



Groundwater level prediction using machine learning algorithms in a drought-prone area

Quoc Bao Pham^{1,2} · Manish Kumar³ · Fabio Di Nunno⁴ · Ahmed Elbeltagi⁵ · Francesco Granata⁴ · Abu Reza Md. Towfiqul Islam⁶ · Swapan Talukdar⁷ · X. Cuong Nguyen^{8,9} · Ali Najah Ahmed¹⁰ · Duong Tran Anh¹¹

Received: 17 May 2021 / Accepted: 30 January 2022 / Published online: 3 March 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Groundwater resources (GWR) play a crucial role in agricultural crop production, daily life, and economic progress. Therefore, accurate prediction of groundwater (GW) level will aid in the sustainable management of GWR. A comparative study was conducted to evaluate the performance of seven different ML models, such as random tree (RT), random forest (RF), decision stump, M5P, support vector machine (SVM), locally weighted linear regression (LWLR), and reduce error pruning tree (REP Tree) for GW level (GWL) prediction. The long-term prediction was conducted using historical GWL, mean temperature, rainfall, and relative humidity datasets for the period 1981–2017 obtained from two wells in the northwestern region of Bangladesh. The whole dataset was divided into training (1981–2008) and testing (2008–2017) datasets. The output of the seven proposed models was evaluated using the root mean square error (RMSE), mean absolute error (MAE), relative absolute error (RAE), root relative squared error (RRSE), correlation coefficient (CC), and Taylor diagram. The results revealed that the Bagging-RT and Bagging-RF models outperformed other ML models. The Bagging-RT models can effectively improve prediction precision as compared to other models with RMSE of 0.60 m, MAE of 0.45 m, RAE of 27.47%, RRSE of 30.79%, and CC of 0.96 for Rajshahi and RMSE of 0.26 m, MAE of 0.18 m, RAE of 19.87%, RRSE of 24.17%, and 0.97 for Rangpur during training, and RMSE of 0.60 m, MAE of 0.40 m, RAE of 24.25%, RRSE of 29.99%, and CC of 0.96 for Rajshahi and RMSE of 0.38 m, MAE of 0.24 m, RAE of 23.55%, RRSE of 31.77%, and CC of 0.95 for Rangpur during testing stages, respectively. Our study offers an effective and practical approach to the forecast of GWL that could help to formulate policies for sustainable GWR management.

Keywords Groundwater prediction · Machine learning · Bangladesh · Locally weighted linear regression

1 Introduction

The fast-growing use of groundwater (GW), especially in agrarian-based developing countries like Bangladesh, is now a prime concern for decision-makers on the issue of sustainable groundwater resources (GWR) management [77, 82]. In recent decades, climatic variability (e.g., rainfall infiltration rate, surface runoff, evaporation, and increase in temperature), fast population growth, and over-exploitation have had adverse effects on this precious GW resource in many regions across the globe [18, 50, 86, 104]. This is particularly true for Bangladesh, where high population density and increasing industrial development are

raising water demand [41, 42]. As a result, GW levels (GWL) are decreasing, with a general increase in water stress. GWL monitoring and modeling are essential activities for sustainable GWR management in the water-stressed regions, especially in the northwestern drought-prone area, Bangladesh.

The GWL datasets obtained by continuous monitoring provide valuable information on the aquifer. The efficient management of GW resources needs an accurate estimation of current and predicted demand of GW and the net GW resource of the basin at the regional and local scales [108]. The research on water table depth oscillations underneath the ground surface level is essential to extract data associated with the rate of recharge, discharge, and storage capacity. GW management is solely feasible after

Extended author information available on the last page of the article

obtaining relevant GW resource information followed by an accurate analysis [72]. The applicability of conventional groundwater simulation models in Bangladesh is limited due to a lack of detailed knowledge of aquifer characteristics. In such a case, soft computing machine learning (ML) tools are an optimal alternative that shows higher potential and efficacy than the conventional numerical or physical process-based approaches.

In recent years, with the progress in soft computational data-driven tools, several ML models have been developed and employed for GWL forecasting [73]. For instance, Artificial Neural Networks (ANN) [5, 24, 25, 32, 33, 41, 59, 61, 90], Support Vector Machine (SVM) [35, 105, 109], Adaptive-Network-based Fuzzy Inference System (ANFIS) (Fallah-Mehdipour et al. [30, 51], Extreme learning machine (ELM) [5, 40, 102], Fuzzy logic (Nadiri et al. 2018), and the M5 model [54] are some of the most popular and widely adopted ML models in GWL prediction. Although some benefits of the ANNs and ANFIS models are to estimate the unknown variables via simple training tools and identify the complex nonlinear association between the predictors and objective parameters, they exhibit some shortcomings, including their over-fitting problem and “black box” nature [28]. Besides, SVM has various kernel functions and, for each performance measure, all of them should be assessed and the optimal one chosen, limiting its use [89]. Apart from these standalone models, tree-based models such as Random Tree, Random Forest, and REP Tree do not require preprocessing of datasets and are simply mapped data features, but they have weak predictability, notably inhibiting scientists from employing them in robust scientific work [34, 91].

A good number of studies have employed hybrid ensemble ML and genetic models for different hydro-meteorological applications [9, 52, 64, 92], Nourani and Mousavi [68, 77, 88, 103, 107]. For example, Moosavi et al. [64] compared the performance of hybrid ensemble Wavelet-ANFIS and Wavelet-ANN models for predicting GWL in Iran and found that Wavelet-ANFIS was the optimal model. Similarly, Suryanarayana et al. [92] compared standalone models such as ANN, SVM, and autoregressive integrated moving average (ARIMA) models with hybrid ML models such as Wavelet-SVM and found that wavelet-SVM provided better precision. Khalil et al. [52] revealed that the Wavelet-ANN ensemble models outperformed single models such as MLR, ANN, in predicting short-term GWL in Iran. Nourani and Mousavi [68] predicted the GWL using a hybrid ensemble of ANN-RBF and ANFIS-RBF models combined with a threshold-based wavelet-denoising model and found that ANFIS-RBF performed better than ANN-RBF owing to the modeling effect being denoised. Yu et al. [107] compared the hybrid wavelet-ANN and wavelet-SVM to the ANN and SVM benchmark models for monthly GWL forecasting in

northwest China to show that wavelet-SVM yielded better outcomes. Rezaie-balf et al. [77] assessed the performance of wavelet coupled multivariate adaptive regression splines (wavelet-MARS) and M5 model tree (wavelet-MT) to MARS and MT standalone models for GWL forecasting and found that the forecast precision of wavelet-MARS was superior to other models in one, three, and six months ahead prediction. Yadav et al. [103] developed an ensemble model to forecast the monthly GWL in the municipal region of Bengaluru, India, and found that the hybrid ensemble SVM-HSVM yielded better performance than the single ANN and SVM models. Sharafati et al. [88] proposed a novel ML model for predicting GWL in Rafsanjan aquifer, Iran and showed that the ensemble Gradient Boosting Regression (GBR) has high accuracy. According to the literature review, all of these hybrid ensemble ML models outperformed the standalone model in terms of prediction. The major benefit of hybrid models is that they often uncover complicated mathematical nonlinear relationships between the objective and predictive parameters.

The promising results of these hybrid ML models offered inspiration for assessment of various model-tool couplings in different hydrological issues [64, 90, 100, 105–107]. However, this study demonstrates the ability of some ensemble ML models, such as Bagging, decision stump, random tree, and REP Tree, to predict GW levels. The use of meta-based ensemble classifiers (Bagging) and tree-based ensemble classifiers (random tree, REP Tree, decision stump) to improve the prediction accuracy of a single classifier is relatively new in hydrological studies. The Bagging model is a common hybrid model used in various hydro-meteorological modeling tasks, such as landslide susceptibility mapping [69], GW potential prediction [67], flood susceptibility mapping [43], suspended sediment load prediction [83], and evapotranspiration studies [37, 81]. The key benefit of the Bagging algorithm is that it can solve nonlinear complex problems better than other algorithms [53]. The Bagging uses bootstrap aggregation, in which classifiers are trained in parallel as the bags from each training input data set. Random tree and decision stump, like Bagging, can assist with over-fitting problems and have high accuracy. Based on the previous literature, the two objectives have been set to tackle the research gaps. The objectives are to (1) improve the groundwater predictionability by developing bagging based seven hybrid machine learning algorithms, and (2) validate the the predicted models at training and testing periods. However, based on above discussion, followings could be the most significant novelties of the present study:

- *General* The work contributes to the robustness of knowledge by designing and using methodologies for

groundwater level prediction in an unstudied area of Bangladesh with greater GW consumption.

- *Regional* Enhanced understanding of groundwater forecast in Bangladesh's northwestern drought-prone region. This effort would provide a significant foundation for earth scientists, government officials, and stakeholders to enhance water resource management and sustainable agriculture management.
- *Methodical* Proposed seven hybrid ML models for GWL prediction that combine Bagging as the base classifier with locally weighted linear regression (LWLR), SVM, M5P, REP Tree, RT, RF, and decision stump. To the best of our experience, these hybrid ML models have never been used in previous research on GWL forecasting.

2 Study area and datasets

The northwestern region of Bangladesh is geographically located between 24°N–26°N latitude and 88°E–89°E longitude [47], covering an area of 34,600 km² (Fig. 1). The study area shares its border with India on the north and west part, while it is bordered by the Brahmaputra-Jamuna River to the east and the Ganges River to the south. The northern portion of the study area is situated in the Himalayan piedmont fans having an average height higher than 90 m. The southern portion of the study area is characterized by a swamp having an average height of less than 10 m. The study area is dominated by the subtropical monsoon climate and has three distinct seasons, such as winter (November–February with cool-dry weather and no rainfall; pre-monsoon (March–May with hot and dry; and monsoon (June–October with heavy rainfall [47]. In the study area, monthly average temperature, evapotranspiration, and rainfall, respectively, are 10–22 °C, 4 mm, and 125 mm. The study area comprises two divisions: Rangpur division and Rajshahi division. The meteorological stations are distributed in the mentioned division. For the present study, the meteorological data for two stations were available: one station from Rangpur division, and another from Rajshahi division (Fig. 1).

The physiography of the study area is characterized by alluvial plain with slightly elevated Pleistocene terraces, and the slope is facing toward the south and southeast [75]. The study area is dominated by the Rangpur Saddle under the Indian platform in the sub-section of Bengal basin, Bangladesh (Akhter et al. 2019). About 80% of the study area is composed of alluvial soil, while the rest of the area is composed of Barind clay. The study area is mostly dominated by agricultural land because of the presence of

fertile soil [45]. The groundwater has mainly been extracted for domestic and agricultural purposes, and a little amount is used for drinking uses [44, 62].

The daily rainfall, relative humidity, and temperature data of Rajshahi and Rangpur stations were collected from Bangladesh Meteorological Department, Dhaka (BMD 2018) (www.bmd.gov.bd). The ground water level data of Rangpur and Rajshahi (Fig. 2) were measured from the Bangladesh Water Development Board (BWDB), Dhaka (www.bwdb.gov.bd). Overall, the dataset covers a period from January 1981 to December 2017.

In general, the fluctuation of groundwater levels relies mainly on the hydro-geological and meteorological variables such as the amount of rainfall, temperature, humidity, the lithological composition of rocks, the level of the drainage capacity of the region, and aquifer characteristics [82, 108]. For that reason, this study has chosen the three meteorological variables, e.g., rainfall, temperature, and humidity for investigating the fluctuation of GWL in the study area.

Furthermore, Tables 1 and 2 provide a descriptive statistic, for both the training and testing stages, of the collected data for Rajshahi and Rangpur stations respectively, where SEM indicates the Standard Error of the Mean, σ the standard deviation and Q1 and Q3, respectively, the first and third quartile. The standard deviation of rainfall is large, as seen in Tables 1 and 2. High standard deviation numbers indicate greater uncertainty than low standard deviation values, implying that deviations from the normal distribution should not be overlooked. The fact that the standard deviation values are so large indicates that the rainfall is very unpredictable, changeable, and irregular, or it may be attributed to the length of datasets used or the dataset quality.

3 Methods

3.1 Bagging model

The Bagging is (short form for “Bootstrap Aggregating”) method consists of two main stages [11, 15]. The first step consists of bootstrapping the samples gained from the raw data, in order to get different sets of training data. Then, multiple models are generated from these training datasets. Final predictions are obtained combining the outcomes of multiple models (Fig. 3).

In this study, different regression trees were combined to obtain a single output by using weighted average. In the model development stage, various exercise datasets with similar size have been chosen randomly from an issue domain. Then, a regression tree model has been developed for each input dataset. Each tree is different and leads to a

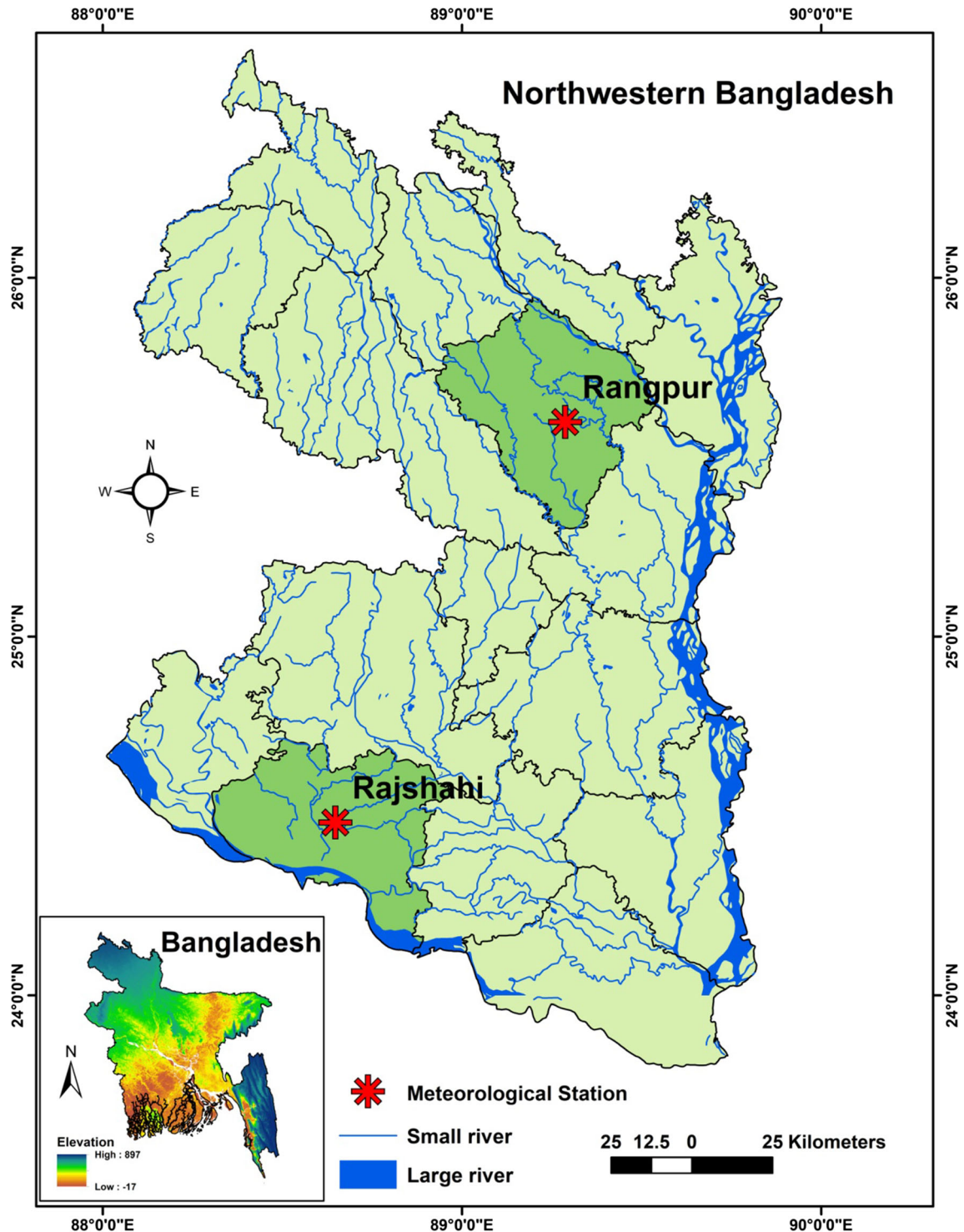


Fig. 1 Location of the study area and two groundwater station

different prediction based on the alteration provided in the training dataset. One of the important benefits of the bagging procedure is that it removes the existing instability present in the regression tree growth. This was done by removing the initial training datasets rather than novel training dataset sampling for each time step. Finally, all the

forecasts of the regression tree have been weighted average. The description of the parameters selected in this method are Batch size=100, bag Size percent=80, classifier =REPTree, max depth=0, numbers of executions slots=1, number of iteration=100, and random seed=1.

Fig. 2 The time-series plot of groundwater level for Rajshahi and Rangpur stations

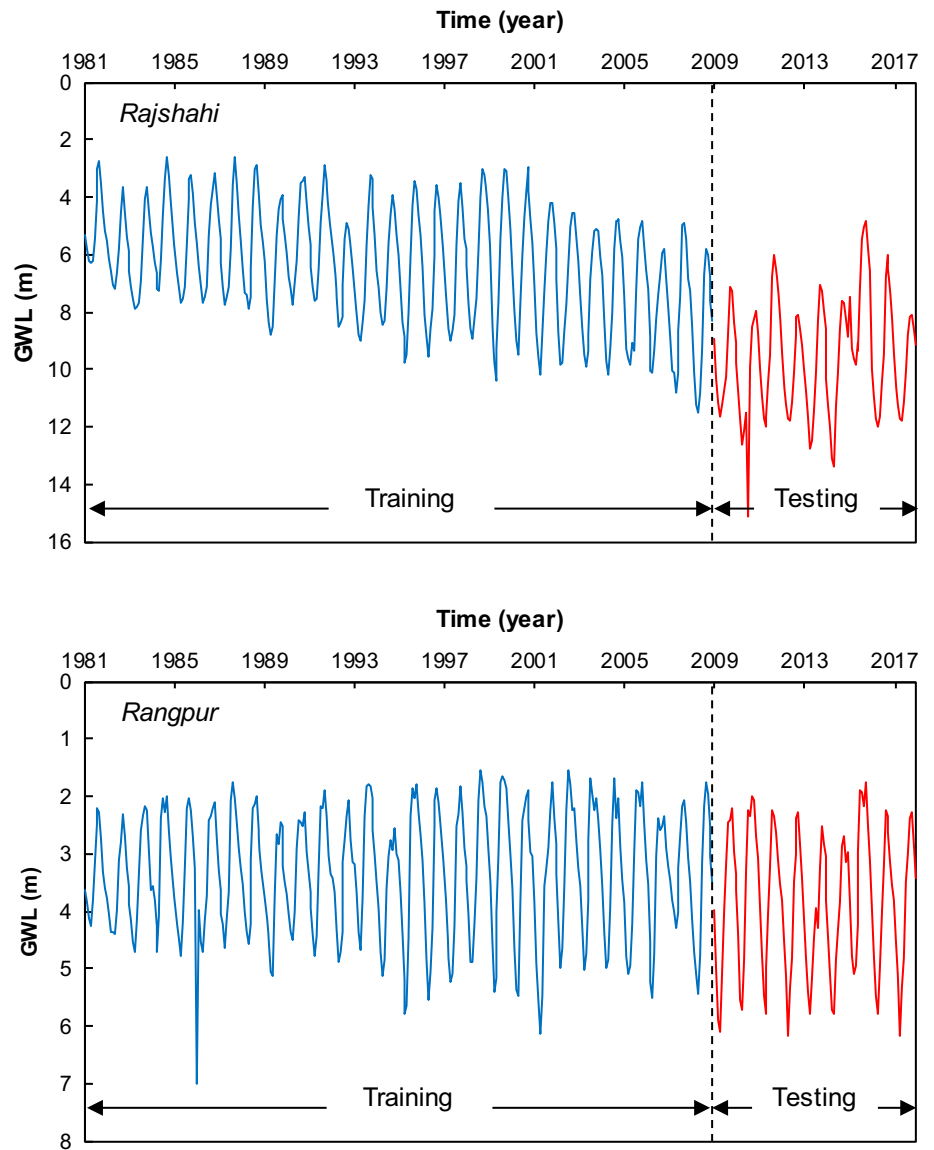


Table 1 Descriptive statistics for Rajshahi station

Stage	Variable	Mean	SEM	σ	Minimum	Q1	Median	Q3	Maximum
Training	T (°C)	25.72	0.24	4.40	15.00	22.46	27.90	29.18	31.67
	R (mm)	124.90	7.67	140.68	0.00	7.00	72.00	220.70	763.00
	H (%)	77.98	0.47	8.66	50.80	72.88	79.15	85.34	90.25
	GWL (m)	6.36	0.11	1.97	2.58	4.82	6.24	7.73	11.47
Testing	T (°C)	26.10	0.46	4.76	15.55	22.35	28.22	29.89	31.61
	R (mm)	103.50	10.60	109.90	0.00	5.30	77.50	178.00	454.00
	H (%)	79.10	0.70	7.37	58.96	76.40	81.21	84.60	88.83
	GWL (m)	9.53	0.19	2.01	4.83	8.07	9.54	11.20	15.13

Table 2 Descriptive statistics for Rangpur station

Stage	Variable	Mean	SEM	σ	Minimum	Q1	Median	Q3	Maximum
Training	T (°C)	26.05	0.24	4.40	15.58	22.61	28.21	29.54	32.23
	R (mm)	197.10	12.20	223.70	0.00	7.30	107.50	346.80	1314.00
	H (%)	80.69	0.35	6.56	55.64	77.51	82.29	85.54	90.90
	GWL (m)	3.34	0.06	1.07	1.55	2.38	3.24	4.17	7.01
Testing	T (°C)	26.24	0.46	4.74	15.66	22.38	28.39	29.92	31.86
	R (mm)	168.10	18.70	194.70	0.00	2.00	112.50	297.50	832.00
	H (%)	79.52	0.52	5.48	63.20	76.79	80.74	83.71	89.29
	GWL (m)	3.82	0.11	1.21	1.74	2.83	3.80	4.84	6.15

Fig. 3 Typical architecture of Bagging algorithm. The algorithms are different for the way regression trees are built [38]

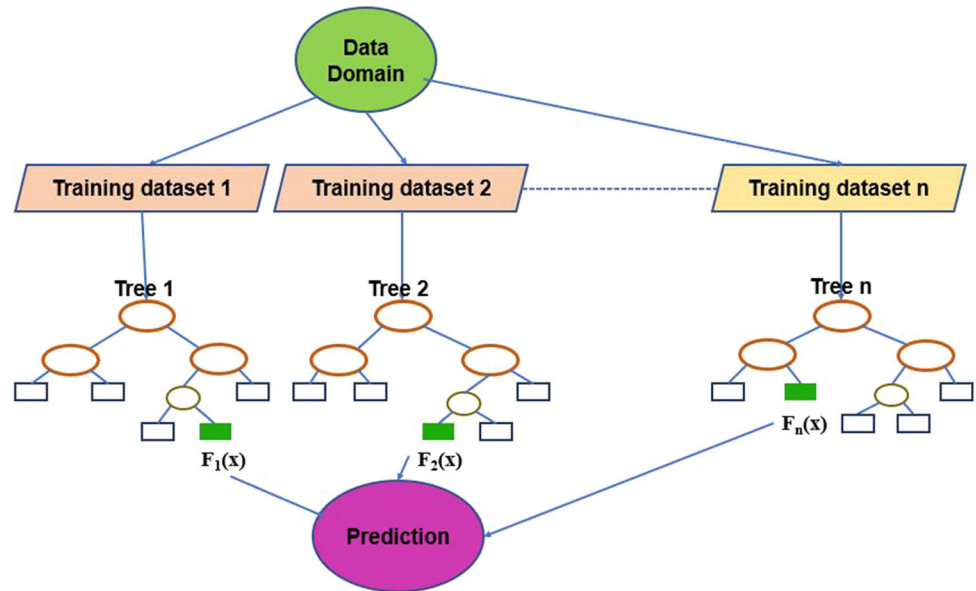
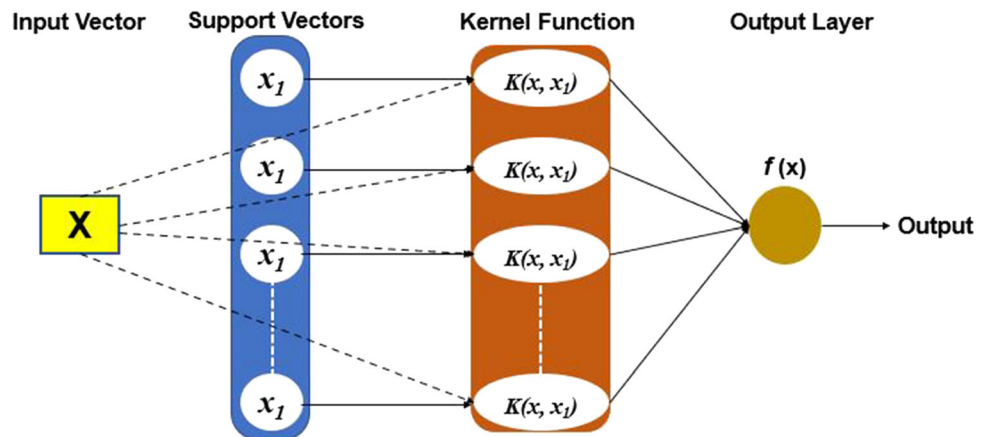


Fig. 4 A schematic diagram of the SVM algorithm [85]



3.2 Locally weighted linear regression (LWLR) method

Locally weighted regression (LWR) is motivated by the classification of the instance-based methods [7]. In this method, the regression model is processed until the output value of the new vector is presented. This is done to

perform all learning accurately at the prediction time. LWR is an advanced form of M5 models in a way that it fits linear as well as nonlinear regression for the particular areas of instance in space. Distances according to the query are used to allocate the weights to the training datasets, and regression equations are generated based on the weighted data. There are a wide variety of distance-based weighting

methods which can be used in LWR based on the choice of the problem. In this study, the parameters used in this method are as follows: Batch size=100, KNN=0, nearest neighbor search algorithm=linear NN.

3.3 SMO-SVM algorithm

Support Vector Machine (SVM) algorithms are basically structured on statistical learning theories and developed from the conceptual optimization hypothesis [94, 95] (Fig. 4). It is generally used in order to achieve the best generalization capability for both the empirical relations and confidence intervals of machine learning. These SVMs have always been shown to perform extremely well and efficiently for optimization and regression studies [96], Collobert et al. [20]). These proved to be highly robust in nature for extremely noise mixed data in comparison to the other local models and algorithms which use traditional chaotic methods. The SVM estimation function (F) in any given regression scenario can be defined as:

$$F(x) = W.T_f(x) + b \quad (1)$$

where, W is the weight age vector; T_f represents the non-linear transfer function, which projects the input vectors toward a very high dimension feature space; and b is the constant variable.

A convex dual optimization problem was introduced by Vapnik [94] that created an insensitivity loss function. Several algorithms have been developed and suggested for solving these dual optimization problems in the SVM (Adamala and Srivastava [2]). In this current study, the Sequential Minimal Optimization (SMO) algorithm is used [70]. The programming codes from the Library for Support Vector Machines (LIBSVM) are used for the calibration and validation of all the datasets [17]. In this work, the Batch size=100, $C=0.1$, kernel used=poly kernel were used in prediction model.

3.4 Random forest (RF)

The Random Forest model was proposed by Breiman [11, 15]. Nowadays, this method is a very effectively applied on the various fields such as hydrology, land use classification, irrigation scheduler, evaporation measurement, forest and crops classification [29]. It is an ensemble of different regression tree models, built from different subsets obtained using bootstrap resampling technique. The original information's concealed outcomes are based on number choices made by categorization of tree divisions, and the decision trees' overall average forecast is considered as the regression model's final predictions (Dietterich et al. [26]). A basic regression tree has the disadvantage of being sensitive to the training data used to build the tree,

especially if the training data are minimal [21]. There may be slight differences in exercise data between trees and predictions [14]. These disadvantages can be solved by advancing random forest technique.

Rodriguez et al. [78] proposed cluster methods based on the alternate forest methodology, which used features derived from the alternate sub-spaces of a large dataset to improve the accuracy of inadequate regression and classification models [12]. By randomly splitting the input distinguishing usual into P distinguishing subsets (one factor has been specified by handler), the method operates similarly to the pseudo-random approach, and then applies a key factor analysis to these sub-section's characteristics. Every regression technique in the group has generated a different set of data. An important module remains in the information directive to maintain the diversity of data in the information. The parameters selected in this method are Batch size=100, bag Size percent=100, max depth=0, numbers of executions slots=1, number of iteration=100, and random seed=1.

3.5 REPTree

The REPTree decision algorithm is a very rapid method of learning with reduced error pruning tree. It generates a decision/regression tree, which is pruned using back-fitting with decreased error [49]. The inaccuracy of the decision tree model is reduced with the help of the “reduce error pruning method,” so the mistake rising from alteration is decreased. For numeric characteristics, the algorithm only examines values once. It is primarily the method of constructing a common set of instructions for decision making using a forecaster variable quantity [12], Birendra [10]. The REPTree process is a basic decision tree beginning method that designs and utilizes condensed error trimming to create a tree of regression using variance data [84]. The numerical ranges in the model have been established [22]. In this case, the execution of several knowledge algorithms in the WEKA environment [98] was employed. The REPTree model is commonly used for hydrological system processes, surface runoff, and other disciplines. REPTree model provides accurate information in a variety of fields such as ecological planning, flood susceptibility, soil erosion, climate and hydrological processes, and it is very effectively applied in a variety of aspects such as irrigation planning, flood analysis, rainfall prediction, and evaporation, among others. This model has recently been employed by machine learning programming, and it is efficiently used by researchers and data scientists in python programming.

$$S = \sum_{E \in \text{leaves}} q_c U_c \quad (2)$$

where q_c is well-described as the class forecast and U_c is the leaf-inside variance. The description of parameters selected in this study for implementing this model was: Batch size=100, Initial count=0, number of folds=3, random seed=1, minimum proportion of the variance=0.001, minimum number=2, and max depth=1.

3.5.1 Random tree (RT)

The RT algorithm was constructed using a conventional technique and then refined using decision tree on an arbitrary column subdivision. This model was created using an approach similar to traditional trees and a single significant exclusion [97]. The model is highly fast and adaptable for a tree novice, and it may be used to teach a wide range of challenges [16, 66]. Supervised classifier is a collaborative learning model that is used to build more distinct beginnings. It provides a bagging impression to conceptualize an arbitrary standard of information for creating decision tree models. Each node in a typical tree is separated using the finest fragmented in entire variables. All nodes are separated by the finest between a subset of forecasters that was arbitrarily chosen at node.

Decision-making tree which each of the leaves acquires an optimized linear model to local subspace, which leaf has an explained. The recital of only decision trees has presented Random Forests to advance significantly: tree variety is created by two make random directions (Amit et al. [6]). The description of the parameters selected in this model are Batch size=100, max depth=0, Min Number=1, Min variance proportion=0.001, and random seed=1.

3.6 M5P

The model tree approach creates a physical representation of data and divides it into linear classes [71]. M5, like most result tree learners, creates a tree by splitting data into predicted values. M5 utilizes attributes to minimize the variation of intra-subset class values of instances lying inside each subdivision, rather than utilizing information-theoretical metrics to choose characteristics. When each characteristic of this knot has been evaluated to establish the expected error reduction, the inconsistency is determined by the standard deviation of the standards that the root-to-branch node achieves [93]. The property that maximizes the projected mistake reduction is designated. If the standards of all instances entering a node differ slightly, the division fails, and a rare occurrence continues (Goyal et al. [36], Ajmera et al. [4]). Decision Tree regression models have been thoroughly investigated in the field of machine learning. Quinlan [71] pioneered the use of algorithm trees and the M5 algorithm to solve issues

involving constant knowledge. The leaves follow the usual decision tree form, but instead of discrete class names, they use linear functions. During model forecasting, a soft process can be utilized to compensate for gaps between end-to-end linear models. The technique measures the forecast value and sieves it on the routine route, smoothing it at each node in accordance with the linear node value anticipated for that node using the Leaf Model. Quinlan explains the procedure (1992). The SDE is an estimated as:

$$SDR = sd(T) \sum \left| \frac{T_i}{T} \right| \times sd(T_i) \quad (3)$$

where, T are sets of examples that reach the node, “SD” denotes standard deviation and T_i are the sets subsequent from splitting the node as per a provided attribute and split value. The Batch size and Minimum number of instances were 100 and 4 for implementing this model in this study.

3.7 Decision stump

This forest algorithm aims to combine several classifiers of oblique decision stumps into a single collaborative approach. Each essential estimate meanings such as bias and variance have dissolved in the display of collaboratively developed research models (Dietterich et al. [27]). Bias is the regular error stated, while variance is the error derived from model variability using sample arbitrariness information. The success of collaborative classifiers is then reflected in the reduction oinvariance in the research algorithm. Bagging and Random forest [11, 14, 15] were used to reduce the variance of a learning model without introducing too much aggregate in bias. Boosting [31] and Arcx4 [13] both attempted to reduce bias at the same time.

These are simple classifiers in which just one theory or element is responsible for the final choice. While most general intelligence is assigned to an individual occurrence of this specified for trial information by popular classifier, a model of DS is a single point decision tree reliant on the coexistence function. This technique generates simple binary decision stumps for problems with a large number of variables and insignificant organization [99]. The batch size in this model was 100 in this investigation. Table 3 shows the parameters utilized in this investigation for each method.

3.8 Statistical analysis

Actual data of discharge (D) and modeled values were compared through the period of this study. To evaluate the accuracy of models, the following statistical indicators have been selected: Root Mean Square Error, Mean Absolute Error, Relative Absolute Error, Root Relative Squared Error, and Correlation Coefficient [57]. All

Table 3 The parameters of the machine learning algorithms used for groundwater level prediction

Model name	Description of parameters
Bagging tree (BT)	Batch size-80, bag Size percent=100, Classifier=REPTree, numbers of executions slots=1, number of iteration=100, random seed=1
locally weighted linear regression (LWLR)	Batch size-100, KNN=0, Nearest neighbor search algorithm=linear NN
Reduced error pruning tree (REPTree)	Batch size-100, Initial count=0, number of folds=3, random seed=1, minimum proportion of the variance =0.001, minimum number=2, max depth=1
Improved support vector machine (SMO-SVM)	Batch size-100, C=0.1, kernel used=polykernel
Random forest (RF)	Batch size-100, bag Size percent=100, max depth=0, numbers of executions slots=1, number of iteration =100, random seed=1
M5 Pruned M5P	Batch size-100, Minimum number of instances=4
Random tree (RT)	Batch size-100, seed=1, minimum variance proportion=0.001
Decision stump tree (DST)	Batch size-100

parameters are defined as follows: D_A^i is an observed or actual value D_p^i is simulated or foreseen value, D^- is the mean value of reference samples, and N is the total number of data points.

3.8.1 Root mean square error

RMSE refers to the sample standard deviation of the variations between expected and real values. It comes from the following formula:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (D_A^i - D_p^i)^2} \tag{4}$$

3.8.2 Mean absolute error (MAE)

The MAE assesses the extent of errors in a series of predictions without taking their sign into account. It’s an estimate of the absolute differences between expected and observed values over the test sample. It is defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |D_p^i - D_A^i| \tag{5}$$

3.8.3 Relative absolute error (RAE)

The total absolute error is normalized by dividing it by the total absolute error of the basic indicator in the RAE.

$$RAE = \left| \frac{D_A^i - D_p^i}{D_p^i} \right| \times 100 \tag{6}$$

3.8.4 Root relative squared error (RRSE)

The total squared error is normalized by dividing it by the total squared error of the basic indicator in the RSE. The error is reduced to the same dimensions as the quantity being predicted by taking the square root of the relative squared error.

$$RRSE = \frac{\sqrt{\sum_{i=1}^N (D_p^i - D_A^i)^2}}{\sqrt{\sum_{i=1}^N (D_A^i - D^-)^2}} \tag{7}$$

3.8.5 Correlation coefficient

The Correlation Coefficient (CC) is a calculation of how accurately the model replicates experimental results. It is defined as follows:

$$CC = \sqrt{1 - \frac{\sum_{i=1}^N (D_p^i - D_A^i)^2}{\sum_{i=1}^N (D^- - D_A^i)^2}} \tag{8}$$

4 Results and discussion

The previously mentioned algorithms were used to create seven different hybrid models for predicting groundwater levels. Tables 4 and 5 outline the features of all models

considered, as well as the values of the performance metrics for Rajshahi and Rangpur in the training and testing phases, respectively.

4.1 Training stage

Figures 5 and 6 show the comparisons, for the training stage, between the predicted and measured values of the groundwater level for the Rajshahi and the Rangpur stations, respectively. The comparison is reported in the form of a time series on the left, for the period January 1981–December 1989, characterized by marked fluctuation of the groundwater level, while on the right all the data are reported in the form of measured versus predicted values charts. Fluctuation of the GWL follows a clear seasonal trend. In particular, groundwater reaches its maximum depth, corresponding to a condition of groundwater scarcity, in April, as a consequence of the dry season that affects the area in the period from November to February. Mean GWLs, for the training stage period (January 1981–December 2008), equal to 8.88 m and 4.94 m were computed for the Rajshahi and the Rangpur stations, respectively. Otherwise, after the monsoon season, which leads heavy rains, groundwater reaches its lower depth, corresponding to a condition of greater groundwater availability, in September, with mean GWLs of 3.86 m and 2.12 m, respectively, for the Rajshahi and the Rangpur stations.

The best performances were obtained with the Bagging-RT model, which provided accurate forecasting for both Rajshahi (CC=0.96, RMSE=0.60 m, Fig. 5a and b) and Rangpur (CC=0.97, RMSE=0.26 m, Fig. 6a and b). Bagging-RF showed a slight performance decrease. However, the forecasts were still very good for both the stations (Rajshahi—CC=0.93, RMSE=0.76 m, Fig. 5c and d; Rangpur—CC=0.96, RMSE=0.31 m, Fig. 6c and d). Model Bagging-RepTree was less accurate than the latter but also exhibited good performance for Rangpur (CC=0.91, RMSE=0.47 m, Fig. 6e and f). However, it should be

noted that unlike the Bagging-RT and Bagging-RF models, Bagging-RepTree was unable to predict the marked fluctuation of the groundwater level measured for the Rangpur station in January 1986, with groundwater that reaches a depth of 7.01 m (Fig. 6e and f). For Rajshahi station, Bagging-RepTree model led to a constant overestimation of the GWL (Fig. 5e and f) with the metrics that reached a CC equal to 0.85, usually considered as the minimum values for a proper prediction [63]. The forecasting capabilities of Bagging-M5P, Bagging-Decision Stump, Bagging-LWLR and Bagging-SVM were very similar and below a minimum level of accuracy to provide reliable predictions with the worst performance for Rajshahi station (CC between 0.66 and 0.78, RMSE between 1.23 m and 1.78 m).

Overall, predictions were significantly affected by the chosen forecasting models, while Bagging-RT and Bagging-RF showed the best results in this study. Furthermore, worse performances were observed for Rajshahi. This should be related to a fairly constant lowering of the groundwater that affects the latter, passing from a mean GWL equal to 4.86 m in 1981 to 8.55 m in 2008, with a mean lowering equal to 0.14 m per year. Rangpur instead highlights an almost constant mean GWL during the training stage, passing from 3.28 m in 1981 to 3.51 m in 2008, with low fluctuation of different signs during the years.

The anomalous positive peak measured for the Rangpur station (Figs. 6a, c and e), nevertheless it took place in the dry season, seems to be related mainly to external factors, e.g., water pumping, as it did not show a seasonal component, being measured only for the year 1986 in the training stage period. However, Bagging-RT and, with less accuracy, Bagging-RF models were able also to provide a prediction of this singular event.

Table 4 Performance metrics for GWL prediction in the training stage

Hybrid models	Rajshahi					Rangpur				
	RMSE (m)	MAE (m)	RAE (%)	RRSE (%)	CC	RMSE (m)	MAE (m)	RAE (%)	RRSE (%)	CC
Bagging-M5P	1.23	1.00	60.64	62.61	0.78	0.60	0.46	50.04	56.21	0.83
Bagging-RF	0.76	0.59	36.08	38.75	0.93	0.31	0.23	25.51	28.94	0.96
Bagging-REPTree	1.06	0.85	51.32	53.74	0.85	0.47	0.35	38.13	43.39	0.91
Bagging-RT	0.60	0.45	27.47	30.79	0.96	0.26	0.18	19.87	24.17	0.97
Bagging-decision stump	1.58	1.31	79.42	80.18	0.60	0.79	0.64	70.59	73.85	0.68
Bagging-LWLR	1.49	1.24	74.82	75.83	0.66	0.71	0.55	60.83	66.42	0.75
Bagging-SVM	1.78	1.32	80.02	79.90	0.60	0.78	0.60	66.08	72.55	0.70

Table 5 Metrics for GWL prediction in the testing stage

Hybrid models	Rajshahi					Rangpur				
	RMSE (m)	MAE (m)	RAE (%)	RRSE (%)	CC	RMSE (m)	MAE (m)	RAE (%)	RRSE (%)	CC
Bagging-M5P	1.36	1.03	61.73	67.91	0.74	0.69	0.54	52.35	57.89	0.83
Bagging-RF	0.71	0.51	30.71	35.64	0.95	0.42	0.29	28.30	35.48	0.94
Bagging-RepTree	1.04	0.75	44.91	52.16	0.87	0.52	0.38	36.37	43.54	0.90
Bagging-RT	0.60	0.40	24.25	29.99	0.96	0.38	0.24	23.55	31.77	0.95
Bagging-decision stump	1.52	1.17	70.07	76.04	0.69	0.90	0.74	70.76	57.12	0.66
Bagging-LWLR	1.36	1.01	60.89	67.88	0.76	0.82	0.67	64.40	68.28	0.73
Bagging-SVM	1.58	1.22	73.58	78.86	0.63	0.93	0.77	74.36	77.26	0.64

4.2 Testing stage

Figures 7 and 8 provide a comparison between measured and predicted GWLs of the Rajshahi and the Rangpur stations, respectively, for the testing stage period: January 2009–December 2017. During this period, GWL reached its maximum during April, as for the training stage with, however, greater mean GWLs, computed for the testing stage period, equal to 11.92 m and 5.79 m for the Rajshahi and the Rangpur stations, respectively. This shows how in recent years the condition of water scarcity in the months following the dry season has become even more dramatic. Instead, groundwater reaches its lower depth during the month of September for Rajshahi, with mean GWL equal to 7.07 m, and October for Rangpur, with mean GWL equal to 2.34 m. In terms of testing, Bagging-RT produced very accurate results for both stations, demonstrating that it was the best performing hybrid model among seven models. In addition, there was no significant change in metrics between Rajshahi (CC=0.96, RMSE=0.60 m, Fig. 7a and b) and Rangpur (CC=0.95, RMSE=0.38 m, Fig. 8a and b). The Bagging-RF model also forecasts well in Rajshahi (CC=0.95, RMSE=0.71 m, Fig. 7c and d) and Rangpur (CC=0.94, RMSE=0.42 m, Fig. 8c and d). While both models are capable of accurately reconstructing the time-series pattern, Bagging-RT captures the positive and negative variations of the groundwater with greater precision than Bagging-RF. GWL fluctuation shows a marginal underestimation of positive peaks and an overestimation of negative peaks for the latter. This is particularly true for the Bagging-REPTree model, which performs worse than the Bagging-RF model in both Rajshahi (CC=0.87, RMSE=1.04 m, Fig. 7e and f) and Rangpur (CC=0.90, RMSE=0.52 m, Fig. 8e, f). In particular, Bagging-REPTree was unable to predict the marked fluctuation of the groundwater level measured for the *Rajshahi* station in July 2010, with GWL equal to 15.13 m (Fig. 7e, f). As stated in the training stage, this anomalous peak, which also happened monsoon

season, could be related to external factors. However, Bagging-RT and, partially, Bagging-RF models were able to provide a forecasting of this event. The forecasting capabilities of Bagging-M5P, Bagging-Decision Stump, Bagging-LWLR and Bagging-SVM exhibit a marked decrease in the performance with the worst performance for Rajshahi station (CC between 0.63 and 0.74, RMSE between 1.36 m and 1.58 m).

Overall, the best predictions for both stations and training and testing stages were obtained by means of the Bagging-RT model. It should be noted that, except for Bagging-RT and Bagging-RF that are the best performing models, the worst performances were observed for the Rajshahi. This result highlights how the less performing models are more affected by the lowering of the groundwater level in providing reliable forecasts. As for the testing stage, a fairly constant lowering of the groundwater was observed for Rajshahi, passing from a mean GWL equal to 8.55 m in 2008 to 9.88 m in 2017, with a mean lowering equal to 0.15 m per year. Rangpur instead highlights a more marked lowering of the groundwater compared to the training stage but in any case, of lesser entity compared to the Rajshahi, passing from 3.52 m in 2008 to 3.93 m in 2017, with a mean lowering equal to 0.05 m per year.

In order to further evaluate the prediction accuracy of the seven models, two Taylor diagrams [3] are reported in Figs. 9 and 10, respectively for the stations of Rajshahi and Rangpur. The advantage of the Taylor representation is that it allows to compare the similarity between measured and predicted values in one diagram [1]. The Bagging-RT model was closer to the target measured values in comparison with the other six models. The predictive results can be also evidenced by considering the high values of Correlation Coefficient (CC) computed for both Rajshahi and Rangpur stations. This result provides a further demonstration of how the Bagging-RT model can provide accurate groundwater level prediction.

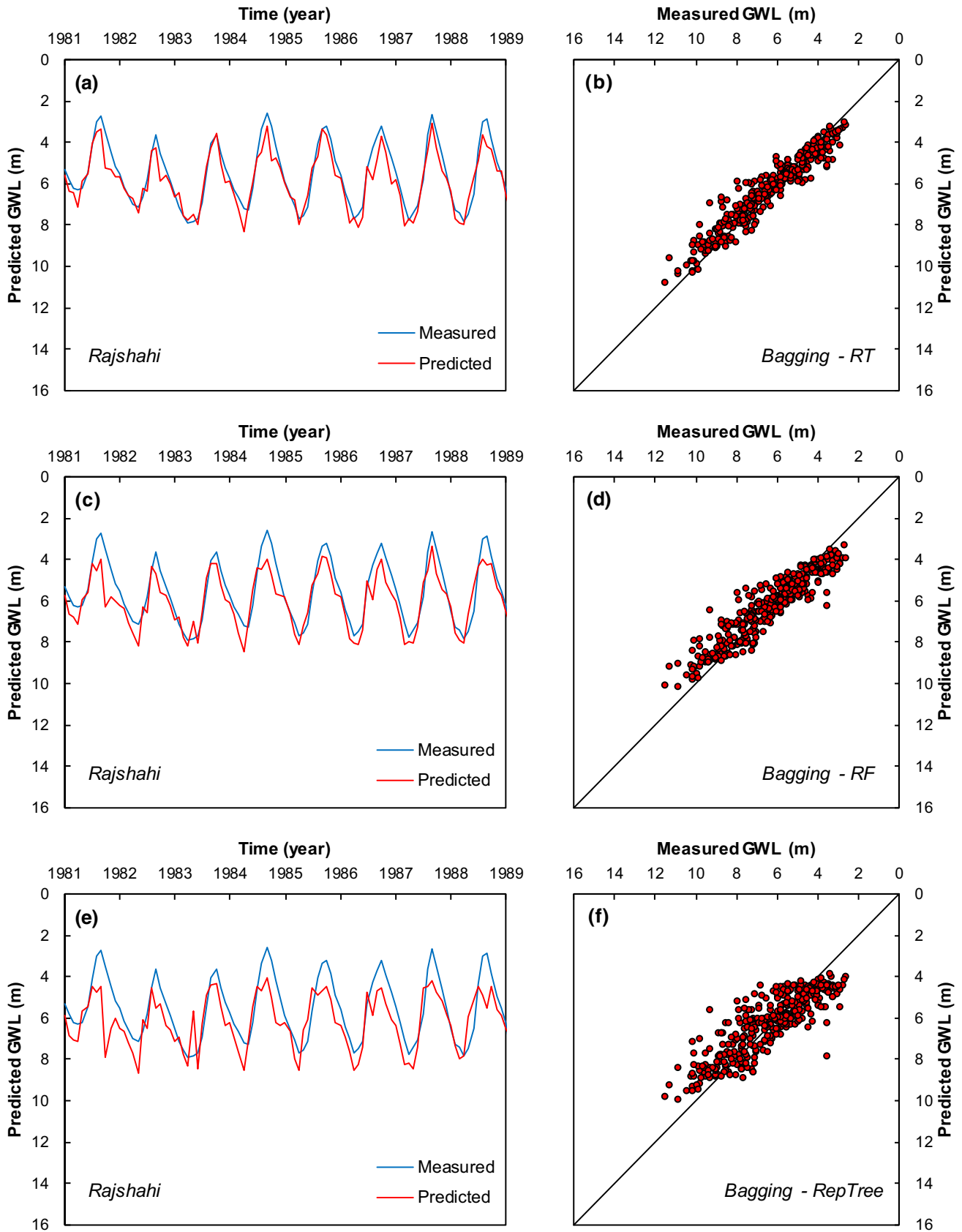


Fig. 5 Comparisons between measured and predicted values for Rajshahi—Training stage

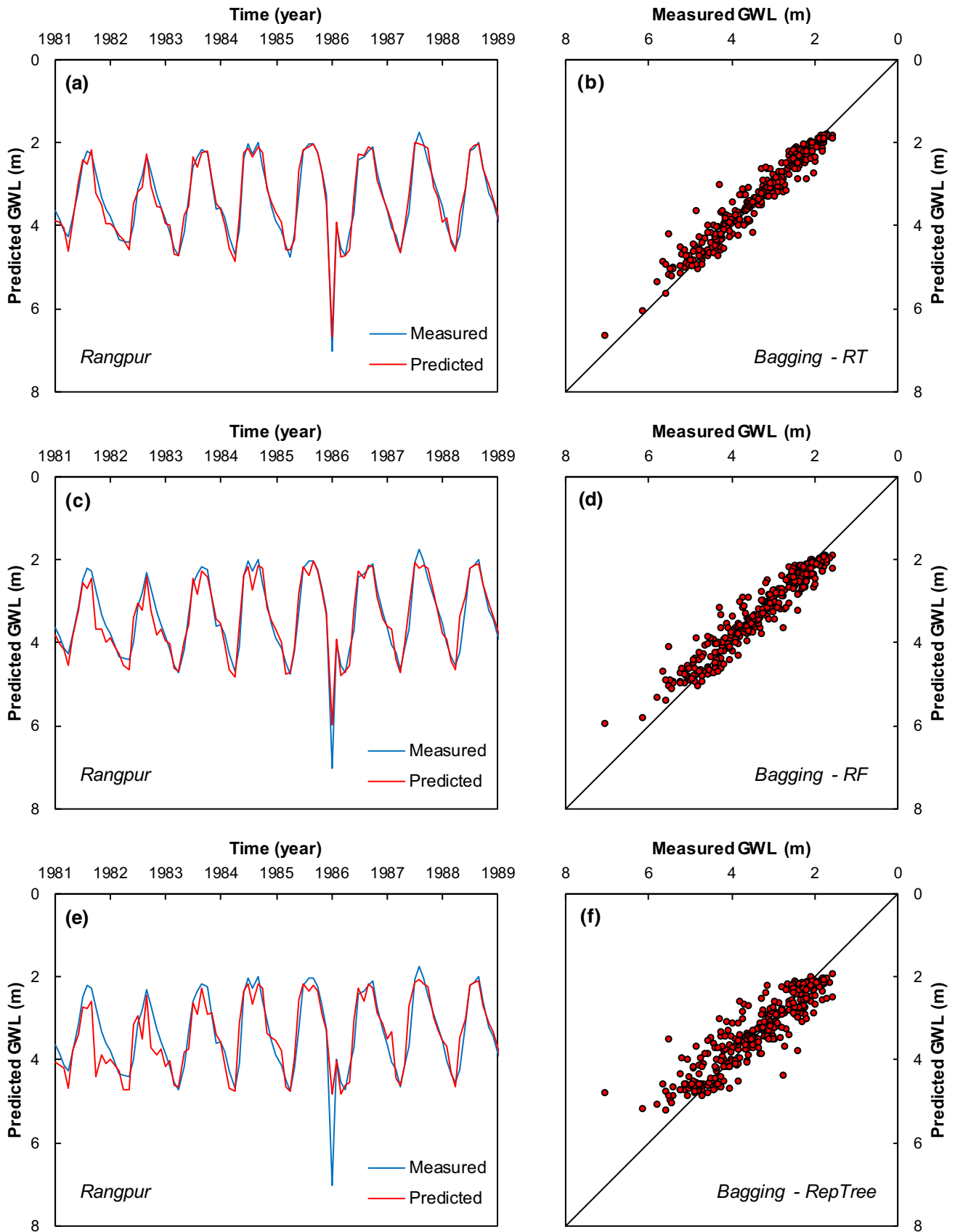


Fig. 6 Comparisons between measured and predicted values for Rangpur—Training stage

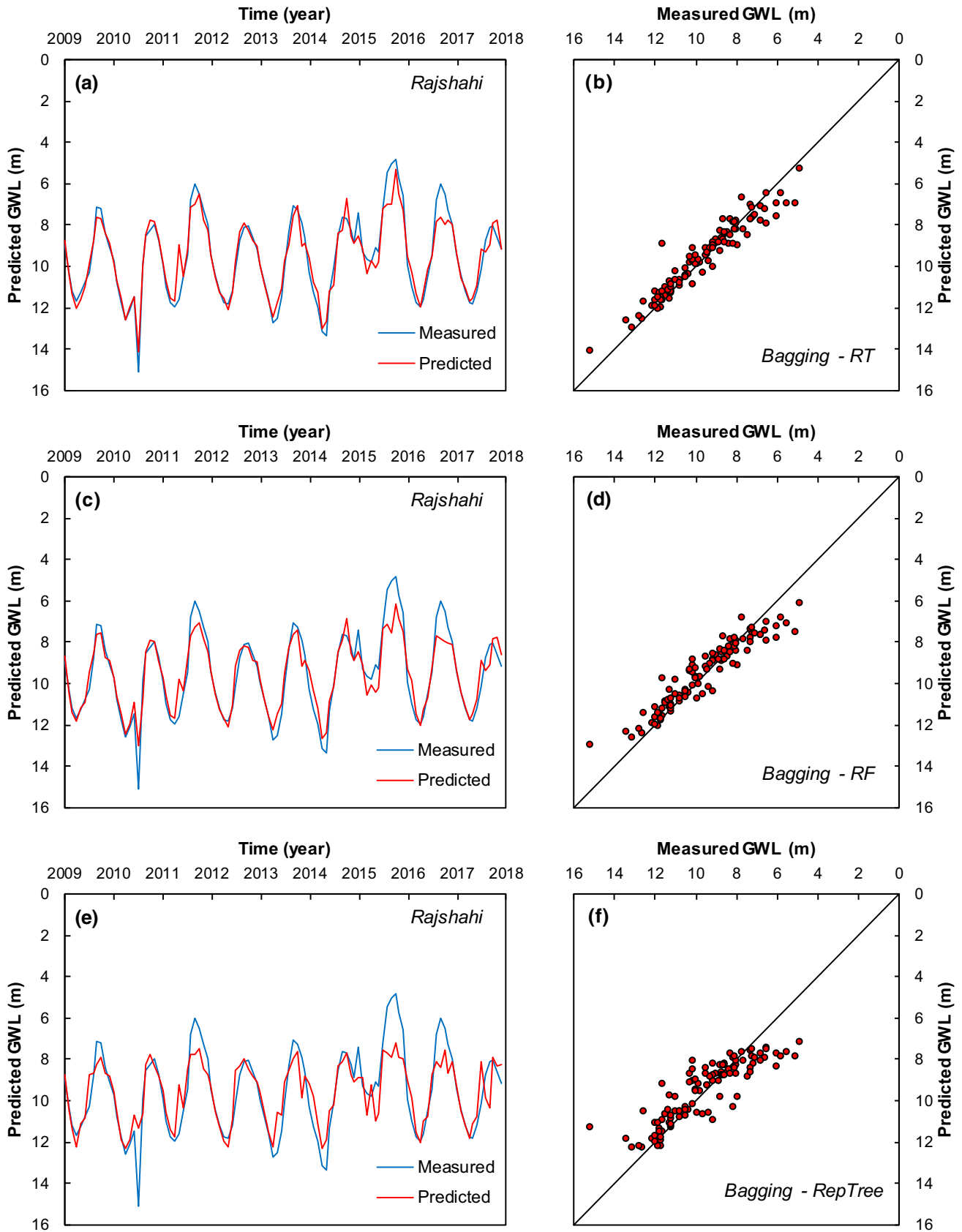


Fig. 7 Comparisons between measured and predicted values for Rajshahi—Testing stage

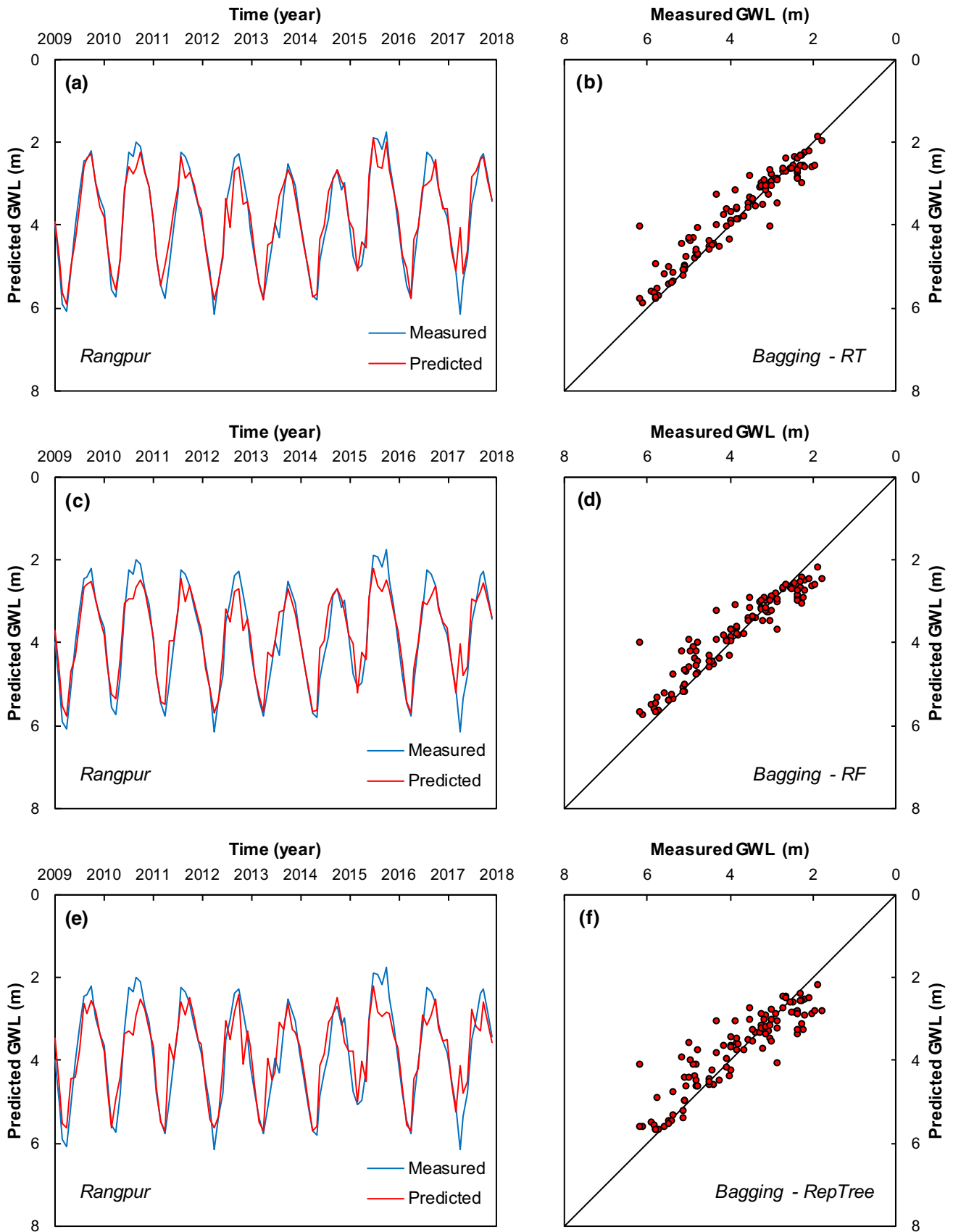


Fig. 8 Comparisons between measured and predicted values for Rangpur—Testing stage

4.3 Multicollinearity statistics and sensitivity analysis

Table 6 shows the tolerance and variable important factor for all inputs. Temperature showed high degree of tolerance while rainfall was characterized by the lowest degree. In contrast, rainfall was the highest effective variable in GWL prediction, followed by humidity. Moreover, Table 7 and Fig. 11 show the outcomes of a regression analysis to identify the most effective input parameters in predictive data-driven models. The results of the performed regression analysis on all input parameters proved that temperature and humidity, having highest absolute standard coefficients ($\beta=0.14$ and 0.47), were identified as the most influential input parameters, respectively, for simulation of GW3L.

5 Discussion

The comparative analysis of GWL prediction models indicate that all the considered models performed quiet well. In particular, hybrid ensemble models such as

Bagging-RT and Bagging-RF models showed the highest predictive performance than the other models. It is evident that a hybrid model will perform better than a standalone model [76]. Many researchers used a hybrid model and reported that the predictive performances of the hybrid ML models can outperform basic models [18, 52, 88, 104]. Likewise, in this research, the performance of the hybrid models, i.e., Bagging-RT and Bagging-RF are also better than the single Bagging classifier. Bagging ensemble efficiently decreases both ambiguity and biases in the method [81]. Generally, common statistical tools have over-fitting and bias issues. ML-based ensemble tools can simply overcome these issues [43, 46]. Importantly, our findings suggest that Bagging ensembles with RT and RF models can reflect complex nonlinear relationships between GWL and input parameters, but the lack of statistically significant testing with hybrid ensemble models limits quantitative hypothesis findings. Earlier research suggested that ML methods could be more useful in GWL studies than the traditional numerical model (Nourani and Moosavi, [68, 77, 103]. Therefore combining ML methods in hybrid models such as the Bagging-RT and Bagging-RF models, the accuracy of the Bagging model has improved, and the

Fig. 9 Taylor diagrams for the seven used models at Rajshahi station

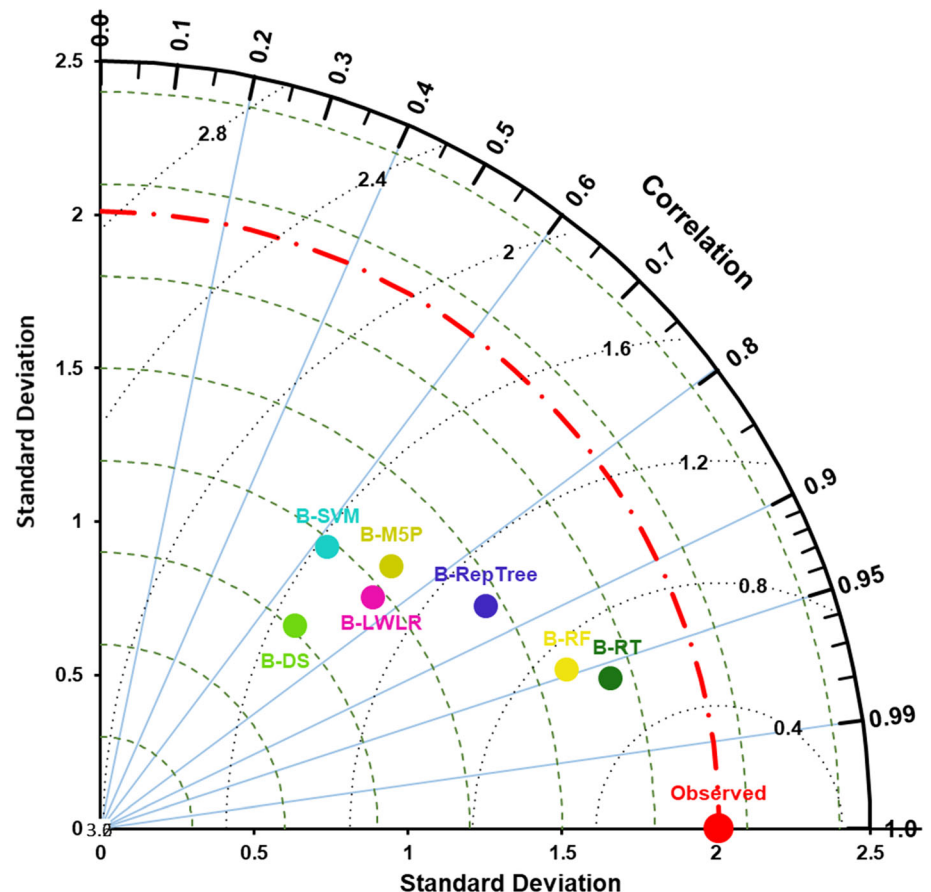


Fig. 10 Taylor diagrams for the seven used models at Rangpur station

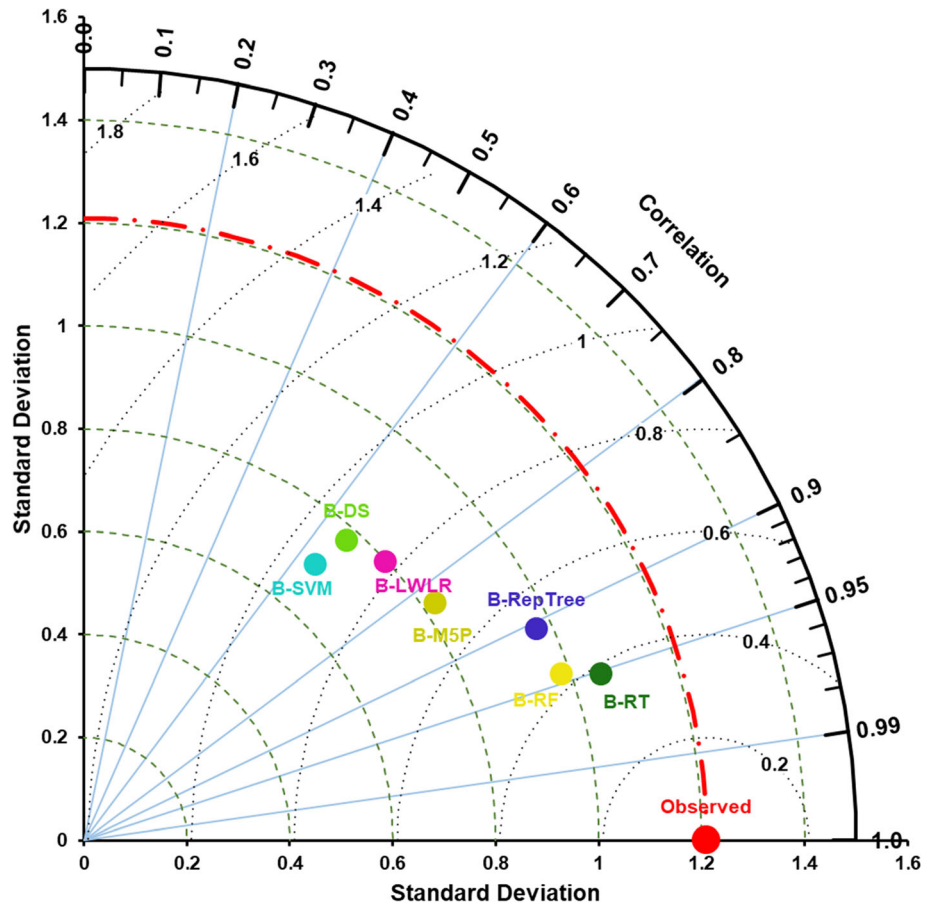


Table 6 Tolerance and variables important factor (VIF)

	Temp	Rainfall	Humidity
Tolerance	0.616	0.404	0.589
VIF	1.622	2.475	1.697

Bagging-RF model now has the highest forecasting precision. The Bagging-RF model has some advantages, e.g., high prediction precision, a smaller number of user-friendly parameters, and capability of escaping over-fitting issues [81]. Likewise, the RT method can deal with regression and classification issues. The main advantage of RT is that it can handle nonlinear heterogeneous events between the input and model output parameters as thus to achieve high model accuracy [60]. Similarly, Bagging-RT

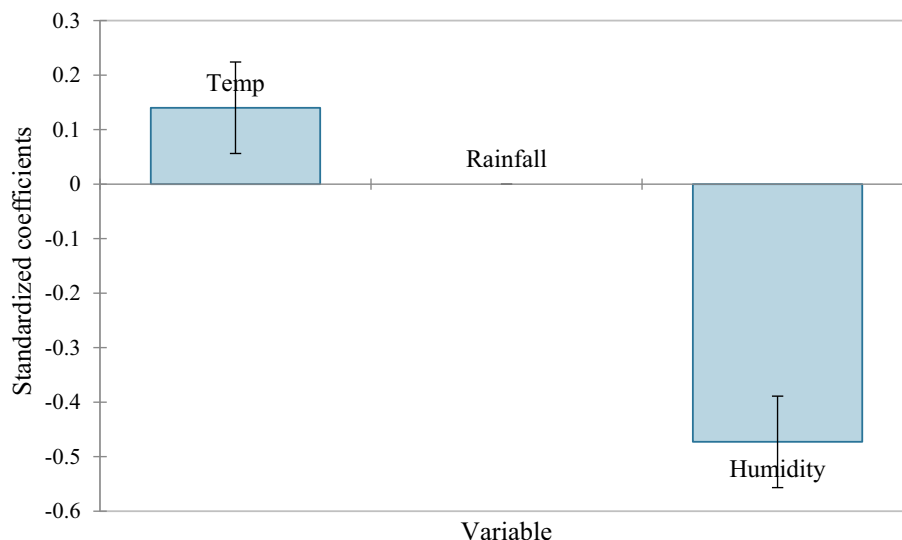
hybrid ensemble model aids addressing the presence of over-fitting issues compared to the other traditional models [56]. On the other hand, the M5P, decision stump, LWLR, and SVM models have required many hyper-parameters, which need to be carefully tuned for groundwater level modeling.

RT was identified as the most successful approach of the proposed seven hybrid methods used in this study, followed by RF, decision stump, M5P, SVM, and REP tree methods, implying that RT was the most capable tool in reducing the variance, biases, and noise of the GWL estimation. One of the key benefits of the Bagging-RT paradigm is that it employs bootstrap aggregation, which allows for the collection of acceptable input parameters and training dataset volume. This was warranted by the reasons that Bagging-RT and Bagging-RF models needed less time for calculation, with a minimum error, while other hybrid

Table 7 The regression analysis for identifying the most effective parameters

Source	Value	Standard error	<i>t</i>	Pr> <i>t</i>	Lower bound (95%)	Upper bound (95%)
Temp.	0.140	0.043	3.281	0.001	0.056	0.224
Rainfall	0.000	0.000				
Humidity	-0.473	0.043	-11.075	<0.0001	-0.557	-0.389

Fig. 11 The standardized coefficients of input variable for sensitivity analysis



models, for example, Bagging-M5P, Bagging-REPTree, Bagging-decision stump, Bagging-LWLR, and Bagging-SVM require a higher memory, maximum datasets, and more time for computation, followed by Rahman et al. [74] and Salam and Islam [81]. Besides, the disadvantage of traditional models is that these models require more time, larger datasets, and more input variables which are not appropriate for data-scarce drought-prone region, such as northwest Bangladesh [82]. Our findings are in line with Sharafati et al. [88] and Jajarmizadeh et al. [48], who stressed the importance of using nonlinear approaches for preprocessing input parameters for prediction. This research demonstrated that hybrid ensemble tools can proficiently minimize noise, variance, and over-fitting issues of input datasets, leading to an improved and more stable modeling [80, 100]. Our results, however, can vary because they are dependent on the sites and the range of GW input parameters. Following that, hybrid machine learning tools are capable and vital techniques for GWL prediction. The employed models can be used in other areas worldwide in similar hydro-climatic conditions for their better applicability. However, these GWL predictions will help policymakers for devising plans to lessen the GWL depletion and to manage GWR.

A myriad number of studies suggested that hybrid ML models have improved the performance of the base models in most cases. For instance, Avand et al. [8] showed that the Bagging coupled with the decision tree method improved the prediction ability of the GW potential zone. Chen et al. [19] demonstrated that the application of the J48 decision tree has better performance when coupled with Bagging and Dagging to identify GW potential zones in Wuqi County, China. In another research, Nguyen et al. [67] reported that Dagging and Bagging improved the performance of the logistic regression model for GW

potential mapping. Though the literature review has mainly stated the efficacy of hybrid ML methods, these methods exhibited various performances for various issues in diverse regions. For example, Pham et al. [69] showed that the RF-CDT (Rotational forest-Credal Decision Tree) ensemble model performed better coupled with CDT and Bagging, Dagging, and Decorate techniques for landslides prediction, while Roy and Shaha [79] reported that multi-layer perceptron neural network (MLP)-Dagging models performed better with MLP, MLP-Bagging for gully erosion prediction. Unlike others, findings have also been observed for flood susceptibility mapping using the hybrid ML models [43, 46, 69]. From previous studies, it can be concluded that the standalone and hybrid ML models are significantly case- and location-specific, and that their performances profoundly rely on the local environment that the training data are generated upon, implying that the use of various techniques in various areas should be sustained to get the best model for each hydro-climatic setting [81].

Though hybrid methods have led to better forecasting performance and generalization, the time-wasting variable tuning may limit their uses in other areas for various purposes. The input variables of the hybrid models generated in our research were physically adjusted based on a trial-and-error basis [67]. However, there are multiple alternative of executing variable optimization (e.g., gradient-based optimization, particle swarm optimization, mayfly optimization and Gray wolf optimization) that can significantly enhance the model-generating procedure. Nonetheless, the use of software like WEKA simplifies the development of these hybrid methods [39] and does not need extensive programming proficiency.

In recent decades, especially the drought-prone north-western region of Bangladesh has been experiencing the

problem of reduction in the GWL since the prevailing GW reserves are quickly being exploited due to the lack of balance between demand and supply of GW in this region.

The long-term GWL fluctuation is a function of many-fold parameters including annual rainfall, annual groundwater withdrawal, and local surface geological conditions. Precipitation has a great impact on GWL, being the primary source for the recharge of the groundwater [46]. The annual rainfall in the Rajshahi and Rangpur regions had never surpassed 1400 mm and 1800 mm in the last three decades, which is 45 and 25% less than the country's average of 2550 mm [110]. The decreasing precipitation trend, which is consistent with the GWL lowering in the study area, should be taken as a key factor [23]. In the present study region, the rainfall has reduced 10–55 mm/year over the past 12 years, which would possibly lead to a decrease of natural recharge [47], even more due to surface geology that characterizes the shallow alluvial aquifers in the region, which affects the response time of the aquifers and the groundwater recharge pathways [101]. In particular, in the layers of superficial clay of considerable thickness, which are characterized by very low hydraulic conductivity, the lowering effect of the groundwater storage is relevant [87].

A further factor that has led to a lowering of the GW table is the agricultural one. The abuse of existing GWL, which are mostly being extracted for paddy farming, rice mill operations and other industrial purposes, is an alarming issue that affects in the Rajshahi region. Groundwater provides 79% of the water needed for rice irrigation.

The application of the Bagging-RT hybrid model allows to get an accurate prediction of the GWL fluctuations, leading to a better management of the GW resources in the northwestern region as well as providing suitable techniques for devising operational strategies. If the integration of remote sensing tools and hybrid Bagging-RT is properly operated, it can assist water managers and other practitioners linked the management of the GW to get precise results. This has become enormously crucial due to the limited GW reserves and the enormous demand for GW in the agricultural field. Our study has used the time-series dataset for assessing the variations in GWL and cycles of these variations. This provides an understanding of the hydrological and hydro-geological patterns. Any oscillations happening in the natural patterns and any changes in the GWL-triggered due to external factors (e.g., over-exploitation of GW resources, and lack of water recharge) may be observed using these tools coupled with the data collected from Remote sensing and hybrid ML model.

Despite substantial advances in the GWL model using ML tools in the past decades [76, 80], there are yet several crucial problems that should be addressed to enhance reliability between ML tools and their theoretical and

physical-based complements [55]. Many studies have paid increasing attention to develop the hybrid ensemble methods for identifying various cycles (such as trends, periodicities, or level alterations) in time-series datasets (e.g., Rezaie-balf et al. 2017, [58, 74] or to produce multi-model ensemble tools [65] for enhancing model performance, mostly for long-term GWL prediction. Our study developed seven hybrid ensemble models for long-term prediction of GWL. The developed hybrid models showed high performance for most of the cases. However, model performance for the short-term and mid-term prediction can probably be further improved by integrating time-series datasets of GW recharge, discharge, pumping, and evapotranspiration. Coupling these related input parameters may enhance the model performance for the mid- and short-term prediction.

For this research, we only considered GWL, relative humidity, rainfall, and mean temperature as model input parameters, which are usually regarded as input parameters for GWL prediction using ML methods [76]. Besides, related studies commonly overlook integrating some vital parameters including bore well depths, and hydro-geological features, e.g., aquifer porosity and permeability, and closeness to sink or source tube wells. Further investigation is essential for thoroughly integrating these important hydro-geological parameters in ML tools to enhance their performance for short and mid-term prediction and to confirm these hybrid methods agree well with the theoretical and numerical-based methods. Such understanding could give further numerical insights and would probably lead to a higher acceptance of ML methods in the sustainable management of GWR.

6 Conclusion

Multiple groundwater level prediction models were built and compared in this study to explore potential knowledge of GWL fluctuations. GWL datasets from two observed wells in Bangladesh's northwestern area were collected and used to train and test the seven developed hybrid ML models for long-term prediction. The CC, MAE, RAE, RRSE, RMSE, and Taylor diagrams were used to compare output with seven ML models. The inclusion of Bagging as a base classifier significantly improved the efficiency of seven hybrid versions, according to the results. Based on their accuracy matrices, the Bagging-RT model outperformed the other hybrid models in GWL estimation. Thus, the Bagging-RT model provides (i) a potential application of historical GW datasets; (ii) accurate forecasting for GW level; and (iii) a potential way of exploring new knowledge for hydrologic field experts and practitioners, as well as other parts of the world in similar hydro-climatic settings.

Future research should focus on finding a solution for coupling hydrogeologic features knowledge with novel deep learning models in order to achieve more satisfactory results.

Authors' contributions QBP contributed to project administration, conceptualization, writing—original draft, formal analysis, visualization. AE contributed to software, formal analysis, writing—original draft, visualization. MK, FDN, FG, ARMTI, ST contributed to writing, review and editing. XCN, ANA, DTA contributed to supervision, writing, review, editing.

Funding No external funding.

Availability of data and materials The data that support the findings of this study are available from the first author, [Quoc Bao Pham, phambaoquoc@tdmu.edu.vn], upon reasonable request.

Declarations

Ethics approval Not applicable.

Consent for participate Not applicable.

Consent for publish Not applicable.

Conflict of interests This manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. There are no conflicts of interest to declare.

References

- Abba SI, Pham QB, Usman AG, Linh NTT, Aliyu DS, Nguyen Q, Bach Q-V (2020) Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *J Water Process Eng*. <https://doi.org/10.1016/j.jwpe.2019.101081>
- Adamala S, Srivastava A (2018) Comparative evaluation of daily evapotranspiration using artificial neural network and variable infiltration capacity models. *Agric Eng Int CIGR J J 20* (1):32–39
- Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and non-linear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour Res* 48(1):1–14
- Ajmera TK, Goyal MK (2012) Development of stage discharge rating curve using model tree and neural networks: an application to Peachtree Creek in Atlanta. *Expert Syst Appl* 39 (5):5702–5710
- Alizamir M, Kisi O, Zounemat-Kermani M (2018) Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrol Sci J* 63(1):63–73
- Amit Y, German D (1997) Shape Quantization and Recognition with Randomized Trees. *Neural Comput* 9(7):1545–1588
- Atkeson CG, Moore AW, Schaal S (1996) Locally weighted learning for control. In: Aha DW (ed) *Lazy learning*. Springer, Dordrecht
- Avand M, Janizadeh S, Tien Bui D, Pham VH, Ngo PTT, Nhu V-H (2020) A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int J Digital Earth* 13:1408–1429
- Beg AH, Islam MZ (2016) Advantages and limitations of genetic algorithms for clustering records. In: 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp. 2478–2483
- Bharti B, Ashish Pandey SK, Tripathi DK (2017) Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. *Hydrol Res* 48(6):1489–1507
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and Regression Trees*. CRC Press, Boca Raton, FL, USA
- Breiman L (1998) Arcing classifiers. *Ann Stat* 26(3):801–849
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Busch JR, Ferrari PA, Flesia AG, Fraiman R, Grynberg SP, Leonardi F (2009) Testing statistical hypothesis on random trees and applications to the protein classification problem. *J Appl Stat* 3:542–563
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol TIST* 2(3):1–27
- Chen Y, Xu W, Zuo J, Yang K (2019) The fire recognition algorithm using dynamic feature fusion and IV-SVM classifier. *Cluster Comput* 22(3):7665–7675
- Chen W, Zhao X, Tsangaratos P, Shahabi H, Ilia I, Xue W, Wang X, Ahmad BB (2020) Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *J Hydrol* 583:124602
- Collobert SB (2001) SVM Torch support vector machines for large-scale regression problems. *J Mach Learn Res* 1 (2001):143–160
- Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88(11):2783–2792
- Daud MNR, Corne DW (2007) Human readable rule induction in medical data mining: A survey of existing algorithms. WSEAS European Computing Conference, Athens, Greece
- Dey NC, Saha R, Parvez M, Bala SK, Islam AKMS, Paul JK et al (2017) Sustainability of groundwater use for irrigation of dry-season crops in northwest Bangladesh. *Groundw Sustain Dev* 4:66–77
- Di Nunno F, Granata F (2020) Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network. *Environ Res* 190:110062. <https://doi.org/10.1016/j.envres.2020.110062>
- Di Nunno F, Granata F, Gargano R, de Marinis G (2021) Forecasting of extreme storm tide events using NARX neural network-based models. *Atmosphere* 12(4):512. <https://doi.org/10.3390/atmos12040512>
- Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach Learn* 40(2):139–157
- Dietterich T, Kong EB (1995) Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Technical report <http://datam.i2r.atar.edu.sg/datasets/krbd/>
- Dumitru C, Maria V (2013) Advantages and disadvantages of using neural networks for predictions. *Ovidius University Annals, Economic Science Series*, pp. 13
- Elbeltagi A, Kumari N, Dharpure JK et al (2021) Prediction of combined terrestrial evapotranspiration index (Ctei) over large river basin based on machine learning approaches. *Water (Switzerland)* 13:1–18. <https://doi.org/10.3390/w13040547>

30. Fallah-Mehdipour E, Haddad OB, Mariño MA (2013) Prediction and simulation of monthly groundwater levels by genetic programming. *J Hydro-Environ Res* 7(4):253–260
31. Freund E, Rossmann J (1999) Projective virtual reality: Bridging the gap between virtual reality and robotics. *IEEE Trans Robot Autom* 15(3):411–422
32. Gang C, Shouhui W, Xiaobo X (2016) Review of spatio-temporal models for short-term traffic forecasting. In: 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE). IEEE, pp 8–12
33. Ghorbani MA, Zadeh HA, Isazadeh M, Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-015-5096-x>
34. Gong M, Bai Y, Qin J, Wang J, Yang P, Wang S (2020) Gradient boosting machine for predicting return temperature of district heating system: a case study for residential buildings in Tianjin. *J Build Eng* 27:100950
35. 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(1):375–391
36. Goyal MK, Ojha CSP (2011) Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree. *Water Resour Manag* 25(9):2177–2195
37. Granata F (2019) Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric Water Manag* 217:303–315. <https://doi.org/10.1016/j.agwat.2019.03.015>
38. Granata F, Gargano R, de Marinis G (2020) Artificial intelligence based approaches to evaluate actual evapotranspiration in wetlands. *Sci Total Environ* 703:135653. <https://doi.org/10.1016/j.scitotenv.2019.135653>
39. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newslett* 11(1):10
40. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501
41. Husna NE, Bari SH, Hussain MM, Ur-Rahman MT, Rahman M (2016) Ground water level prediction using artificial neural network. *Int J Hydrol Sci Technol* 6(4):371–381
42. Islam ARMT, Mehra B, Salam R, Siddik NA, Patwary MA (2020) Insight into farmers' agricultural adaptive strategy to climate change in northern Bangladesh. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-020-00681-6>
43. Islam ARMT, Talukdar S, Mahato S et al (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front* 12:101075. <https://doi.org/10.1016/j.gsf.2020.09.006>
44. Islam ARMT, Ahmed N, Bodrud-Doza M, Chu R (2017) Characterizing groundwater quality ranks for drinking purposes in Sylhet district, Bangladesh, using entropy method, spatial autocorrelation index, and geostatistics. *Environ Sci Pollut Res* 24(34):26350–26374
45. Islam MS, Islam ARMT, Rahman F, Ahmed F, Haque MN (2014) Geomorphology and land use mapping of northern part of Rangpur District, Bangladesh. *J Geosci Geomat* 2(4):145–150
46. Islam ARMT, Karim MR, Mondol MAH (2021) Appraising trends and forecasting of hydroclimatic variables in the north and northeast regions of Bangladesh. *Theoret Appl Climatol* 143(1–2):33–50. <https://doi.org/10.1007/s00704-020-03411-0>
47. Jahan CS, Mazumder QH, Islam ATMM, Adham MI (2010) Impact of irrigation in Barind area, NW Bangladesh—an evaluation based on the meteorological parameters and fluctuation trend in groundwater table. *J Geol Soc India* 76(2):134–142
48. Jajarmizadeh M, Lafdani EK, Harun S, Ahmadi A (2015) Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran. *KSCE J Civ Eng* 19:345–357
49. Joseph KS, Ravichandran T (2012) A comparative evaluation of software effort estimation using REPTree and K* in handling with missing values. *Aust J Basic Appl Sci* 6:312–317
50. Kalhor K, Emaminejad N (2019) Sustainable development in cities: Studying the relationship between groundwater level and urbanization using remote sensing data. *Groundw Sustain Dev* 9:100243
51. Kasiviswanathan KS, Saravanan S, Balamurugan M, Saravanan K (2016) Genetic programming based monthly groundwater level forecast models with uncertainty quantification. *Model Earth Syst Environ* 2:27
52. Khalil B, Broda S, Adamowski J, Ozga-Zielinski B, Donohoe A (2015) Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural network models. *Hydrogeol J* 23:121–141
53. Khatibi R, Nadiri AA (2021) Inclusive Multiple Models (IMM) for predicting groundwater levels and treating heterogeneity. *Geosci Front* 12:713–724. <https://doi.org/10.1016/j.gsf.2020.07.011>
54. Kisi O (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 528:312–320
55. Koch J, Berger H, Henriksen HJ, Sonnenborg TO (2019) Modelling of the shallow water table at high spatial resolution using random forests. *Hydrol Earth Syst Sci* 23(11):4603–4619. <https://doi.org/10.5194/hess-23-4603-2019>
56. Landwehr N, Hall M, Frank E (2005) Logistic model trees. *Mach Learn* 59(1–2):161–205
57. Malone BP, Minasny B, McBratney AB (2017) Using R for digital soil mapping, vol 35. Springer International Publishing, Cham, Switzerland
58. Malekzadeh M, Kardar S, Saeb K, Shabanlou S, Taghavi L (2019) A novel approach for prediction of monthly ground water level using a hybrid wavelet and nontuned self-adaptive machine learning model. *Water Resour Manag* 33:1609–1628. <https://doi.org/10.1007/s11269-019-2193-8>
59. Mirarabi A, Nassery HR, Nakhaei M, Adamowski J, Akbarzadeh AH, Alijani F (2019) Evaluation of data-driven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems. *Environ Earth Sci* 78(15):489. <https://doi.org/10.1007/s12665-019-8474-y>
60. Mishra AK, Ratha BK (2016) Study of random tree and random forest data mining algorithms for microarray data analysis. *Int J Adv Electr Comput Eng* 3(4):5–7
61. Mohanty S, Jha MK, Raul SK, Panda RK, Sudheer KP (2015) Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. *Water Resour Manag* 29(15):5521–5532
62. MPO (Master Plan Organization) (1987) Groundwater resources of Bangladesh, Technical Report no 5. (Dhaka: Master Plan Organization) Hazra, USA; Sir M MacDonald, UK; Meta, USA; EPC, Bangladesh
63. Moore DS, Notz WI, Flinger MA (2018) The basic practice of statistics. W.H Freeman and Company, New York
64. Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N (2013) A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. *Water Resour Manag* 27(5):1301–1321

65. Nadiri AA, Naderi K, Khatibi R, Gharekhani M (2019) Modelling groundwater level variations by learning from multiple models using fuzzy logic. *Hydrol Sci J* 64(2):210–226
66. Najock D, Heyde CO (1982) The number of terminal vertices in certain random trees with an application to stemma construction in philology. *J Appl Probab* 19:675–680
67. Nguyen PT, Ha DH, Avand M et al (2020) Soft computing ensemble models based on logistic regression for groundwater potential mapping. *Appl Sci* 10:2469. <https://doi.org/10.3390/app10072469>
68. Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet – artificial Intelligence models in hydrology: a review. *J Hydrol* 514:358–377
69. Pham BT, Phong TV, Nguyen-Thoi T, Parial KK, Singh S, Ly H-B, Nguyen KT, Ho LS, Le HV, Prakash I (2020) Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int.* <https://doi.org/10.1080/10106049.2020.1737972>
70. Platt JC (1999) Using analytic QP and sparseness to speed training of support vector machines. *Adv Neural Inf Process Syst* 11:557–563
71. Quinlan JR (1992) Learning with continuous classes. *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, pp. 343–348
72. Raghavendra N, Deka PC (2015) Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. *Cogent Eng* 2:999414
73. Raghavendra NS, Deka PC (2014) Forecasting monthly groundwater table fluctuations in coastal aquifers using support vector regression. In: *International Multi Conference on innovations in engineering and technology (IMCIET-2014)* (61–69). Elsevier Science and Technology, Bangalore
74. Rahman ARMS, Hosono T, Quilty JM, Das J, Basak A (2020) Multiscale groundwater level forecasting: coupling new machine learning approaches with wavelet transforms. *Adv Water Resour* 141:103595
75. Rahman MS, Islam ARMT (2019) Are precipitation concentration and intensity changing in Bangladesh overtimes? Analysis of the possible causes of changes in precipitation systems. *Sci Total Environ* 690:370–387
76. Rajae T, Ebrahimi H, Nourani V (2019) A review of the artificial intelligence methods in groundwater level modeling. *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2018.12.037>
77. Rezaie-balf M, Naganna SR, Ghaemi A, Deka PC (2017) Wavelet coupled MARS and M5 model tree approaches for groundwater level forecasting. *J Hydrol* 553:356–373. <https://doi.org/10.1016/j.jhydrol.2017.08.006>
78. Rodriguez JJ, Kuncheva LI, Carlos J (2006) Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 28(10):1619–1630
79. Roy J, Saha S (2021) Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India. *Adv Space Res* 67:316–333
80. Sahoo S, Russo TA, 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
81. Salam R, Islam ARMT (2020) Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2020.125241>
82. Salam R, Islam ARMT, Islam S (2020) Spatiotemporal distribution and prediction of groundwater level linked to ENSO teleconnection indices in the northwestern region of Bangladesh. *Environ Dev Sustain* 22(5):4509–4535. <https://doi.org/10.1007/s10668-019-00395-4>
83. Salih SQ, Sharafati A, Khosravi K, Faris H, Kisi O, Tao H, Ali M, Yaseen ZM (2019) River suspended sediment load prediction based on river discharge information: application of newly developed data mining models. *Hydrol Sci J* 65(4):624–637
84. Senthil Kumar AR, Ojha CSP, Goyal MK, Singh RD, Swamee PK (2012) Modelling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic and decision tree algorithms. *J Hydrol Eng.* [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000445](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000445)
85. Seyam M, Othman F, El-Shafie A (2017) Prediction of stream flow in humid tropical rivers by support vector machines. In: *MATEC Web of Conferences*, vol 111. EDP Sciences, p 01007.
86. Shahid S, Hazarika MK (2010) Groundwater drought in the northwestern districts of Bangladesh. *Water Resour Manag* 24:1989–2006
87. Shamsudduha M, Taylor RG, Ahmed KM, Zahid A (2011) The impact of intensive groundwater abstraction on recharge to a shallow regional aquifer system: evidence from Bangladesh. *Hydrogeol J* 19:901–916
88. Sharafati A, Asadollah SBHS, Neshat A (2020) A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *J Hydrol* 591:125468
89. Sheikh Khozani Z, Bonakdari H, Zaji AH (2018) Estimating shear stress in a rectangular channel with rough boundaries using an optimized SVM method. *Neural Comput Appl* 30:1–13. <https://doi.org/10.1007/s00521-016-2792-8>
90. Shiri J, Kisi O, Yoon H, Lee KK, Nazemi AH (2013) Predicting groundwater level fluctuations with meteorological effect implications – a comparative study among soft computing techniques. *Comput Geosci* 56:32–44
91. Song Y, Zhou H, Wang P, Yang M (2019) Prediction of clathrate hydrate phase equilibria using gradient boosted regression trees and deep neural networks. *J Chem Thermodyn* 135:86–96
92. Suryanarayana C, Sudheer C, Mahammood V, Panigrahi BK (2014) An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. *Neurocomputing* 145:324–335
93. Torgo L (1997). Functional models for regression tree leaves. In: *Machine learning*, Proceedings of the 14th International Conference (D. Fisher, ed.). Morgan Kaufmann, pp. 385–393.
94. Vapnik VN (1995) *The nature of statistical learning theory*. Springer-Verlag, New York Inc
95. Vapnik VN (1998) *Statistical learning theory*. Wiley
96. Vapnik VN, Golwicz S, Smola AJ (1997) Support vector method for function approximation, regression estimation and signal processing. In: *Mozer M, Jordan M, Petsche T (eds) Advances in Neural Information Processing Systems*, 9. MIT Press, Cambridge, MA, USA, pp 281–287
97. Verbyla DL (1987) Classification trees: a new discrimination tool. *Can J For Res* 17(9):1150–1152
98. Witten IH, Frank E (2000) *Data mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann, San Francisco, CA
99. Witten IH, Frank E, Trigg L, Hall M, Holmes G, Cunningham SJ (1999) *Weka: practical machine learning tools and techniques with Java implementations*. *Emerging Knowledge Engineering and Connectionist-Based Info. Systems*, pp. 192–196
100. Wöhling T, Burbery L (2020) Eigenmodels to forecast groundwater levels in unconfined river-fed aquifers during flow recession. *Sci Total Environ* 747:141220
101. WARPO (Water Resources Planning Organization) (2000) *National Water Management Plan. Volume 2: Main Report*; Water Resources Planning Organization, Ministry of Water Resources: Dhaka, Bangladesh, 2000

102. Yadav B, Ch S, Mathur S, Adamowski J (2017) Assessing the suitability of extreme learning machines (ELM) for groundwater level prediction. *J Water Land Dev* 32(1):103–112
103. Yadav B, Gupta PK, Patidar N, Himanshu SK (2019) Ensemble modelling framework for groundwater level prediction in urban areas of India. *Sci Total Environ* 712:135539
104. Yadav B, Gupta PK, Patidar N, Himanshu SK (2020) Ensemble modelling framework for groundwater level prediction in urban areas of India. *Sci Total Environ* 712:135539
105. Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (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
106. Yosefvand F, Shabanlou S (2020) Forecasting of groundwater level using ensemble hybrid wavelet–self-adaptive extreme learning machine-based models. *Nat Resour Res*. <https://doi.org/10.1007/s11053-020-09642-2>
107. Yu PS, Yang TC, Chen SY, Kuo CM, Tseng HW (2017) Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J Hydrol* 552:92–104
108. Zannat F, Islam ARMT, Rahman MA (2019) Spatiotemporal variability of rainfall linked to ground water level under changing climate in northwestern region, Bangladesh. *Eur J Geosci EURAASS* 1(1):35–58
109. 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
110. Zinat MRM, Salam R, Badhan MA, Islam ARMT (2020) Appraising drought hazard during Boro rice growing period in western Bangladesh. *Int J Biometeorol* 64(10):1697–1697

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Quoc Bao Pham^{1,2} · Manish Kumar³ · Fabio Di Nunno⁴ · Ahmed Elbeltagi⁵ · Francesco Granata⁴ · Abu Reza Md. Towfiqul Islam⁶ · Swapan Talukdar⁷ · X. Cuong Nguyen^{8,9} · Ali Najah Ahmed¹⁰ · Duong Tran Anh¹¹

✉ Duong Tran Anh
ta.duong@hutech.edu.vn

¹ Institute of Applied Technology, Thu Dau Mot University, Binh Duong Province, Vietnam

² Faculty of Natural Sciences, Institute of Earth Sciences, University of Silesia, Będzińska Street 60, 41-200 Katowice, Sosnowiec, Poland

³ Department of Soil and Water Conservation Engineering, G. B. Pant University of Agriculture and Technology, Pantnagar 263145, India

⁴ Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, DICEM, Via Di Biasio, 43, 03043 Cassino, Frosinone, Italy

⁵ Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

⁶ Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

⁷ Department of Geography, Faculty of Natural Science, Jamia Millia Islamia, New Delhi, Delhi, India

⁸ Center for Advanced Chemistry, Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

⁹ Faculty of Environmental and Chemical Engineering, Duy Tan University, Da Nang 550000, Vietnam

¹⁰ Institute of Energy Infrastructure (IEI), Civil Engineering Department, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, 43000 Selangor, Malaysia

¹¹ HUTECH University, 475A, Dien Bien Phu, Ward 25, Binh Thanh District, Ho Chi Minh City, Vietnam