Earth Science Informatics (2020) 13:1237–1250
https://doi.org/10.1007/s12145-020-00508-y

RESEARCH ARTICLE

Predicting groundwater depth fluctuations using deep learning, extreme learning machine and Gaussian process: a comparative study

Deepak Kumar[1] · Thendiyath Roshni[1] · Anshuman Singh[1] · Madan Kumar Jha[2] · Pijush Samui[1]

Received: 7 April 2020 / Accepted: 19 August 2020 / Published online: 23 August 2020

Abstract

Groundwater depth has complex non-linear relationships with climate, groundwater extraction, and surface water flows. To understand the importance of each predictor and predictand (groundwater depth), different artificial intelligence (AI) techniques have been used. In this research, we have proposed a Deep Learning (DL) model to predict groundwater depths. The DL model is an extension of the conventional neural network with multiple layers having non-linear activation function. The feasibility of the DL model is assessed with well-established framework models [Extreme Learning Machine (ELM) and Gaussian Process Regression (GPR)]. The area selected for this study is Konan basin located in the Kochi Prefecture of Japan. The hydro-meteorological and groundwater data used are precipitation, river stage, temperature, recharge and groundwater depth. Identical set of inputs and outputs of all the selected stations were used to train and validate the models. The predictive accuracy of the DL, ELM and GPR models has been assessed considering suitable goodness-of-fit criteria. During training period, the DL model has a very good agreement with the observed data ($RMSE = 0.04$, $r = 0.99$ and $NSE = 0.98$) and during validation period, its performance is satisfactory ($RMSE = 0.08$, $r = 0.95$ and $NSE = 0.87$). To check practicality and generalization ability of the DL model, it was re-validated at three different stations (E2, E3 and E6) of the same unconfined aquifer. The significant prediction capability and generalization ability makes the proposed DL model more reliable and robust. Based on the finding of this research, the DL model is an intelligent tool for predicting groundwater depths. Such advanced AI technique can save resources and labor conventionally employed to estimate various features of complex groundwater systems.

Keywords Groundwater-fluctuation modeling · Extreme learning machine · Deep learning · Gaussian process regression · Model evaluation

Communicated by: H. Babaie

✉ Thendiyath Roshni
roshni@nitp.ac.in

Deepak Kumar
decage007@gmail.com

Anshuman Singh
asingh@nitp.ac.in

Madan Kumar Jha
madan@agfe.iitkgp.ac.in

Pijush Samui
pijush@nitp.ac.in

[1] Department of Civil Engineering, National Institute of Technology Patna, Ashok Rajpath, Bihar 800005, India

[2] Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, West Bengal 721302, India

Introduction

From an ecological viewpoint, groundwater plays a crucial role in maintaining the health of the ecosystem (Fan et al. 2008). In arid and semi-arid region, groundwater depth influences the pattern of vegetation and soil health and result in soil salinization and land desertification. Mining of groundwater in arid and semi-arid regions lead to a low yield of aquifer potential resulting in groundwater decline and higher pumping cost. The groundwater in such regions is also affected by anthropogenic activities, which leads to groundwater contamination in the form of elevated level of arsenic and salinity. Therefore, the precise prediction of groundwater depth is necessary for holistic groundwater management and risk assessment.

In a developed country like Japan, the groundwater consumption rate has increased significantly that lead to gradual decline in groundwater and low yield aquifer. This resulted in groundwater contamination problem such as seawater intrusion and land subsidence (Jha et al. 1999). Hence, to overcome such aforementioned problems, proper spatial and temporal variation of groundwater depth is required especially in agricultural areas and natural reserves in the perspective of planning and management of groundwater resources (Bierkens 1998). The groundwater is a complex hidden resource, which depends upon earth properties below and hydro-meteorological characteristics. Groundwater system consists of different hydrological and geological properties. Various physical models are available to study the groundwater depth and to simulate the groundwater fluctuations. The basic framework of physical model is that it consists of governing equations. These governing equations simplify the physical phenomena to predict the different groundwater features based on different boundary conditions (Wang et al. 2011). Though physical models are viable tools, these models have limitations in terms of complexity and require large number of predictors, which are always not possible to collect. In such circumstances, when different predictors of data are limited and prediction is crucial than understanding the underlying causative mechanism, artificial intelligence (AI) techniques are generally used (Tapoglou et al. 2014; Jha and Sahoo 2015). The AI methods with appropriate tools may handle noisy data evolving from dynamic and complex nonlinear systems.

In recent times, applications of deep learning show groundbreaking advances in many fields, such as image recognition (Ji et al. 2012), forecasts in weather, traffic and other aspects (Lv et al. 2014), natural language processing (Hirschberg and Manning 2015), recommendation systems (Elkahky et al. 2015), speech recognition (Amodei et al. 2016), climate forecasting (Scher 2018), flood and typhoon forecasting (Jiang et al. 2018) and forest cover projection (Ye et al. 2019). Moreover, applications of deep learning are found in various real-world engineering problems relating to multi magnitude data e.g. bearing capacity of concrete-steel columns (Zhang et al. 2019), low-flow hydrological time series forecasting (Sahoo et al. 2019), forecasting monthly rainfall forecasting using sequential modeling (Kumar et al. 2019) and determination of soil liquefaction (Njock et al. 2020).

Furthermore, Krishna et al. (2008) has investigated ANN and feed forward neural network with Levenberg Marquardt models prediction of groundwater levels in coastal aquifer in Andhra Pradesh. Nourani et al. (2008) employed artificial neural network (ANN) methodology for estimating the groundwater levels in some piezometers placed in an aquifer in North Western Iran. Shiri et al. (2013) investigated the abilities of Gene Expression Programming (GEP), Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Networks (ANN) and Support Vector Machine (SVM) techniques for groundwater level forecasting. Roshni et al. (2020) have forecasting groundwater-level using Emotional neural network (EANN) and the results are compared with Generalised Regression Neural Network (GRNN) and the conventional Feedforward Neural Network (FFNN) model in coastal aquifer. In recent times, different AI techniques coupled or hybrid combination as wavelet analysis (WA), gained more attention over conventional machine learning techniques in order to obtain better accuracy (Suryanarayana et al. 2014; Yu et al. 2018). The conventional AI technique may not perform well for the non-stationary predictors and often require pre-processing of input variables (Cannas et al. 2006). The wavelet-based nonlinear Volterra model was investigated for predicting the long-term groundwater level by Maheswaran and Khosa (2013). Furthermore, a wavelet-adaptive neuro-fuzzy inference (WA-ANFIS) model was also compared with WA artificial neural network (WA-ANN) and ANFIS to predict the groundwater depth (Moosavi et al. 2013). In this sense, GPR is closely related to generalized least squares, which has been used extensively for regression analysis in hydrology (e.g., Sun et al. 2014; Raghavendra and Deka 2016; Karbasi 2018). Alizamir et al. (2018) has investigated the long-term groundwater variability using extreme learning machine (ELM).

Review of recent literature on Deep Learning (DL) in various fields and its promising results motivated the authors to explore its efficacy and to apply it as an advanced artificial intelligence (AI) technique to obtain greater prediction accuracy. In this study, deep learning model was proposed to predict the groundwater depth in the Konan basin of Japan. To compare the practicality and generalization ability of deep learning, two different framework models ELM and GPR were used to compare the prediction accuracy of groundwater depth. The basic framework of ELM is a single layered feed forward neural network (SLFNs), whereas the GPR model works on the probabilistic framework that uses Lazy learning and measure the similarity between points through the kernel function to predict the value for an unseen data. The prediction of this model is not just an estimate for that point, but also has uncertainty associated with it. The novelty of this work is to develop a generalized regional based deep learning concept for the unconfined aquifer through testing on other stations in the central aquifer of Konan basin.

Theoretical background

Deep learning

Deep Learning (DL) is a subfield of machine learning, which works on feature engineering via successive layers of meaningful representation (Chollet 2017). Deep learning is a

framework with a set of learning algorithms developed for deep structured neural networks (including but not limited to: feed forward neural networks with multiple hidden layers and recurrent neural networks). The layers contributing to the model is called the depth of the model. Deep learning also called as layered representation learning or hierarchical representation learning. The deep learning actually does not have a concrete formulation as other machine learning model. Similar to conventional neural network, the basic framework of deep learning is backpropagation method. Deep learning attempts to learn high-level features in data by using structures composed of multiple non-linear transformations through representation learning and overcome the basic problems of data pre-processing, overfitting and computational load (Schmidhuber 2015). In addition, its layers act as a multistage information distillation, where information is purified over successive filtering operations (Chollet 2017).

In recent time, deep learning has wide-spread attention, mainly by outperforming alternative machine learning methods such as kernel machines for image classification (Schölkopf et al. 1998). At this point, supervised learning is about mapping inputs and targets, which is basically done by multiple exposures of input and target, in order to minimize the loss of target and the predicted value. The basic difference between conventional supervised learning and deep learning is that it consists of deep sequence of simple data transformation and learns through exposure to data transformation. Figure 1 shows how the learning is achieved.

During the learning process, firstly data get exposed through layers in which transformation carried out by layer and henceforth the layers are parametrized by its weights and this weight is termed as parameters of the layers. Furthermore, it finds the sets of value for the weights of all layers in the network in order to correctly map inputs to the associated targets via the objective function (loss function). The objective function assigns the distance score between predicted values and true targets, and this assigned score work as a feedback signal in order to adjust the value of weights of the layers through optimizer for the next epoch, which results in lower the loss score (Chollet 2017). The optimizers serve as backpropagation algorithm and called as heart algorithm in the deep learning. The optimizers continue update network weights through backpropagation based on the loss score during the learning process, to minimize loss between the predicted and target values. This process iterates continuously until meet termination criteria.

Extreme learning machine

Extreme learning machine (ELM) is a state-of-art approach developed by G.-B. Huang et al. (2004). ELM is a single layer feedforward neural network (SLFNs). It has extremely fast learning time capability and reaches smallest training error with smallest norm of weight makes ELM most popular among ML models (G.-B. Huang et al. 2004). Figure 2 shows the brief topological structure of ELM having an input layer, feature optimization space and an output layer. Interested readers can find more detail in G.-B. Huang et al. (2004, 2015).

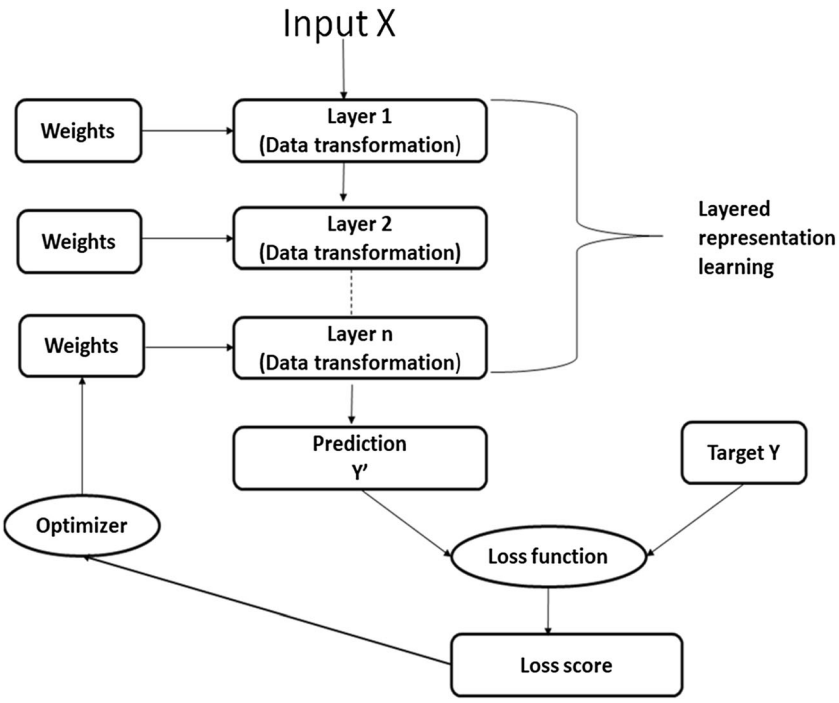

Fig. 1 The simple architecture of multi-layer deep learning (Karbasi 2018)

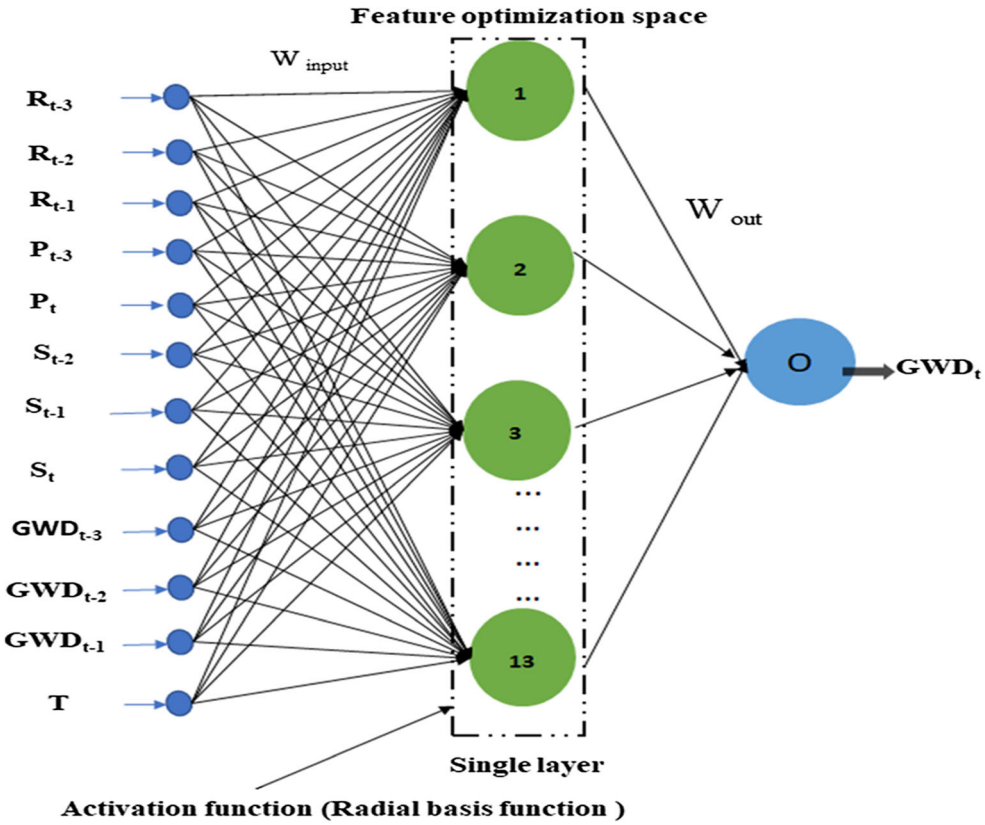

Fig. 2 Simple architecture of extreme Learning machine

The ELM model works on the principle of interpolation capability considering universal approximation (Huang et al. 2015). The Eq. (1) represents basic formulation of ELM model consists of the N arbitrary distinct input samples$(x_l, y_l) \in R^n \times R^n$, with standard single layer feedforward networks consisting M hidden nodes and an activation $\phi(.)$ function.

$$\sum_{j=1}^{M} \beta_j \phi\left(\mathrm{x}_l, \mathrm{b}_j, \mathrm{w}_j\right) = y_l \qquad l = 1, 2, 3, \ldots, N \tag{1}$$

where,$b_i \in R$is the random bias assign to each j^{th}hidden neurons and $w_j \in R$ is their respective weight vector to the hidden neuron and input node. Whereas, $\beta_j$denotes j^{th}hidden node weight vector connection to input of the hidden node to the output node with the input sample x_l. The results of hidden node produce $\phi(\mathrm{x}_l, \mathrm{b}_j, \mathrm{w}_j)$ as the output from each hidden node. ELM network randomly assigns each input to the hidden nodes. Hence Eq. (2) can be rewritten as Eqs. (3 and 4).

$$H\beta = Y \tag{2}$$

where,

$$H = \begin{bmatrix} \phi(\mathrm{x}_1,,,,\mathrm{b}_1,,,,\mathrm{w}_1)\ldots\phi(\mathrm{x}_1,,,,\mathrm{b}_M,,,,\mathrm{w}_M) \\ \vdots \quad \vdots \\ \phi(\mathrm{x}_N,,,,\mathrm{b}_1,,,,\mathrm{w}_1)\ldots\phi(\mathrm{x}_N,,,,\mathrm{b}_M,,,,\mathrm{w}_M) \end{bmatrix}_{N\times M} \tag{3}$$

The output is estimated through the least square solutions by the stated linear system, which is calculated using Eq. (4). Thereafter, the final output (Y) can be calculated using Eq. (2).

$$\begin{aligned} H\beta &= \left(\beta_1{}^T \beta_2{}^T \cdots \beta^T{}_L\right)^T_{m\times M} \\ \beta &= H^{\dagger} Y \end{aligned} \tag{4}$$

Where, $\mathrm{H}^{\dagger}$represents the inverse of the matrix (H) considering Moore–Penrose generalization. In this process of Moore–Penrose generalization, hidden biases are selected randomly and the weights output are analytically determined.

Gaussian process regression (GPR)

GPR is a probabilistic regression framework transfers over function, that describes the random variables either scalar or vectors or both in the form of Gaussian distribution (Rasmussen 2004). GPR prediction is quantified by its mean and covariance functions of inputs features: $x \sim \mathrm{N}(\mu, \sigma^2)$assuming that $D = \{(x_i, y_i) | \mathrm{i} = 1 | 2 | \ldots | \mathrm{N}\}$ as input training data set to the model and where N denote the pair's vector of input $(X_\mathrm{n} \in R^L)$ with output y_n considering noisy scalar. The noise uncertainty in the model output is result of an external factor in the form of truncation or observation errors (ε) (Wan and Sapsis 2017). The fundamental assumption is that noise is stationary and gaussian dispersed and

additive in nature. The output of the GPR model is calculated by Eq. (5).

$$y = \phi(x) + \varepsilon, \varepsilon \approx \mathrm{N}\left(0, \mathrm{s}^2{}_{noise}\right) \tag{5}$$

where, $s^2{}_{noise}$ denotes the noise variance. The GPR model consists ϕ function as the latent variable and the input values X play the role of indexing. This latent variable consists the finite collection of inputs $\{\phi(x_1), \ldots, \phi(x_l)\}$ with unique indices upholding Gaussian distribution (Hoang et al. 2016). The major advantage of using GPR over other machine models is that it considers prior Gaussian assumption which helps in interpolation and can be conveniently specified by the function of mean M(x) and covariance$C(x, x^{'})$:

$$\begin{aligned} M(\mathrm{x}) &= \psi[\phi(x)] \\ C\left(x, x^{'}\right) &= \psi\left[(\phi(x) - M(\mathrm{x}))\left(\phi\left(x^{'}\right) - M\left(x^{'}\right)\right)\right] \end{aligned} \tag{6}$$

where, $\psi[\cdot]$ denotes approximation. The function of mean is required only in the case missing or unobserved section of the input space (set to zero) and the process is entirely governed by covariance function which can be calculated using Eq. (7).

$$C\left(x, x^{'}\right) = \mathrm{q}_1 exp\left(\frac{\left\|x - x^{'}\right\|}{2q_2}\right) \tag{7}$$

where, $\|.\|$ is the norm defined on the input space, an interesting point to be noted here as the distant pairs of inputs x and $x^{'}$ increases, the covariance function decreases rapidly signify that weak correlation between $\phi(x)$and $\phi(x^{'})$. Whereas, q_1 is the hyper-parameter, which define the maximum allowable for the covariance and whereas q_2 denotes the rating decay in correlation, when the distance between points increases. In addition, the third hyper-parameter (q_3) which denotes the unknown variance ($s_{noise}{}^2$) of Eq. (6) not explicitly shown in Eq. (7). In general, hyper-parameter (q_1, q_2, q_3) are grouped together as a vector in the form of random realization vector (Q). This realization is most articulate with the training dataset and henceforth used to make a prediction (Kottek et al. 2006).

Material methods

Study area and field dataset description

Location of observation wells selected

Dwindling stream flows and deteriorating surface water quality forced Japanese nationals to explore more and more groundwater, which results in extensive depletion of groundwater depth. The area selected for this study is the Konan basin, where groundwater exploitation is a serious issue (Fig. 3). The Konan basin is located in the Kochi Prefecture of Japan. The whole basin is surrounded by mountains in the north and the Pacific Ocean in the south. East and west part of the study area is bounded by the intermittent Koso river and the perennial Monobe river. The Konan basin is underlain by alluvial unconfined and leaky aquifers. The spatial variability of aquifer parameters is ascertained by the large variation of hydraulic conductivity and transmissivity (Jha et al. 1999). The stations selected for the study are site E-2, E-3, E-4 and F-6. Site E-2 is towards the western part of the basin and near to the perennial Monobe river and stations E-3, E-4 and F-6 are in the central part of the Konan groundwater basin. According to the lithology available in Jha et al. (1999), all the stations selected are in the unconfined aquifer. Detailed information about geology of the basin is given in Jha et al. (1999). The stations selected in the center part of the aquifer are thickly populated and agricultural zone, where the water consumption is mainly utilized for domestic needs and agricultural purposes (greenhouse and fish culture). As daily exploitation of groundwater is more in the study area, a clear understanding of shorter time period groundwater fluctuation is of great interest to the water resources engineers for proper water resources planning and management.

Wavelet coherence analysis

Wavelet Coherent Analysis (WCA) has been generally applied in spatial and temporal analysis of hydrological problems. It inspects the phase lag and coherence between two time series data. Detailed description of the WCA approach is furnished in Grinsted et al. (2004) and Roshni et al. (2019). The WCA for the signal comparison was performed with the help of MATLAB functions (Grinsted et al. 2004). In the present study, the coherence and phase lags of groundwater depths of the selected stations E-2, E-3, E-4 and F-6 were analyzed by WCA and the results are shown in Fig. 4. Very short breakdowns are visible for the shorter time period at E-2, F-6 station and F-6, E-3 stations, but majority shows a very good coherence between signals. This confirms the fact that E-2, E-3, E-4 and F-6 observation wells are tapping same unconfined aquifer, which is in accordance with the litholog results of Jha et al. (1999) and Roshni et al. (2019).

In total, 84 data samples have been collected to establish a data set for training and verifying the ML models. Each data sample contains the information of the groundwater depth as the predictand variables and the information of the 4 predictors (rainfall, river stage, groundwater recharge, temperature) have been employed as influencing factors to estimate the groundwater depth. The statistical characteristics of the variables used in this study are provided in Table 1. Figure 5 shows the histogram plot of all attributes, which have been considered for the prediction of groundwater depth.

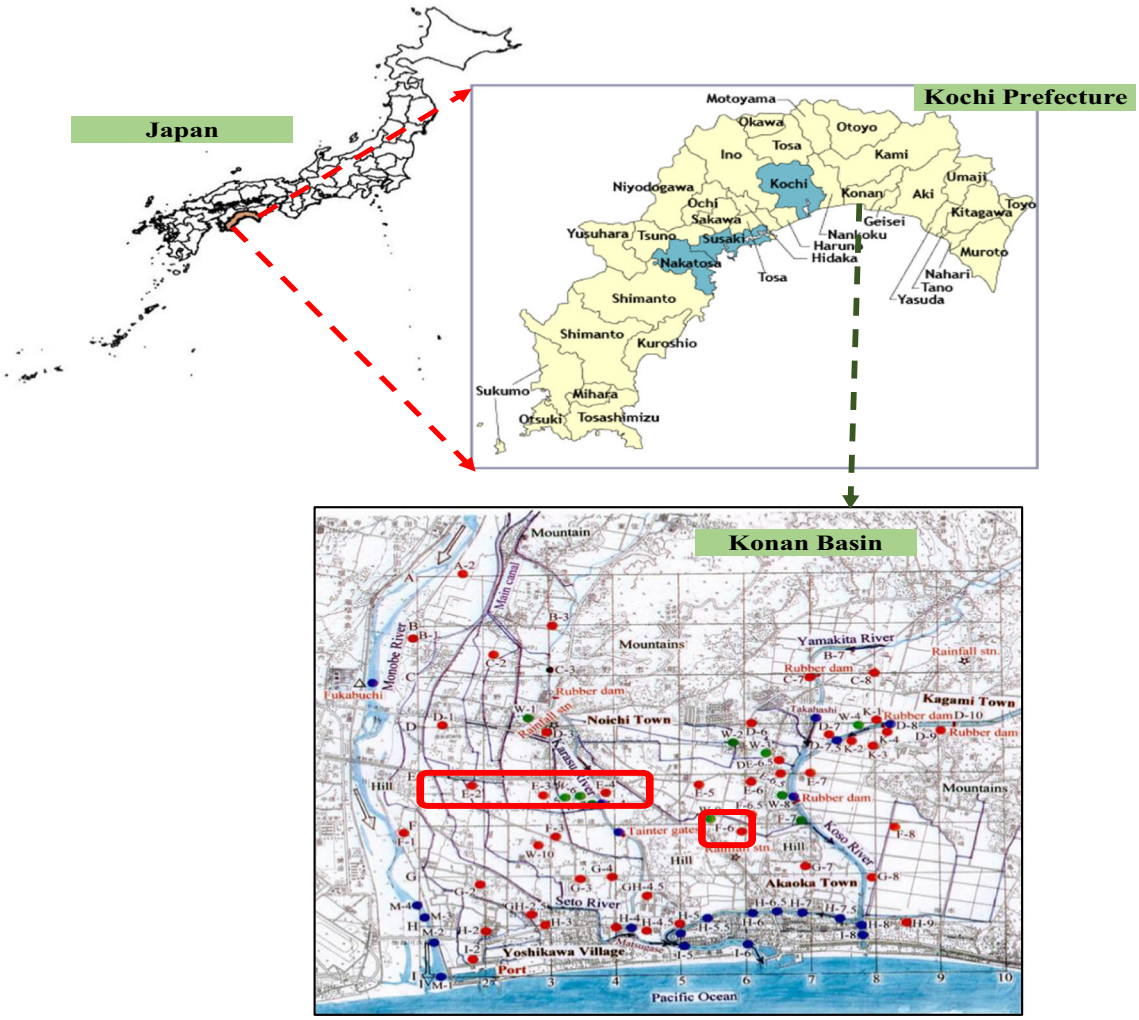

Fig. 3 Map of study area with the location of observation wells

Development of predictive models

The algorithm for deep learning models was developed using *Keras Deep Learning library* in Intel(R) Core (TM) i5-6200U 2.40 GHz under a computer operating system. To predict the groundwater depth, the two-other machine learning regression models were considered (i.e., ELM and GPR) using sub-routines of R package (Karatzoglou et al. 2004; Gosso and

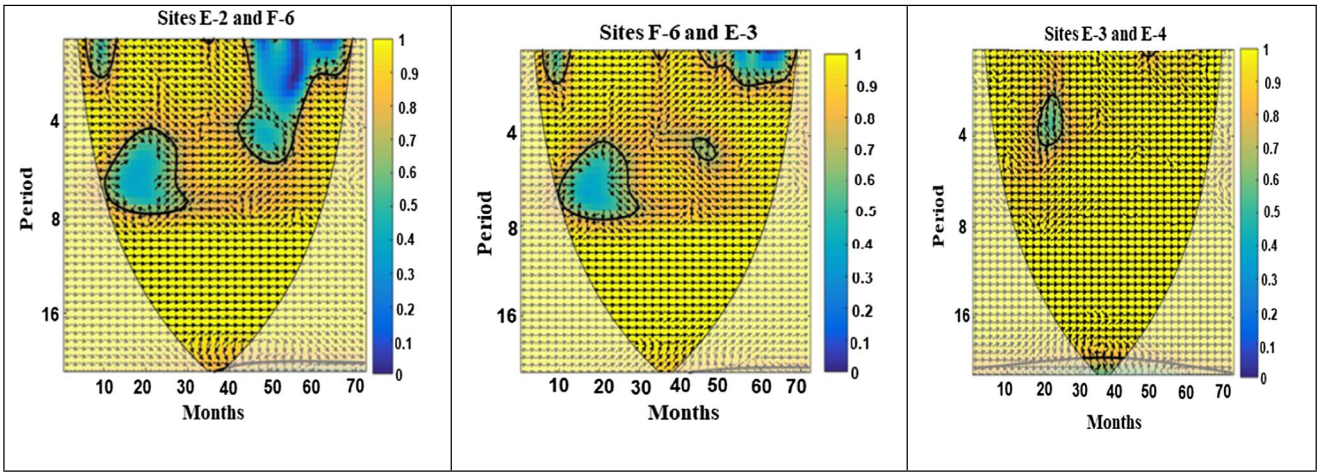

Fig. 4 WCA between time series analysis of groundwater depth between a) stations E-2 and F-6 (b) stations F-6 and E-3and (c) Stations E-3 and E-4

Table 1 Statistical description of the input variables of model

Statistic	Rainfall (mm)	River stage (m)	Temp (°C)	GWD E2 m	GWD E3	GWD E4	GWD F6	GWD E5	E2 Recharge mm.month^{-1}	E3 Recharge	E4 Recharge	F6 Recharge
Minimum	7	11.5	5.0	5.1	4.7	4.6	3.1	4.1	−278.1	−145.5	−150	−273.4
Maximum	1213	12.7	28.1	8.6	8.7	8.8	5.7	8.1	117.1	117.8	138.6	124.8
1st Quartile	82	11.7	9.7	6.2	6.1	5.7	3.8	5.0	−163.9	−77.8	−83.9	−97.5
Median	186	11.8	17.8	7.3	7.3	7.2	4.8	6.4	−94.4	−39.6	−42.5	−56.9
3rd Quartile	290	12.1	23.3	7.8	7.8	7.7	5.1	6.9	−40.0	−10.4	−21.5	−27.0
Mean	223	11.9	16.7	7.0	6.9	6.8	4.5	6.1	−104.3	−41.7	53.4	−61.9
Standard deviation	197	0.3	7.4	1.0	1.1	1.1	0.7	1.1	82.8	48.3	42.6	62.9
Skewness	2.1	0.7	−0.1	−0.5	−0.6	−0.4	−0.3	−0.2	−0.1	0.4	0.3	−0.2
Kurtosis	6.5	−0.1	−1.4	−1.1	−1.0	−1.1	−1.2	−1.2	−0.5	0.6	0.7	1.7

Gosso 2012) in open source R software. The structure of the models include an input matrix (x) defined by x = (recharge, temperature, precipitation, river stage, groundwater depth) represented as the predictor variables, while the groundwater depth as the target variable (y). An important step is the determination of input variables. Appropriate input combination is selected based on the relative percentage error (RPE) method proposed by (Lin et al. 2009).

$$RPE = \frac{E(\mathrm{L}) - \mathrm{E}(\mathrm{L}+1)}{E(L)} \times 100 \tag{8}$$

where, E(L) and E ($L+1$) are the roots mean square errors (*RMSE*s) for L and $L+1$ lag lengths of predictor, respectively for the models. The lag length is terminated once RPE is less than 5%.

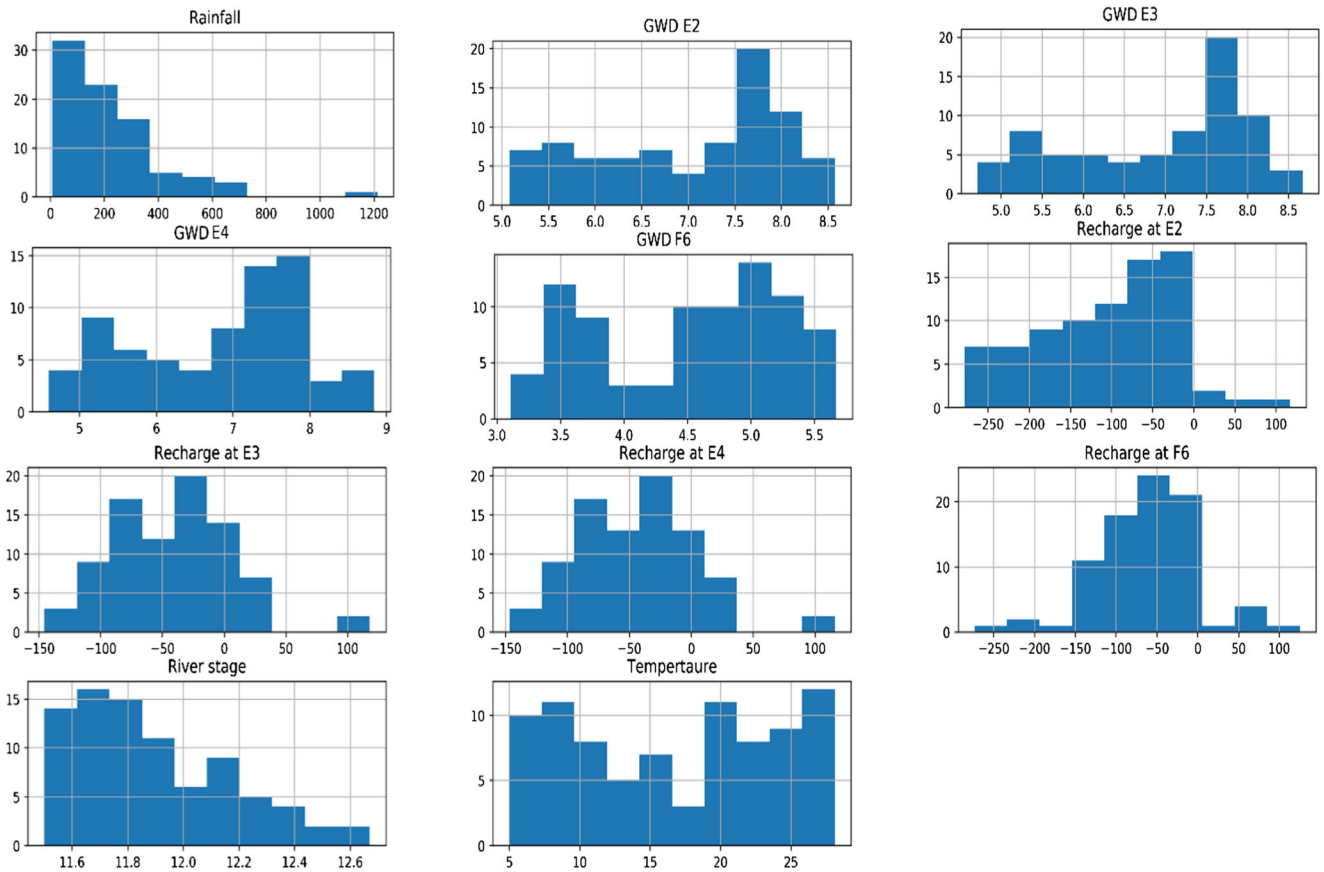

Fig. 5 Histogram plot of input variables

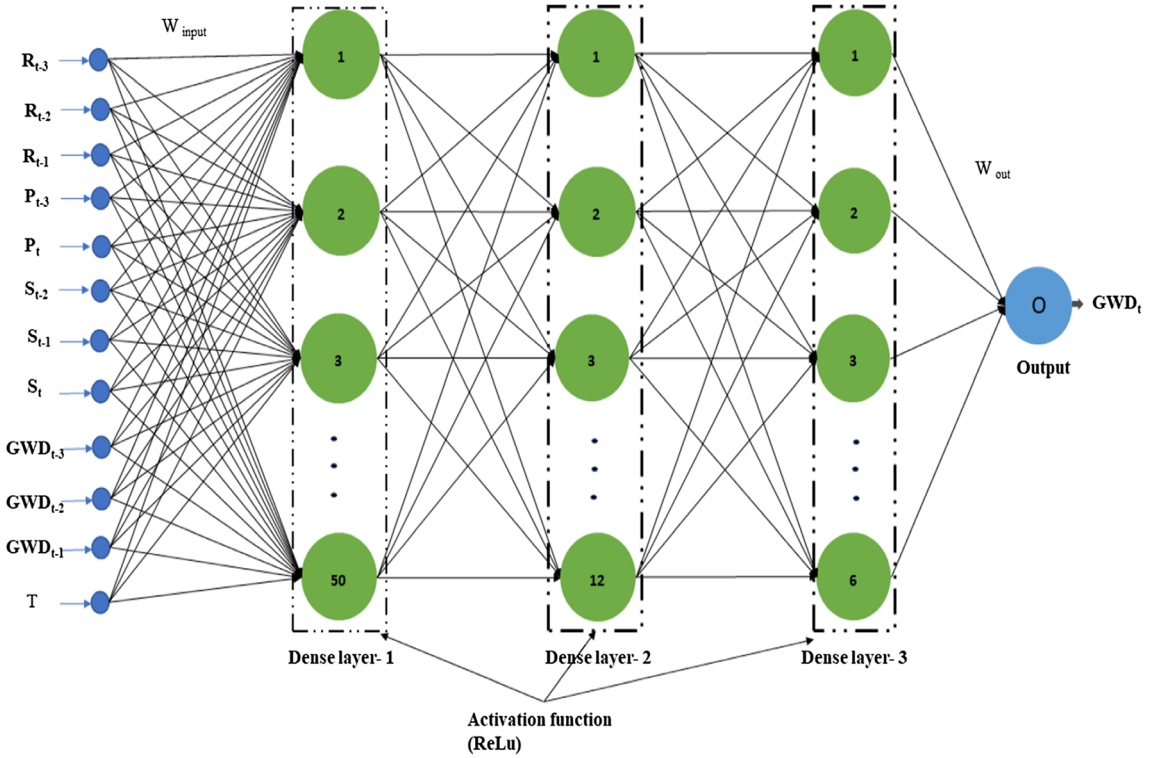

Fig. 6 The proposed architecture of multi-layer deep learning

Based on RPE method, the combination of three lags were selected as optimum inputs for the model development. Consequently, the antecedent 3-months, 2- months and 1-month time series of the recharge (*R*), precipitation (*P*), river stage (*S*), groundwater depth (*GWD*) along with current temperature (*T*) data were selected as inputs (i.e. R_{t-3}, R_{t-2}, R_{t-1}, P_t, P_{t-3}, S_{t-2}, S_{t-1}, S_t, GWD_{t-3}, GWD_{t-2}, GWD_{t-1}, T) to predict the groundwater depth (GWD_t).

In any modelling process, the major task is to find the appropriate training dataset and testing dataset. There is no rule of thumb for data division and it depend upon problem and availability of dataset. The randomly sampled 70% training dataset was used to developed the Deep learning network, ELM and GPR models and tested on remaining dataset. Prior to model development, all datasets were normalized using Eq. (9)

$$Z_{normalised} = \frac{(z - z_{min})}{(z_{max} - z_{min})} \quad (9)$$

where, Z = data value, Z_{min} = minimum value of the whole dataset, Z_{max} = maximum value of whole data and $Z_{normalized}$ = normalized dataset.

In order to develop the deep learning model, three-layer network was selected with *adam* optimizer and mean square error loss function. The selected *adam* optimizer is robust and well suited for non-convex optimization problem especially for deep learning network (Kingma and Ba 2015). The final structure of deep learning model consists three layers consisting 50,12,6 neurons in each layer with activation function *ReLu* at each respective layer and single neurons at output layer (Fig. 6).

The optimal batch size of 25 with 360 epochs were selected to fit the deep learning mode. The optimal batch size and epochs were selected adopting fivefold cross-validation techniques. The final model configuration used was discovered by trial and error. This leaves the door open to explore new and possibly better configurations. For designing the ELM framework network, a single layer network, with 13 neurons was selected to denote the number of the predictor variable. The ELM output layer had one neuron representing the predicted value. In the

Table 2 Statistical performance fitness parameters used to evaluate the model prediction accuracy (*r* = correlation coefficient, *RMSE* = root mean square error, *NSE* = Nash–Sutcliffe efficiency coefficient, *MAE* = mean absolute error, *d* = Index of agreement)

	Training			Testing		
	ELM	GPR	DL	ELM	GPR	DL
r	0.92	1.00	0.99	0.80	0.89	0.95
NSE	0.84	1.00	0.98	0.18	0.78	0.87
RMSE	0.11	0.00	0.04	0.20	0.10	0.08
MAE	0.09	0.00	0.03	0.15	0.08	0.06
d	0.96	1.00	1.00	0.85	0.94	0.97

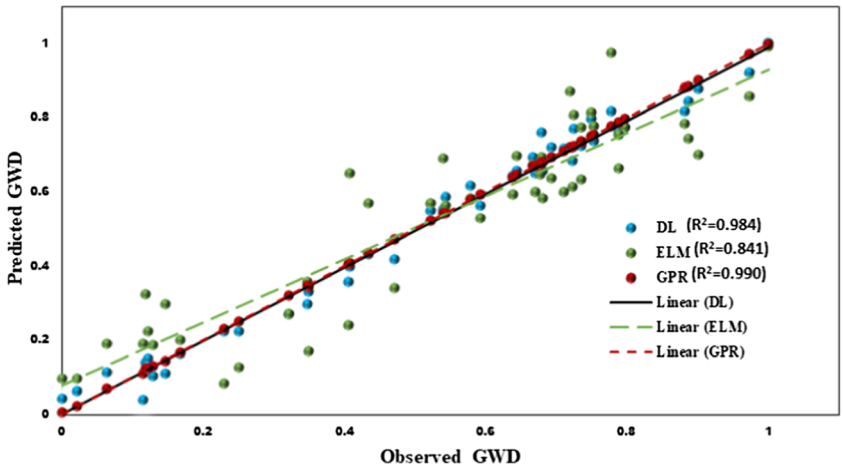

Fig. 7 Scatter plot displaying a statistical comparison with observed and predicted of three model estimates of the groundwater depth for training at station E4

beginning, number of neurons were trialed randomly from 1 to 20 and finally 13 neurons were selected based on the high performance and lowest error. The overall ELM architecture had the combination of 13 (input neurons)-1 (layer)- 1 (output neuron) with the radial basis activation function (Fig. 2). For GPR model, optimal values of the hyper-parameters i.e. covariance function was tuned by maximizing the marginal likelihood based on the trial and error basis. The final configuration of the model was obtained at; sigma (0.1) and variance noise (0.005) with tolerance of termination criterion (0.01).

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{l} (GWD_{E_i} - GWD_{O_i})^2}{l}} \tag{10}$$

$$r = \left(\frac{\sum_{i=1}^{l} \left(GWD_{E_i} - GWD_{\overline{E}_i} \right) \left(GWD_{O_i} - GWD_{\overline{O}_i} \right)}{\sqrt{\sum_{i=1}^{l} \left(GWD_{E_i} - GWD_{\overline{E}_i} \right)^2 \sum_{i=1}^{n} \left(GWD_{O_i} - GWD_{\overline{O}i} \right)^2}} \right) \tag{11}$$

Model performance assessment

Values of control parameters of models were selected initially and thereafter varied in trials till the best fitness measures were produced. To investigate the performance of the proposed models, five statistical indicators including root mean square error (*RMSE*) and coefficient of correlation *(r)*, Nash–Sutcliffe efficiency coefficient (*NSE*), mean absolute error (*MAE*) and index of agreement (*d*) were used and is shown in equations (10, 11, 12, 13 and 14).

$$\mathrm{NSE} = \left(1 - \frac{\sum_{i=1}^{l} (GWD_{Ei} - GWD_{O_i})^2}{\sum_{i=1}^{l} \left(GWD_{O_i} - GWD_{\overline{O}_i} \right)^2} \right) \tag{12}$$

$$\mathrm{MAE} = \left(\frac{100}{l} \sum_{i=1}^{l} \left| \frac{GWD_{O_i} - GWD_{E_i}}{GWD_{Oi}} \right| \right) \tag{13}$$

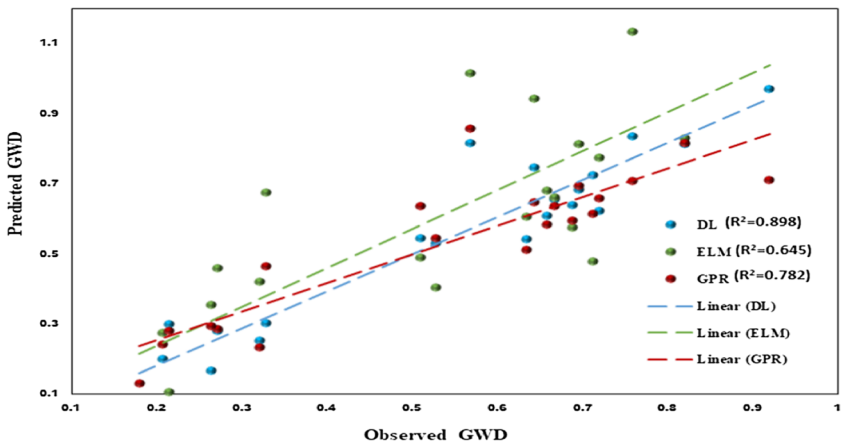

Fig. 8 Scatter plot displaying a statistical comparison with observed and predicted of three model estimates of the groundwater depth for testing at station E4

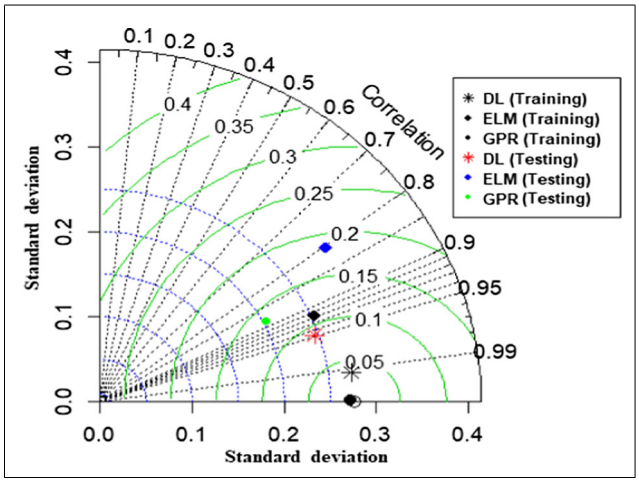

Fig. 9 Sample Taylor diagram displaying a statistical comparison with observations of three model estimates of the groundwater depth for training and testing

$$d = 1 - \left[\frac{\sum_{i=1}^{l} (GWD_{E_i} - GWD_{O_i})^2}{\sum_{i=1}^{l} (|GWD_{E_i}| - |GWD_{O_i}|)^2} \right], 0 \le d \le 1 \quad (14)$$

where, GWD_{E_i} is the i^{th} estimated monthly rainfall using models; GWD_{O_i} is the i^{th}observed monthly groundwater depth; $GWD_{\overline{E_i}}$ is the average of the estimated monthly groundwater depth; $GWD_{\overline{O_i}}$ is the average of the observed monthly groundwater depth and l is the number of observations.

Results and discussion

Groundwater is a hidden heterogenous complex system consisting of numerous constituents of hydrological and geological properties above and below the ground surface. A major concern in groundwater depth determination is to predict this precisely. The key parameter (groundwater depth) is highly dynamic due to nonlinear behavior of associated attributes. In this research, an attempt has been made to apply deep learning model that can mimic and capture the empirical relationship between groundwater depth and its attributes. In addition, a comparative study was also performed, considering well-known soft computing technique (ELM, GPR) to find deep learning adaptability. The predictive accuracies of the investigated models during training and testing periods are shown in Table 2. During training phase, all the selected models showed good relationship for prediction of groundwater depth (Fig. 7). The GPR model outperformed in terms of fitness parameters ($r = 1$, $RMSE = 0$, $NSE = 1$, $MAE = 1$, d = 1). However, while testing GPR model showed significant lower performance ($r = 0.89$, $RMSE = 0.10$, $NSE = 0.78$, $MAE = 0.08$, $d = 0.94$). This clearly indicates that the GPR model underfits for test dataset. As the GPR model estimates the probability distribution over assigned functions to fit the data. Moreover, in this case small fluctuations in the groundwater depth or certain noises will be consumed in the data due to memorizing particular characteristics of the training data instead of discovering a general predictive rule or due to insufficient number of datasets. This might be the reason for lesser predictive performance during the testing phase. Similar case was also found in case of ELM model which also showed underfit condition. This shown that random learning structure fit the linear regression to the output layer but fail during the testing dataset. Whereas, the DL model showed consistent performance during training and testing (Figs. 7 and 8), further it was also found, good relationship between predictor and predictand in terms of fitness criteria (Table 2). ELM has advantages due to random learning structure with simple linear regression on the output layer. However, deep networks have many advantages especially in feature extraction and activation learning function for large data sets (not with ELM) and some layers in deep networks can be randomized for better recognition of patterns.

For the comparative assessment, a mathematics based graphical Taylor diagram was plotted using R package (Lemon et al. 2009). It is used as statistical summaries to evaluate the degree of correspondence between observed and predicted values in terms of Pearson correlation coefficient and standard deviation and root mean square error

Table 3 Statistical performance fitness parameters use to evaluate the model prediction accuracy at different stations (r = correlation coefficient, $RMSE$ = root mean square error, NSE = Nash–Sutcliffe efficiency coefficient, MAE = mean absolute error, d = Index of agreement)

Fitness measure	E2			E3			F6		
	DL	ELM	GPR	DL	ELM	GPR	DL	ELM	GPR
r	0.917	0.813	0.901	0.968	0.904	0.963	0.938	0.876	0.909
NSE	0.823	0.614	0.79	0.907	0.768	0.907	0.848	0.678	0.823
RMSE	0.118	0.174	0.128	0.08	0.126	0.08	0.109	0.149	0.11
MAE	0.092	0.139	0.103	0.066	0.105	0.061	0.086	0.112	0.083
d	0.954	0.894	0.929	0.977	0.938	0.975	0.962	0.915	0.946

Table 4 Different models used for the prediction of groundwater depth by different researchers

Authors	Year	Models	Correlation coefficient (r)
Behzad et al. [14]	2009	Support vector machine (SVM), Artificial neural network (ANN)	SVM ($r = 0.96$), ANN ($r = 0.943$)
Yoon et al. [36]	2011	SVM, ANN	SVM ($r = 0.793$), ANN ($r = 0.733$)
Shiri et al. [12]	2013	Genetic programming (GP), ANN, SVM, Adaptive-neuro fuzzy interface system (ANFIS),	GP ($r = 0.850$), ANN (r = 0.72), SVM ($r = 0.701$), ANFIS ($r = 0.551$),
Gong et al. [37]	2016	ANN, SVM and ANFIS	ANFIS ($r = 0.931$), SVM ($r = 0.921$), ANN ($r = 0.917$)
Sahoo et al. [38]	2017	Multivariate linear regression model (MLR), Multivariate nonlinear regression model (MNLR), ANN	ANN ($r = 0.817$), MNLR ($r = 0.67$), MLR ($r = 0.60$),
Yadav et al. [39]	2017	Extreme learning machine (ELM) and SVM	ELM ($r = 0.949$), SVM (0.912)
Wen et al. [40]	2017	Wavelet-ANN	WA-ANN ($r = 0.921$)
Alizamir et al. [19]	2018	ELM, ANN, RBF, ARMA (2,1)	ELM ($r = 0.998$) ANN ($r = 0.997$) RBF ($r = 0.997$) ARMA ($r = 0.893$)
Yu et al. [13]	2018	WA-SVR, WA-ANN, SVR, ANN	WA-SVR ($r = 0.96$), WA-ANN ($r = 0.89$), SVR ($r = 0.83$) ANN ($r = 0.82$)
Present study	2019	Deep learning (DL), ELM, Gaussian process regression (GPR)	DL ($r = 0.95$), ELM ($r = 0.80$), GPR ($r = 0.89$)

(Taylor 2001). The Pearson correlation is represented by an azimuthal angle (dash-blue contours); RMSE is proportional to the distance from the x-axis (green contours) and radial distance from the origin (blue contours) is the standard deviation of the simulated pattern. Figure 9 shows that the DL model outperformed during training and testing in terms of fitness parameters.

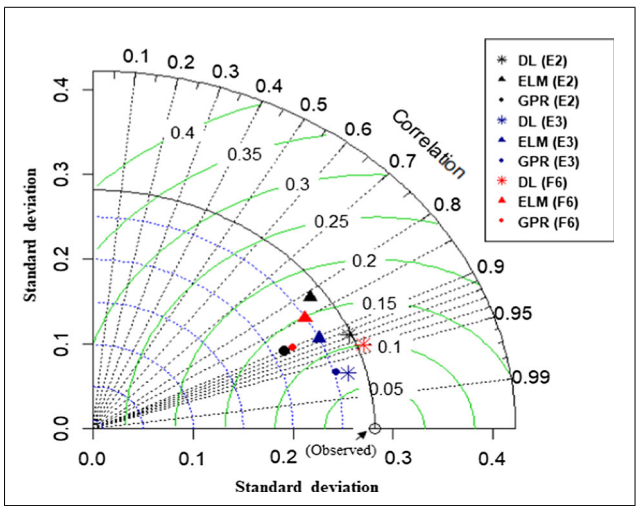

Fig. 10 Taylor diagram displaying a statistical comparison three model at three different stations to estimates the groundwater depth

Furthermore, to compare the practicality and generalization ability of deep learning with others models and to develop generalized model for this unconfined aquifer of Konan basin. The DL models were further validated on others stations (E2, E3 and F6) which are the part of the same unconfined aquifer, selected based on the WCA (Fig. 4). To check the robustness of DL model with others models (ELM and GPR), a comparative performance study has been done to predict the groundwater depth fluctuation. From the analysis of results, the DL model has shown good agreement in terms of fitness measurement irrespective of stations in the Konan basin (Table 3). Figure 10 provide the clear confirmation that DL model has outperformed in comparison to ELM and GPR irrespective of selected station for this Konan basin. The Taylor plot signified that standard deviation of observed groundwater depth is approximately closed to predicted value of groundwater depth by the DL model. In addition, violin plot has been drawn for three models for different stations (Hintze and Nelson 1998). It is hybrid of box plot and density trace which provide better indication of the shape of the error distribution. The density trace gives insightful to error distribution. Figure 11 shows the

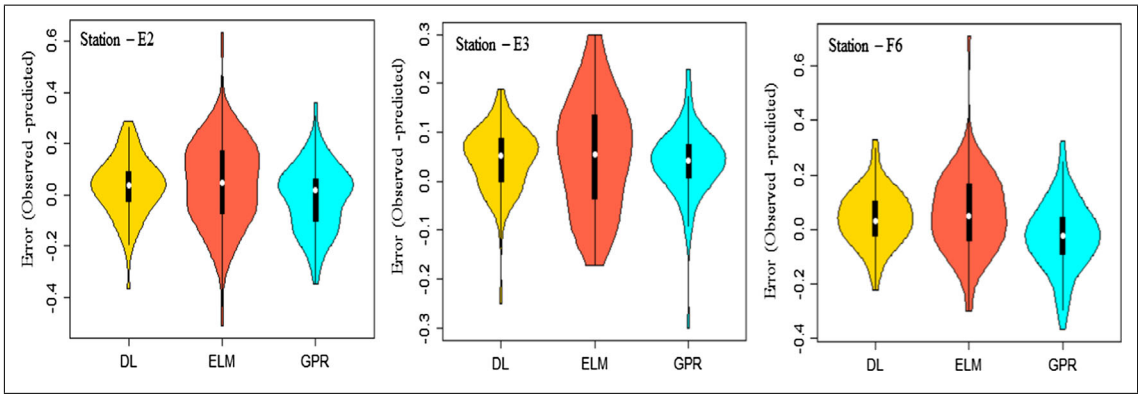

Fig. 11 Violin plot displaying an error (observed-predicted) for three model at three different stations

error distribution of applied models (DL, ELM and GPR) at different stations (E2, E3 and F6). From the analysis, it is clear that DL model shows, error clustered around the median and near the interquartile range, whereas ELM models show larger interquartile range with larger outlier followed by GPR (Fig. 11). The analysis of error signified that DL model is more robust and consistent in terms of prediction of groundwater depth. Furthermore, the finding of this research is compared with the others models used for predict the groundwater depth by the different researchers (Table 4). This research explores, the feasibility of DL model for prediction of groundwater depth considering its attributes (i.e. recharge, precipitation, river stage and temperature). The Interesting finding of this research is that even with the relatively small dataset, deep learning model is well capable of predicting the groundwater fluctuation compared to other models (ELM and GPR) which are inconsistent and insignificant. This research finding implies that, DL model may be used as an optional technique in comparison to physical based model when prediction is important than the underlying groundwater process.

Conclusions

The groundwater depth is an important parameter of water resource and is widely used in the design and planning of land and water management strategies. To avoid painstaking procedures to obtain this parameter, this study proposed three Machine Learning (ML) approaches namely, DL, ELM, and GPR. Groundwater depth and its attributes were used to develop the ML-based models. The dataset consisted of rainfall, recharge, river stage, temperature as predictors to infer the groundwater depth fluctuations. All the proposed models were capable of producing good prediction accuracy in terms of fitness measures. The DL model was found to have the best accuracy followed by GPR and ELM models for developing a generalized model for the study basin. The proposed model was also validated at other three stations (i.e., E2, E3 and F6) of the same unconfined aquifer. Others stations were selected based on the WCA analysis. From the results, it is evident that the developed DL model could be used as a reliable ML model irrespective of different stations in the study basin and it can be considered as generalized model to simulate groundwater depth fluctuations. Furthermore, it was also found that the DL model outperform even with the small dataset. The finding of this study recommends DL as potential alternative method to assist groundwater researchers while studying groundwater fluctuations. The future direction of the current work could be application of the Deep Learning model to solve groundwater problems as well as utilization of other state-of-the-art methods.

Acknowledgements Authors would like to thank Editors and unknown reviewers for improving the quality of the research paper.

Compliance with ethical standards

Conflict of interest We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Abbreviations *RMSE*, Root Mean Square Error; *r*, Coefficient of correlation; *NSE*, Nash–Sutcliffe efficiency coefficient; *d*, index of agreement; *MAE*, mean absolute error; $\phi(.)$, activation function; *w*, weight vector; *b*, random bias; w_j, weight vector to the hidden neuron; β_j, denotes j^{th} hidden node weight vector connection to input of the hidden node to the output node; ε, observation errors; s^2_{noise}, noise variance; ϕ, latent variable function; $\psi[\cdot]$, denotes approximation; q_1, (hyper-parameter); q_2, denotes the rating decay in correlation; $H^{\dagger}$, inverse of matrix; D, input training dataset to the model; μ, mean; σ, standard deviation; C, Covariance; $\psi[\cdot]$, (Approximation); Q, random realization vector; R, recharge,; P, precipitation; S, river stage; GWD, groundwater depth; T, current temperature data; t, time period; Z, data value; Z_{min}, minimum value of the whole dataset; Z_{max}, maximum value of whole data; $Z_{normalized}$, normalized dataset

References

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

Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C et al (2016) Deep speech 2: End-to-end speech recognition in english and mandarin. In: International conference on machine learning, 2016, pp 173–182

Bierkens MF (1998) Modeling water table fluctuations by means of a stochastic differential equation. Water Resour Res 34(10):2485–2499

Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Phys Chem Earth, Parts A/B/C 31(18):1164–1171

Chollet F (2017) Deep learning with Python. Manning Publications Co

Elkahky AM, Song Y, He X (2015) A multi-view deep learning approach for cross domain user modeling in recommendation systems. In: Proceedings of the 24th International Conference on World Wide Web, 2015, pp 278–288

Fan Z, Chen Y, Li H, Ma Y, Kurban A (2008) Determination of suitable ecological groundwater depth in arid areas in northwest part of China. J Arid Land Resour Environ 22(2):1–5

Gosso, A., & Gosso, M. A. (2012). "Package 'elmNN'," ELM Package Version 1.0, July. 17, 2012. [Online]. Available: https://cran.rproject.org/web/packages/elmNN/index.html

Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Process Geophys 11(5/6):561–566

Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. Am Stat 52(2):181–184

Hirschberg J, Manning CD (2015) Advances in natural language processing. Science 349(6245):261–266

Hoang N-D, Pham A-D, Nguyen Q-L, Pham Q-N (2016) Estimating compressive strength of high performance concrete with Gaussian process regression model. Adv Civ Eng 2016:8–8. https://doi.org/10.1155/2016/2861380

Huang G-B, Zhu Q-Y, Siew C-K (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, 2004, IEEE, vol 2, pp 985–990

Huang G, Huang G-B, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

Jha MK, Sahoo S (2015) Efficacy of neural network and genetic algorithm techniques in simulating spatio-temporal fluctuations of groundwater. Hydrol Process 29(5):671–691

Jha MK, Chikamori K, Kamii Y, Yamasaki Y (1999) Field investigations for sustainable groundwater utilization in the Konan basin. Water Resour Manag 13(6):443–470

Ji S, Xu W, Yang M, Yu K (2012) 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231

Jiang GQ, Xu J, Wei J (2018) A deep learning algorithm of neural network for the parameterization of typhoon-ocean feedback in typhoon forecast models. Geophys Res Lett 45(8):3706–3716

Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) Kernlab-an S4 package for kernel methods in R. J Stat Softw 11(9):1–20

Karbasi M (2018) Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model. Water Resour Manag 32(3):1035–1052

Kingma D, Ba JA (2015) Adam: A method for stochastic optimization. arXiv 2014. arXiv preprint arXiv:1412.6980v9 [cs.LG], 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15(3):259–263

Krishna B, Satyaji Rao Y, Vijaya T (2008) Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. Hydrol Process 22(8):1180–1188

Kumar D, Singh A, Samui P, Jha RK (2019) Forecasting monthly precipitation using sequential modelling. Hydrol Sci J 64(6):690–700

Lemon J, Bolker B, Oom S, Klein E, Rowlingson B, Wickham H et al (2009) Plotrix: various plotting functions. R package version 2.7–2. R Project for Statistical Computing Vienna, Austria

Lin GF, Chen GR, Wu MC, Chou YC (2009) Effective forecasting of hourly typhoon rainfall using support vector machines. Water Resour Res 45(8)

Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2014) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16(2):865–873

Maheswaran R, Khosa R (2013) Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics. Comput Geosci 52:422–436

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

Njock PGA, Shen S-L, Zhou A, Lyu H-M (2020) Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/t-SNE model. Soil Dyn Earthq Eng 130:105988

Nourani V, Mogaddam AA, Nadiri AO (2008) An ANN-based model for spatiotemporal groundwater level forecasting. Hydrol Process 22(26):5054–5066

Raghavendra NS, Deka PC (2016) Multistep ahead groundwater level time-series forecasting using gaussian process regression and ANFIS. In: Advanced Computing and Systems for Security, Springer, pp 289-302

Rasmussen CE (2004) Gaussian processes in machine learning. In: Advanced lectures on machine learning, Springer, pp 63-71

Roshni T, Jha MK, Deo RC, Vandana A (2019) Development and evaluation of hybrid artificial neural network architectures for modeling Spatio-temporal groundwater fluctuations in a complex aquifer system. Water Resour Manag 33:2381–2397. https://doi.org/10.1007/s11269-019-02253-4

Roshni T, Jha MK, Drisya J (2020) Neural network modeling for groundwater-level forecasting in coastal aquifers. Neural Comput & Applic 32:12737–12754

Sahoo BB, Jha R, Singh A, Kumar D (2019) Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys 67(5):1471–1481

Scher S (2018) Toward data-driven weather and climate forecasting: approximating a simple general circulation model with deep learning. Geophys Res Lett 45(22):12,616–612,622

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117

Schölkopf B, Smola A, Müller K-R (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5):1299–1319

Shiri J, Kisi O, Yoon H, Lee K-K, Nazemi AH (2013) Predicting groundwater level fluctuations with meteorological effect implications—a comparative study among soft computing techniques. Comput Geosci 56:32–44

Sun AY, Wang D, Xu X (2014) Monthly streamflow forecasting using Gaussian process regression. J Hydrol 511:72–81

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

Tapoglou E, Karatzas GP, Trichakis IC, Varouchakis EA (2014) A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. J Hydrol 519:3193–3203

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192

Wan ZY, Sapsis TP (2017) Reduced-space Gaussian process regression for data-driven probabilistic forecast of chaotic dynamical systems.

Phys D: Nonlinear Phenom 345:40–55. https://doi.org/10.1016/j.physd.2016.12.005

Wang P, Yu J, Zhang Y, Fu G, Min L, Ao F (2011) Impacts of environmental flow controls on the water table and groundwater chemistry in the Ejina Delta, northwestern China. Environ Earth Sci 64(1):15–24

Ye L, Gao L, Marcos-Martinez R, Mallants D, Bryan BA (2019) Projecting Australia's forest cover dynamics and exploring influential factors using deep learning. Environ Model Softw 119:407–417

Yu H, Wen X, Feng Q, Deo RC, Si J, Wu M (2018) Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. Water Resour Manag 32(1):301–323

Zhang N, Shen S-L, Zhou A, Xu Y-S (2019) Investigation on performance of neural networks using quadratic relative error cost function. IEEE Access 7:106642–106652