of the aquifer, making it more robust when the groundwater basin experiences

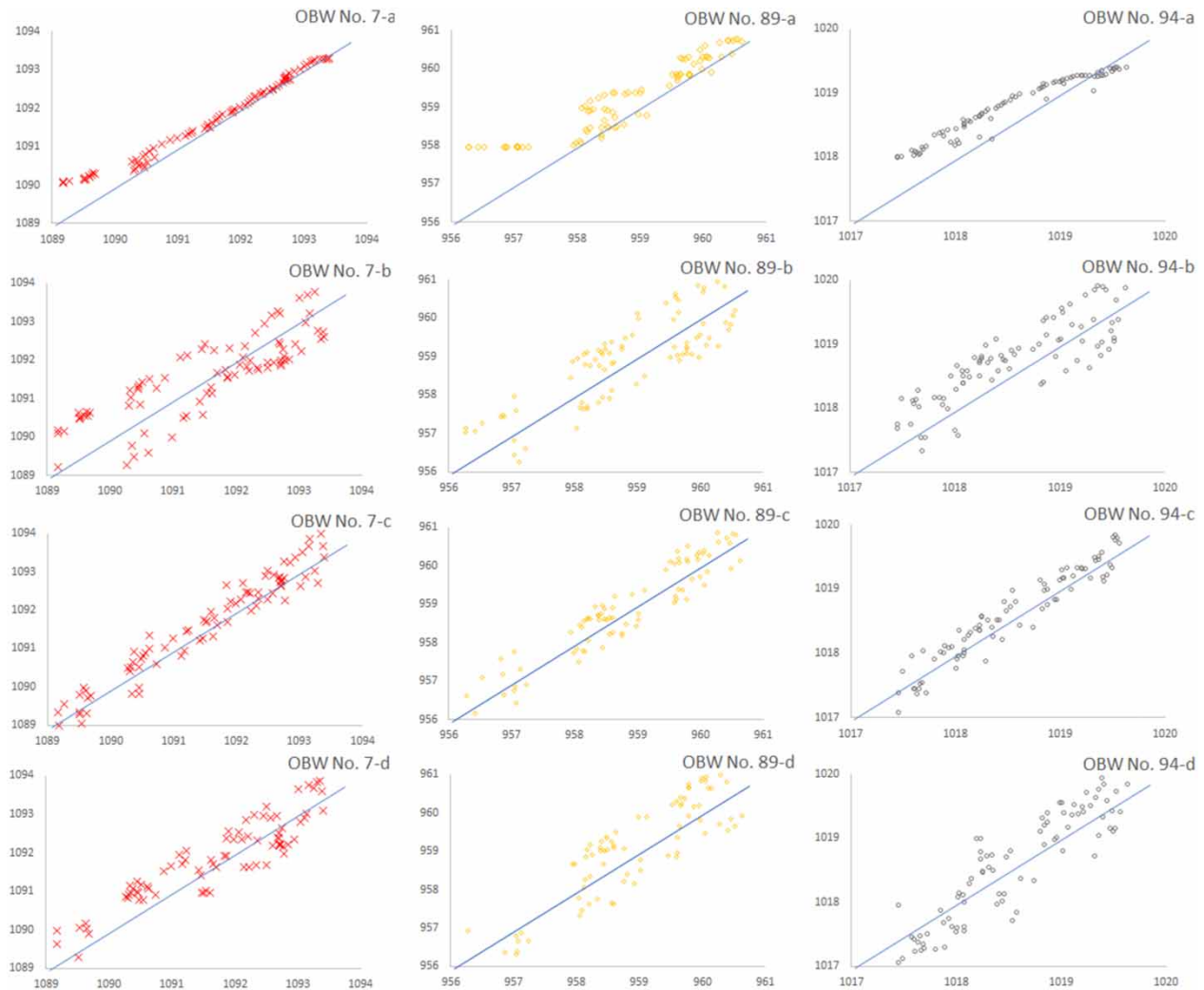


Figure 9 | Correlation of the simulated GWL (m) and observed GWL (m) during the verification/testing period for OBW numbers 7, 94, and 89 for (a) MODFLOW, (b) CNN, (c) SVM, and (d) DT.

changes in recharge due to drought, climate change, or stress from pumping at levels different from the study period, leading to superior performance in scenarios requiring comprehensive flow simulations. In contrast, ML algorithms do not perform well under such altered conditions. However, one of the challenges of MODFLOW is the extensive data preparation required; it can take months to complete a modeling project. Additionally, understanding groundwater flow and hydrogeological sciences is necessary for successful MODFLOW modeling. Unlike ML techniques, MODFLOW simulates groundwater flow by balancing the flow budget, offering a more comprehensive approach.

Turning to ML, users can simulate groundwater with fewer data requirements, and preparing the necessary GWLP data using ML is easier than collecting MODFLOW data (Mohanty *et al.* 2013). On the one hand, ML approaches, particularly when trained with robust datasets, offer flexibility, reduced computational demands, and faster model preparation. On the other hand, ML models can be extremely helpful in cases of aquifers with limited data. Notably, PCA revealed that latitude and longitude contributed significantly, with a combined effectiveness of 22% in improving model predictions, along with the GWL having the greatest effectiveness at 71%. The findings from PCA also imply that perfect feature selection can fill in the gaps between various modeling techniques.

Consequently, by applying management scenarios over time, both NM and ML models are able to predict GWL changes, providing decision-makers with tools to analyze groundwater flow under conditions of drought, climate change, and water stress due to aquifer over-exploitation.

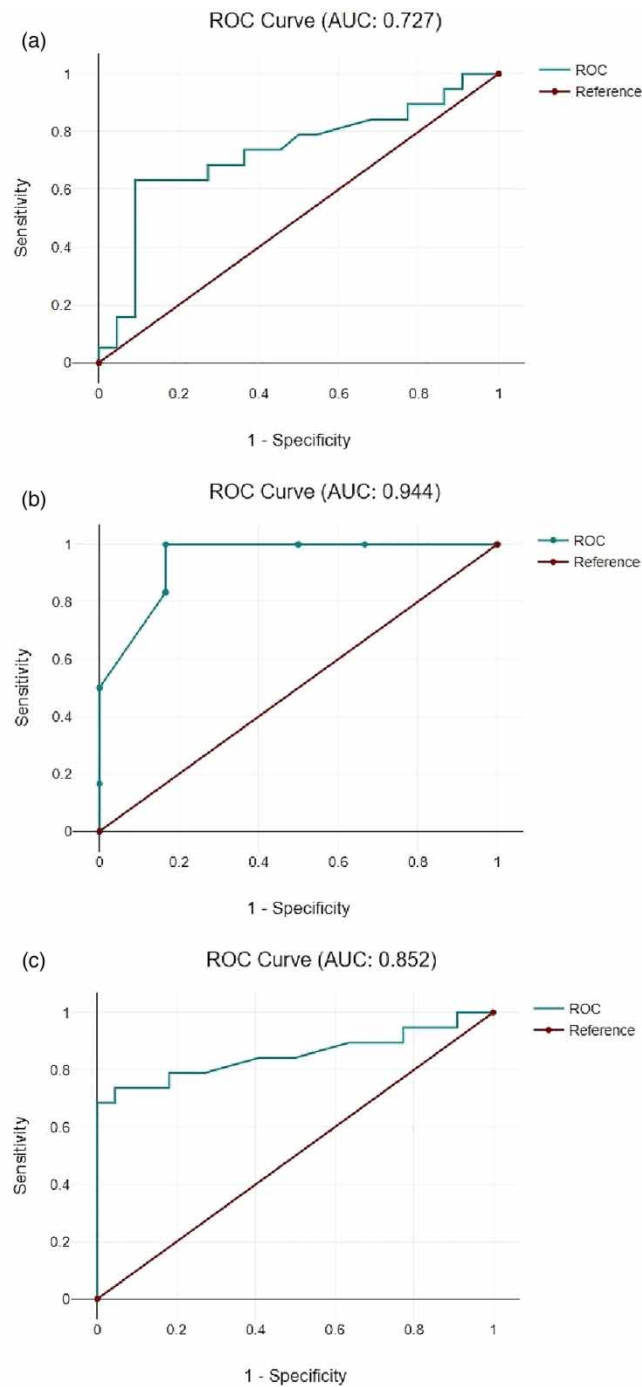


Figure 10 | ROC curve and AUC values of the models in the training steps for (a) CNN, (b) SVM, and (c) DT.

Table 1 | Comparison of performance criteria for numerical and ML models in the validation/testing stage

Criteria	MODFLOW	CNN	SVM	DT
RMSE	0.84	0.26	0.12	0.17
R^2	0.68	0.64	0.90	0.80
GeA	1.02	1.06	1.04	1.04

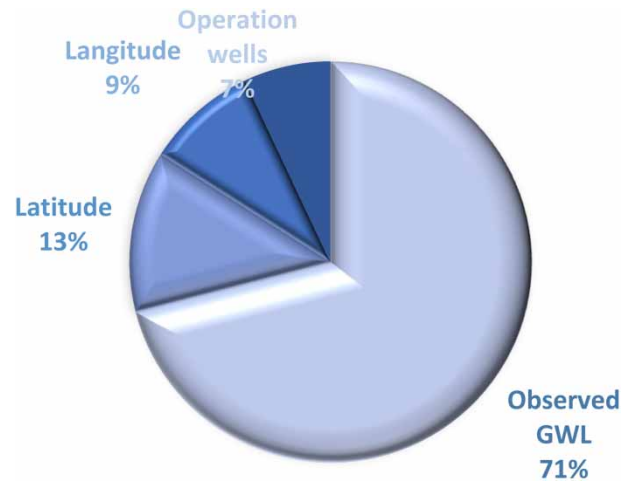


Figure 11 | The effect of each parameter and its percentage impact on the accuracy of the GWLP by the ML models.

5. CONCLUSION

The study area is located in the Yazd-Arkan Plain, one of Iran's most industrialized regions, which has recently experienced a dramatic decline in GWL. Consequently, the implementation of management programs is required to prevent further damage and use groundwater simulators to predict potential impacts. This study used ML models and MODFLOW as an NM to perform a GWLP, utilizing aquifer features and data from 36 OBWs. SVM provides the most accurate modeling for GWLP, with AUC of 0.94, RMSE of 0.12, and R^2 of 0.91. When compared to other models used in the study, MODFLOW outperformed the others in the GeA index. However, all methods produced good GWL modeling results. MODFLOW achieved an RMSE value of 0.84, indicating its effectiveness in simulation. The most efficient features in ML prediction are identified by employing the PCA technique. Four features of OBWs and OPWs, along with their latitude and longitude, were introduced as ML input. According to PCA, these four input features contributed to ML accuracy by 71, 7, 13, and 9%, respectively. PCA also determined that the fifth time period was the most effective for CNN and SVM models, while the sixth time period was the most effective for DT. One of the most notable distinctions between the two approaches is the amount of time required for GWLP. However, physically based models like MODFLOW are not hindered by non-stationary conditions because they incorporate physical modules. Finally, the authors recommend modeling the entire Yazd-Ardakan Plain using a hybrid algorithm for groundwater forecasting, especially for CNN. Future studies could assess the autocorrelation of groundwater variations in the study area to identify patterns or dependencies within the GWL time series. Additionally, the use of the detrended fluctuation analysis (DFA) method could help remove trends and focus on the intrinsic fluctuations of GWL data from the OBWs.

ACKNOWLEDGEMENTS

The authors express their gratitude to the editor and reviewing team for their helpful comments, which helped elevate the quality of this work. They also thank Yazd Regional Water for supplying the data.

FUNDING

The authors assert that they have no known financial or personal relationships that could have potentially influenced the work reported in this paper.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Afrifa, S., Zhang, T., Appiahene, P. & Varadarajan, V. (2022) Mathematical and machine learning models for groundwater level changes: a systematic review and bibliographic analysis, *Future Internet*, **14** (9), 259. <https://doi.org/10.3390/fi14090259>.
- Afzaal, H., Farooque, A. A., Abbas, F., Acharya, B. & Esau, T. (2020) Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning, *Water (Switzerland)*, **12** (1), 5. <https://doi.org/10.3390/w12010005>.
- Almuhaylan, M. R., Ghumman, A. R., Al-Salamah, I. S., Ahmad, A., Ghazaw, Y. M., Haider, H. & Shafiquzzaman, M. (2020) Evaluating the impacts of pumping on aquifer depletion in arid regions using MODFLOW, ANFIS and ANN, *Water (Switzerland)*, **12** (8), 2297. <https://doi.org/10.3390/w12082297>.
- Amaranto, A., Munoz-Arriola, F., Solomatine, D. P. & Corzo, G. (2018) Semi-seasonal groundwater forecast using multiple data-driven models in an irrigated cropland, *Journal of Hydroinformatics*, **20** (6), 1227–1246. <https://doi.org/10.2166/hydro.2018.002>.
- Amaranto, A., Munoz-Arriola, F., Corzo, G., Solomatine, D. P. & Meyer, G. (2019) A spatially enhanced data-driven multimodel to improve semiseasonal groundwater forecasts in the High Plains aquifer, USA, *Water Resources Research*, **55** (7), 5941–5961. <https://doi.org/10.1029/2018WR024301>.
- Amiri, S., Rajabi, A., Shabanlou, S., Yosefvand, F. & Izadbakhsh, M. A. (2023) Prediction of groundwater level variations using deep learning methods and GMS numerical model, *Earth Science Informatics*, **16** (4), 3227–3241. <https://doi.org/10.1007/s12145-023-01052-1>.
- Ayvaz, M. T. & Elçi, A. (2013) A groundwater management tool for solving the pumping cost minimization problem for the Tahtali watershed (Izmir-Turkey) using hybrid HS-Solver optimization algorithm, *Journal of Hydrology*, **478**, 63–76. <https://doi.org/10.1016/j.jhydrol.2012.11.045>.
- Bana Bafroei, H. R. B. & Alimohammadi, S. (2023) Economic comparison of groundwater artificial recharge options using treated wastewater (Yazd-Ardakan Aquifer), *Journal of Water and Wastewater*, **33** (5), 142–150. <https://doi.org/10.22093/wwj.2022.347603.3267>.
- Banadkooki, F. B., Xiao, Y., Malekinezhad, H. & Hosseini, M. M. (2022) Optimal allocation of regional water resources in an arid basin: insights from integrated water resources management, *Aqua Water Infrastructure, Ecosystems and Society*, **71** (8), 910–925. <https://doi.org/10.2166/aqua.2022.029>.
- Beyene, T. D., Zimale, F. A. & Gebrekristos, S. T. (2023) A review on sources of uncertainties for groundwater recharge estimates: insight into data scarce tropical, arid, and semiarid regions, *Hydrology Research*, **55** (1), 51–66. <https://doi.org/10.2166/nh.2023.221>.
- Bouwer, H. (2002) Artificial recharge of groundwater: hydrogeology and engineering, *Hydrogeology Journal*, **10** (1), 121–142. <https://doi.org/10.1007/s10040-001-0182-4>.
- Cai, H., Shi, H., Liu, S. & Babovic, V. (2021) Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: the case of central eastern continental United States, *Journal of Hydrology: Regional Studies*, **37** (September), 100930. <https://doi.org/10.1016/j.ejrh.2021.100930>.
- Chakraborty, A. & Prakash, O. (2023) Identification of clandestine groundwater pollution sources in aquifer polluted with BTEX, *Journal of Water Resources Planning and Management*, **149**. <https://doi.org/10.1061/JWRMD5.WRENG-5078>.
- 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, *Scientific Reports*, **10** (1), 1–13. <https://doi.org/10.1038/s41598-020-60698-9>.
- Chen, Y., Chen, W., Pal, S. C., Saha, A., Chowdhuri, I., Adeli, B., Janizadeh, S., Dineva, A. A., Wang, X. & Mosavi, A. (2022) Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential, *Geocarto International*, **37** (19), 5564–5584. <https://doi.org/10.1080/10106049.2021.1920635>.
- Chenini, I. & Ben Mammou, A. (2010) Groundwater recharge study in arid region: an approach using GIS techniques and numerical modeling, *Computers and Geosciences*, **36** (6), 801–817. <https://doi.org/10.1016/j.cageo.2009.06.014>.
- Choi, C., Kim, J., Han, H., Han, D. & Kim, H. S. (2020) Development of water level prediction models using machine learning in wetlands: a case study of Upo wetland in South Korea, *Water*, **12** (1), 93. <https://doi.org/10.3390/w12010093>.
- Ebrahimi, R. S., Eslamian, S. & Zareian, M. J. (2023) Groundwater level prediction based on GMS and SVR models under climate change conditions: case study – Talesh Plain, *Theoretical and Applied Climatology*, **151** (1–2), 433–447. <https://doi.org/10.1007/s00704-022-04294-z>.
- Gholami, V. (2022) Spatial modeling of groundwater depth fluctuations using co-active neuro-fuzzy inference system (CANFIS) and geographic information system (GIS), *Applied Water Science*, **12** (3), 24. <https://doi.org/10.1007/s13201-022-01572-z>.
- Goodarzi, M. (2024) A machine learning approach for predicting and localizing the failure and damage point in sewer networks due to pipe properties, *Journal of Water and Health*, **22** (3), 487–509. <https://doi.org/10.2166/wh.2024.249>.
- Goodarzi, M. R., Abedi, M. J., Niknam, A. R. R. & Heydaripour, M. (2022) Groundwater quality status based on a modification of water quality index in an arid area, Iran, *Water Supply*, **22** (7), 6245–6261. <https://doi.org/10.2166/ws.2022.225>.
- Goodarzi, M. R., Niknam, A. R. R., Barzkar, A., Niazkar, M., Mehrjerdi, Y. Z., Abedi, M. J. & Pour, M. H. (2023) Water quality index estimations using machine learning algorithms: a case study of Yazd-Ardakan Plain, Iran, *Water (Switzerland)*, **15** (10), 1876. <https://doi.org/10.3390/w15101876>.
- Goodarzi, M. R., Niazkar, M., Barzkar, A. & Niknam, A. R. R. (2024a) Assessment of machine learning models for short-term streamflow estimation: the case of Dez River in Iran, *Sustainable Water Resources Management*, **10** (1), 1–16. <https://doi.org/10.1007/s40899-023-01021-y>.

- Goodarzi, M. R., Sabaghzadeh, M., Barzkar, A., Niazkar, M. & Saghafi, M. (2024b) A comparison of machine learning methods for estimation of snow density using satellite images, *Water and Environment Journal*, **38** (3), 437–449. <https://doi.org/10.1111/wej.12939>.
- Goodarzi, M. R., Vazirian, M. & Niazkar, M. (2024c) Hydraulic conductivity estimation: comparison of empirical formulas based on new laboratory experiments, *Water*, **16** (13), 1854. <https://doi.org/10.3390/w16131854>.
- Harbaugh, A.W, Banta, E. R., Hill, M. C. & McDonald, M. G. (2000) *MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model: User Guide to Modularization Concepts and the Ground-Water Flow Process*, Open-File Report.
- Khan Mohammad, S. & Coulibaly, P. (2006) Application of support vector machine in lake water level prediction, *Journal of Hydrologic Engineering*, **11** (3), 199–205. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:3\(199\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:3(199)).
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998) Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86** (11), 2278–2324. <https://doi.org/10.1109/5.726791>.
- Lenin Sundar, M., Ragunath, S., Hemalatha, J., Vivek, S., Mohanraj, M., Sampathkumar, V., Mohammed, S. A., Parthiban, V. & Manoj, S. (2022) Simulation of ground water quality for Noyyal River basin of Coimbatore city, Tamilnadu using MODFLOW, *Chemosphere*, **306**, 135649. <https://doi.org/10.1016/j.chemosphere.2022.135649>.
- Li, X., Sha, J. & Wang, Z. (2016) A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen, *Hydrology Research*, **48** (5), 1214–1225. <https://doi.org/10.2166/nh.2016.149>.
- Liang, Y., Zhang, X., Sun, Y., Yao, L., Gan, L., Wu, J., Chen, S., Li, J. & Wang, J. (2024) Enhanced groundwater vulnerability assessment to nitrate contamination in Chongqing, Southwest China: integrating novel explainable machine learning algorithms with DRASTIC-LU, *Hydrology Research*, **55** (6), 663–682. <https://doi.org/10.2166/nh.2024.036>.
- Maghsoudi, R., Javadi, S., Shourian, M. & Golmohammadi, G. (2023) Determining the optimal aquifer exploitation under artificial recharge using the combination of numerical models and particle swarm optimization, *Hydrology*, **10** (5), 100. <https://doi.org/10.3390/hydrology10050100>.
- Malekzadeh, M., Kardar, S. & Shabanlou, S. (2019) Simulation of groundwater level using MODFLOW, extreme learning machine and wavelet-extreme learning machine models, *Groundwater for Sustainable Development*, **9**, 100279. <https://doi.org/10.1016/j.gsd.2019.100279>.
- Mallick, J., Talukdar, S., Alsubih, M., Ahmed, M., Islam, A. R. & Thanh, N. V. (2022) Proposing receiver operating characteristic-based sensitivity analysis with introducing swarm optimized ensemble learning algorithms for groundwater potentiality modelling in Asir region, Saudi Arabia, *Geocarto International*, **37** (15), 4361–4389. <https://doi.org/10.1080/10106049.2021.1878291>.
- Ministry of Energy (2012) *Studies on Updating the Comprehensive Water Plan of the Siah Kuh-Rig Zarrin Catchment, Case Study of Yazd-Ardekan*. Tehran: Ministry of Energy.
- Moghaddam, H. K., Moghaddam, H. K., Kivi, Z. R., Bahreinimotlagh, M. & Alizadeh, M. J. (2019) Developing comparative mathematic models, BN and ANN for forecasting of groundwater levels, *Groundwater for Sustainable Development*, **9**, 100237. <https://doi.org/10.1016/j.gsd.2019.100237>.
- Mohammed, K. S., Shabanlou, S., Rajabi, A., Yosefvand, F. & Izadbakhsh, M. A. (2023) Prediction of groundwater level fluctuations using artificial intelligence-based models and GMS, *Applied Water Science*, **13** (2), 1–14. <https://doi.org/10.1007/s13201-022-01861-7>.
- Mohanty, S., Jha, M. K., Kumar, A. & Panda, D. K. (2013) Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India, *Journal of Hydrology*, **495**, 38–51. <https://doi.org/10.1016/j.jhydrol.2013.04.041>.
- Pearson, K. (1901) LIII. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **2** (11), 559–572. <https://doi.org/10.1080/14786440109462720>.
- Quinlan, J. R. (1986) Induction of decision trees, *Machine Learning*, **1** (1), 81–106. <https://doi.org/10.1007/BF00116251>.
- Rahman, A. T. M. S., Hosono, T., Quilty, J. M., Das, J. & Basak, A. (2020) Multiscale groundwater level forecasting: coupling new machine learning approaches with wavelet transforms, *Advances in Water Resources*, **141**, 103595. <https://doi.org/10.1016/j.advwatres.2020.103595>.
- 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. <https://doi.org/10.3390/su12104023>.
- Seifu, T. K., Eshetu, K. D., Woldesenbet, T. A., Alemayehu, T. & Ayenew, T. (2023) Application of advanced machine learning algorithms and geospatial techniques for groundwater potential zone mapping in Gambela Plain, Ethiopia, *Hydrology Research*, **54** (10), 1246–1266. <https://doi.org/10.2166/nh.2023.083>.
- Sherif, M., Ebraheem, A. A., Shetty, A., Sefelnasr, A., Alghafli, K. & Al Asam, M. (2022) 'Evaluation of the effect of the Wadi Bih Dam on groundwater recharge, UAE', *World Environmental and Water Resources Congress 2017*, pp. 509–527.
- Tigabu, T. B., Wagner, P. D., Hörmann, G. & Fohrer, N. (2020) Modeling the spatio-temporal flow dynamics of groundwater-surface water interactions of the Lake Tana Basin, Upper Blue Nile, Ethiopia, *Hydrology Research*, **51** (6), 1537–1559. <https://doi.org/10.2166/nh.2020.046>.
- Vapnik, V. N. (1995) *The Nature of Statistical Learning Theory*. New York: Springer.
- Wunsch, A., Liesch, T. & Broda, S. (2021) Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX), *Hydrology and Earth System Sciences*, **25** (3), 1671–1687. <https://doi.org/10.5194/hess-25-1671-2021>.

- Xu, H., Wang, D., Ding, Z., Deng, Z., Shi, Y., Yu, D., Li, J. & Ni, B. (2020) Application of convolutional neural network in predicting groundwater potential using remote sensing : a case study in southeastern Liaoning, China, *Arabian Journal of Geosciences*, **13** (739). <https://doi.org/10.1007/s12517-020-05585-3>.
- Yifru, B. A., Chung, I-M., Kim, M-G. & Chang, S. W. (2022) Assessing the effect of urbanization on regional-scale surface water-groundwater interaction and nitrate transport, *Scientific Reports*, **12** (1), 12520. <https://doi.org/10.1038/s41598-022-16134-1>.
- Yin, J., Pham, H. V. & Tsai, F. T.-C. (2020) Multiobjective spatial pumping optimization for groundwater management in a multiaquifer system, *Journal of Water Resources Planning and Management*, **146** (4), 1–12. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001180](https://doi.org/10.1061/(asce)wr.1943-5452.0001180).
- Yoon, H., Hyun, Y., Ha, K., Lee, K. K. & Kim, G. B. (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, *Journal of Hydrology*, **396** (1–2), 128–138. <https://doi.org/10.1016/j.jhydrol.2010.11.002>.
- Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O. & Lee, K. K. (2016) A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions, *Computers and Geosciences*, **90**, 144–155. <https://doi.org/10.1016/j.cageo.2016.03.002>.
- Zhang, J., Chen, X., Khan, A., Zhang, Y., Kuang, X., Liang, X., Taccari, M. L. & Nuttall, J. (2021) Daily runoff forecasting by deep recursive neural network, *Journal of Hydrology*, **596** (December 2020), 126067. <https://doi.org/10.1016/j.jhydrol.2021.126067>.

First received 11 September 2024; accepted in revised form 29 November 2024. Available online 12 December 2024