Computers and Electronics in Agriculture 191 (2021) 106568

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional Long Short-Term Memory (BiLSTM) deep neural network

Redvan Ghasemlounia [a], Amin Gharehbaghi [b], Farshad Ahmadi [c,*], Hamid Saadatnejadgharahassanlou [d]

[a] Faculty of Engineering, Dept. of Civil Engineering, Istanbul Gedik Univ, Istanbul, Postal code: 34876, Turkey
[b] Faculty of Engineering, Dept. of Civil Engineering, Hasan Kalyoncu Univ, Şahinbey, Gaziantep, Postal code: 27110, Turkey
[c] Department of Hydrology & Water Resources Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[d] Islamic Azad University, Ahar Branch, Civil Engineering Department, Tabriz, East Azerbaijan, Iran

ARTICLE INFO

Keywords:
Deep neural networks
BiLSTM
Algorithm tuning process
Groundwater level fluctuations
Miandoab plain

ABSTRACT

In this study, mean monthly groundwater level (*GWL*) fluctuations of four different monitoring piezometers are forecasted in an agricultural area by developing new proposed structures of BiLSTM based neural network models. For developing models, 186 monthly measured water table depth during time period (Sep 2001–Feb 2017) are employed. Due to a lack of meteorological variables records, the module of sequence-to-one is employed. By tuning several hyperparameters such as the number of hidden units (*NHU*), kind of state activation function (*SAF*), learning dropout rate (*P-rate*), and network topology pattern (*NTP*), the performance of designed models is improved. Based on results of modelling, in all designed models, the optimal *P-rate* is obtained 0.5, and the running time decrease by increasing *P-rate* in the same model. Since harsh fluctuations of *GWL*, the general Single-LSTM mode (1) and Single-BiLSTM model (2) perform too weak compared to other models. Nonetheless, by deepening models through adding the suitable hidden layers instead of many numbers of nodes, the performance of Single-BiLSTM model is improved strikingly. Based on statistical evaluation metrics, violin plot, and also *TLP* (Total Learnable Parameters) value, novel proposed model (5) (simple Double-BiLSTM model combined by a Multiplication layer (×)) is selected as the superior model. In the piezometer 4 with a range of 4.49 (m), the model (5) yielded in an R^2 of 0.89 and an *RMSE* of 0.17 (m), while in the same physical characteristics, the simple Double-BiLSTM model (3) yielded in an R^2 of 0.77 and an *RMSE* of 0.25 (m).

1. Introduction

Groundwater considers a significant and broadly distributed resource that provides freshwater largely in the semi-arid and arid areas in worldwide. But, nowadays, since a speedy development in municipal districts, farming, industrialization, and particularly growing populace, request for water reserves has been perceptibly intensified.

Accurate prediction of groundwater level (*GWL*) facilitates the assessment of shortage in groundwater resources. Consequently, it induces a rational management of groundwater resources for diverse demands by the water resources managers and also averts the depletion of water stores (Raghavendra and Deka, 2016). Predicting *GWL* fluctuations over a long period in cultivated regions considers a problematic task, due to extremely stochastic, nonlinear interactions, uncertainties in meteorological and hydrogeological parameters, some human breaking in, etc. (Todd and Mays, 2005). Thus, choosing an apt scheme by which to predict a complicated trend of *GWL* fluctuations is regarded as a momentous problem that necessitates developing extraordinarily hi-tech systems (Fallah-Mehdipour et al., 2013; Zare and Koch, 2018). In the last three decades, under dramatic advancements in computer science, various numerical techniques have been developed for water and environmental problems (Zhou et al., 2017; Gharehbaghi et al., 2017; Gharehbaghi, 2017, 2016), especially in the field of groundwater hydrology (Kim et al., 2008). Even so, implementing numerical approaches involves gathering a massive quantity of systematic and consistent information about physical attributes of used sphere and model parameters such as topography and geological coverage, transmissivity and

* Corresponding author.
E-mail addresses: redvan.ghasemlounia@gedik.edu.tr (R. Ghasemlounia), amin.gharehbaghi@hku.edu.tr (A. Gharehbaghi), f.ahmadi@scu.ac.ir (F. Ahmadi).

https://doi.org/10.1016/j.compag.2021.106568
Received 9 July 2021; Received in revised form 9 November 2021; Accepted 12 November 2021
Available online 19 November 2021

Nomenclature

LSTM	Long Short-Term Memory
BiLSTM	Bi-directional Long Short-Term Memory
P-rate	Learning dropout rate
AF	Activation functions
RMSE	Root mean square error (m)
STDV	Standard deviation
CV	Coefficient of variation
y_i	Predicted *GWL* at time *i*
x_i	Observed *GWL* at time *i*
μ_x	Average of observed *GWL* at time *i*
ANNs	Artificial Neural Networks
DNNs	Deep Neural Networks
TLP	Total Learnable Parameters
N	Number of datasets
R^2	Determination coefficient [-]
GWL	Groundwater level (m)
Bi-RNN	Bi-directional Recurrent Neural Network
NTP	Network topology pattern
NHU	Number of hidden units
σ_x	Standard deviation of observed *GWL* at time *i*
σ_y	Standard deviation of estimated *GWL* at time *i*
μ_y	Average of estimated *GWL* at time i

specific storage coefficient, hydrological and climatical dataset, etc. (Yoon et al., 2011).

Formerly, the majority of researchers have employed classical Neural Network (NN) models to predict highly nonlinear *GWL* fluctuations which often excelled on numerical models (Mohanty et al., 2013; Chang et al., 2016; Guzman et al., 2017). Despite everything, since *GWL* time series is a non-stationary process, these models have characteristically encountered severe predicaments such as weights fine-tuning and mostly a lack of enough ability to maintain prior information in the traditional NNs like FFNN (Feed-Forward Neural Network), etc. (Feng et al., 2008).

In this study, to address the above-mentioned constraints and drawbacks, a novel method of large-scale data diagnostic namely, Bi-directional Long Short-Term Memory (BiLSTM) neural network, Deep Neural Networks (DNNs) based, a Machine Learning Models (MLMs), is developed to predict accurately long-term *GWL* fluctuations. In this direction, different layer structures of BiLSTM model are designed, owing to its sophisticated network architect. Also, because nonexistence of exact records of meteorological variables, the module of sequence-to-one is operated in these structures for acquiring the context information of time series *GWL*.

DNNs consider an improvised version of Artificial Neural Networks (ANNs) with a lot of layers among the input and output layers. They can estimate multifaceted nonlinear relationships extremely simpler than shallow networks and extract noteworthy traits from the raw inputs, by making cyclic compositional models. More specifically, deep learning structures have a significant depth termed, Credit Assignment Paths (CAPs), which are latent random links of transformations between inputs to target (Bengio, 2007). At present, DNNs have been utilized in various fields of engineering for predicting intelligent load, wind speed, and solar irradiance (Jalali et al., 2021a,b,c).

LSTM (Long Short-Term Memory), as an eminent representative of DNNs, is an upgraded version of RNNs (Recurrent Neural Networks) which has been developed to pass salient information across time steps (Hochreiter and Schmidhuber, 1997). LSTM is well-matched and flexible enough to approximate long-term nonlinear input–output relationships in the domain of sequential learning tasks. The foremost benefit of LSTM over the conventional FFNN is its ability to reminisce patterns over a long time, owing to its incomparable progressive structure. BiLSTM includes two independent LSTM with a similar fundamental framework and chain-form structure for reiterating units in the hierarchical mode (Graves and Schmidhuber, 2005).

In BiLSTM, the training sequence process contains the forward and backward RNNs wherein connections among units shape a directed loop and make data to stream in the interior of the network so as to the prior information can be well-kept up for future use (Byeon et al., 2015). Both models were represented to address limitations corresponding to the gradients vanishing dilemma and relatively large historical data transformation once backpropagation through time is performed (Abdelhameed et al., 2018). Its interior self-looped cells make it capable enough to implement time series prediction tasks (Sutskever et al., 2014).

Above and beyond, prediction performance also strongly depends on the type of input variables, especially the longer frequency historical data in different locations, which can certainly progress the ability of models (Pelosi et al., 2016; Boithias et al., 2017). Because *GWL* is a correlated, continuous, and probabilistic time series data, thus its time dimension should be taken into account in modelling.

LSTM has been already employed as an enhanced deep learning model in the field of hydrology science (Jeong and Park, 2019; Yin et al., 2020; Lee et al., 2020; Bai et al., 2021; Dikshit et al., 2021; Feng et al., 2020). Zhang et al., (2018) developed LSTM based model to forecast the water table depth in agricultural areas. They reported that the suggested model achieved better R^2 scores (0.789–0.952) compared to the traditional neural networks with relatively small R^2 scores (0.004–0.495). Jeong et al. (2020) predicted *GWL* fluctuations based on the robust training of RNN using corrupted data. They asserted that the optimal cost function is the most effective method in discarding the impact of outliers in the training process for Gangjin–Seongjeon monitoring well. Vu et al. (2021) reconstructed missing *GWL* data through LSTM deep neural network. They declared that the suggested approach entailed no calibration for time-lag based data for processing and operations just relied on *GWL* fluctuations to recuperate missing data in the employed piezometers. Bowes et al. (2019) forecasted groundwater table in a flood-prone coastal city via RNN and LSTM neural networks. The results proved that LSTM neural network outperformed RNN.

Despite LSTM is a rising research field, yet various applications of BiLSTM for more complicated natural data are still rare. The novelty of the current study is developing an innovative layer structure of BiLSTM-based models using the module of sequence-to-one for complex natural phenomena like *GWL* fluctuations. To the best of the authors' knowledge, the last studies were limited in predicting *GWL* fluctuations with module of sequence-to-sequence regression under the simple architecture of DNNs models. No research has been reported in the literature on examining the performance of unique proposed models (i.e., 4 and 5) by the authors. One of the advantage of new suggested models is that newly proposed models against the expensive, difficult, and time-consuming module of sequence-to-sequence regression, do not require large numbers of hydrological and meteorological variables records as indispensable input data for modelling.

The specific objectives and scope of the present paper are as follows: (1) to develop different layer structures of BiLSTM based neural network models for accurate predicting of *GWL* fluctuations with time series characteristics using mean monthly observed long-term water table depth of 4 different piezometers without applying meteorological parameters in Miandoab plain, (2) to determine an optimal value of hyperparameters such as type of state activation function (*SAF*), number of hidden units (*NHU*), and learning dropout rate (*P-rate*) for better configurating of designed models to decrease the effect of underfitting or overfitting problems, and (3) to compare and analyze results of

prediction for selecting the best model by using evaluation metrics.

2. Study area and observational data

2.1. Study area

Miandoab plain with an area of around 1100 km^2, situated in the South of Urmia Lake (one of the largest and saltiest water lakes on the earth) at Miandoab city, West Azerbaijan province, the Northwest of Iran. It has 36°58′10″N longitude and 46°06′10″E latitude coordinates with 4311 ft above sea level. Miandoab plain is a division of Azerbaijan-Alborz structural region with very prolific soil, a semi-arid and cold climate (Norouzi and Moghaddam, 2020). Fig. 1 demonstrates the geographical position of the study area.

Peak rainfall in this district befalls predominantly from January to May. The average yearly rainfall and temperature were reported approximately 284 mm and 12 °C (ranging from −10 °C in December to 30 °C in August), respectively (EARWO, 2020). In the summer, the real evapotranspiration (ETa) ratio is high, so surface and groundwater supplies are utilized as complementary reserves alongside rainfall. The average yearly potential evapotranspiration (ETPot) is 900–1500 mm and the maximum of irrigation happens in the spring and summer (April–September) (Jalilvand et al., 2019), hence this area is extremely prone to water dearth and groundwater plays a crucial role in diverse objectives.

Zarrineh and Simineh are the most important rivers which regard as the main surface water supplies to irrigate this district that after providing this plain, bypass into Urmia Lake. The irrigation information in this plain is merely reported for the surface irrigation and there are no exact statistics about the water amount abstraction from rivers or wells.

Miandoab plain aquifer is unconfined and comprises old and recent alluvial terraces, alluvial fans, and fluvial sediments. The depth of aquifer differs from nearby a few meters in the margins of the plain to over 75 m around Miandoab city. The stream direction of groundwater is from the Southeast on the way to the Northwest. Also, in the Southeast and the Northeast of aquifer, *GWL* is high and on the way to the outlet becomes lower (Norouzi and Moghaddam, 2020).

2.2. Data description and statistical analysis

In this research, to estimate *GWL* fluctuations in the study area, 186 monthly observed *GWL* datasets of 4 different observation piezometers with different statistical attributes during time period (Sep 2001–Feb 2017) were attained from West Azerbaijan Regional Water Authority. Fig. 2 displays the position of observation piezometers in Miandoab plain along with Urmia Lake, basin, and two main rivers.

Table 1 provides descriptive statistics of *GWL* for employed piezometers in Miandoab plain which *STDV* and *CV* represent the standard deviation and coefficient of variation, respectively.

In the term of mathematical field, stationary systems are a stochastic process in which joint possibility distribution does not alter with variations of space or time (Shiri et al., 2013). Thus, on the basis of statistical parameters in Table 1, it could be deduced that datasets employed in the calibration and validation are stationary. Fig. 3 depicts the long-term time series of monthly recorded *GWL* of piezometers between Sep 2001-Feb 2017.

From Fig. 3 and Table 1, it is clear that the piezometer 2 has the most range of *GWL* during the study period.

3. Methodology

According to results of Table 1, it is fathomable that the oscillations trend of *GWL* in applied piezometers is too unbalanced and instable particularly in the observation piezometer 2, attributable to complex nature and nonlinearity by the high ranges, skewness, and kurtosis. Consequently, it is necessary to operate robust precise methods for modelling.

As aforementioned, different layer structures of BiLSTM based neural network models are developed to forecast *GWL* in 4 observation piezometers in the study area. In this direction, first of all, all datasets are normalized to zero mean and unit variance based on the recommendation of Lawrence et al. (1997). Then, the standardized recorded datasets are divided into two subgroups. The first 130 samples between Oct 2002–Sep 2012 (70% of total datasets) are applied for training of designed models (calibration) and the remaining 56 samples between

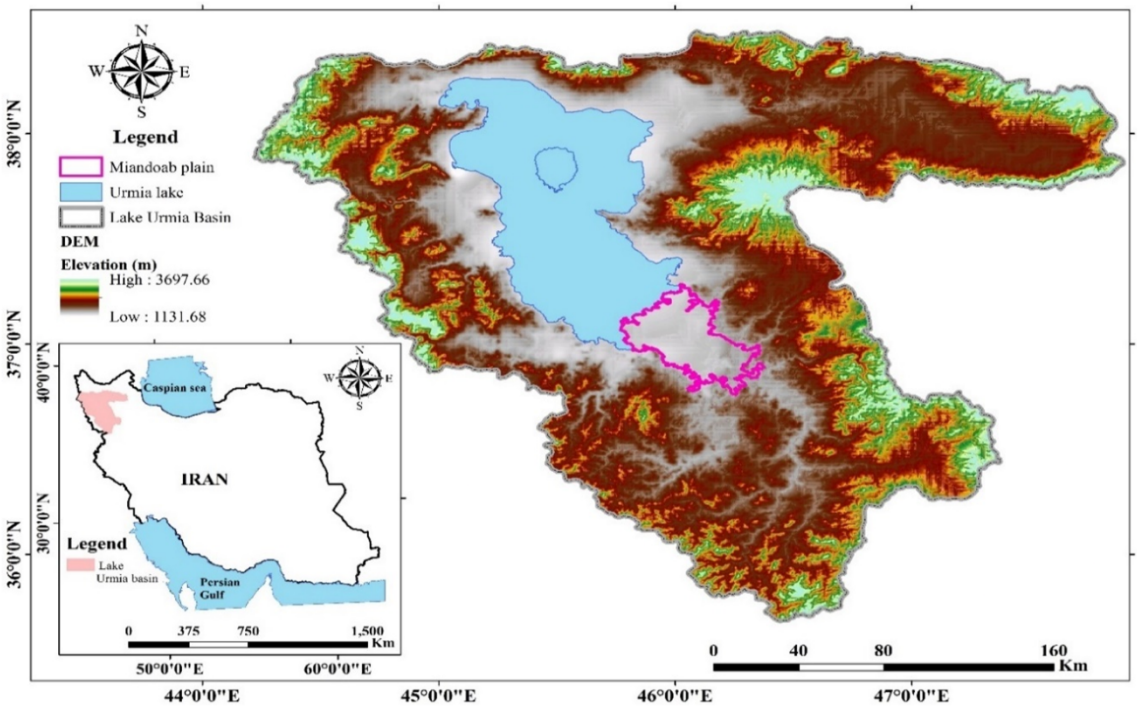

Fig. 1. Geographical position map of the study region.

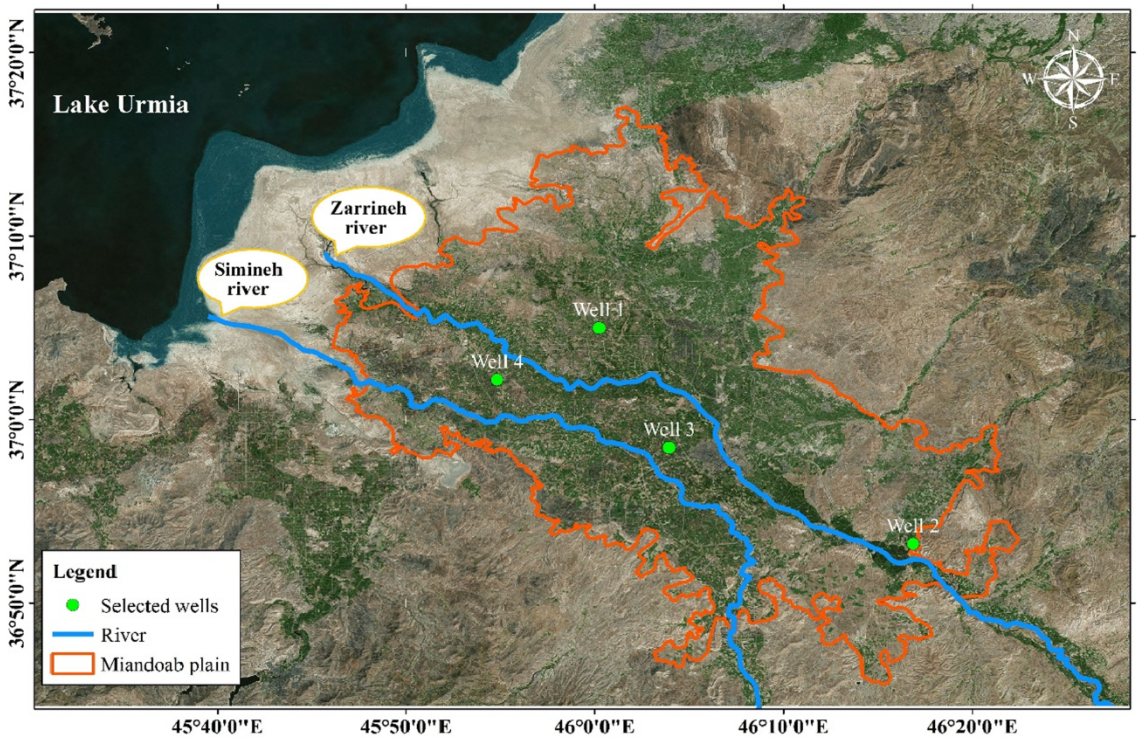

Fig. 2. Sites of observation piezometers in Miandoab plain, West Azerbaijan Province, Iran.

Table 1
Statistical parameters of *GWL* (m) the employed piezometers.

Piezometer	*Data period*	*Mean*	*Max*	*Min*	*Range*	*STDV*	*CV*	*Skewness*
1	Whole period	3.23	6.51	1.25	5.26	0.987	0.305	1.25
2	Whole period	12.55	17.21	7.92	9.29	2.21	0.18	0.17
3	Whole period	5.11	8	2.65	5.35	1.29	0.25	0.21
4	Whole period	5	7.2	2.71	4.49	0.84	0.17	−0.05

Oct 2012–Sep 2017 (30% of total datasets) are applied for testing (validation).

3.1. LSTM neural network

Recurrent Neural Network (RNN) is a kind of neural network that integrates prior information for predicting the forthcoming state of a variable by the input data with given dependencies. In RNN, the memory cell takes into consideration an unfolded loop cell which permits the prior information to pass into for forecasting of the next step. Though, its systematized mode does not allow an operative evaluation for long-term dependencies, since its learning process brings about exploding gradients in backpropagation through time (Graves, 2013).

Long Short-Term Memory (LSTM) was developed by adding hidden units to the memory cell of RNN. This memory block is designed purposefully to store information over long time periods (Hochreiter & Schmidhuber, 1997). It holds four sections: a CEC (Constant Error Carousel) cell in the company of three particular multiplicative units namely gates. CEC cell performs in a such way without using any activation function (*AF*) in the entire chain so that the gradient does not disappear when backpropagation through time is operated to train LSTM. The input, forget, and output gates control the range in which new input flows into a CEC cell. Under this process, information accumulates in a cell plus a steady gradient computation inside of the memory block, and the output of the cell flows into the rest of network (Graves and Schmidhuber, 2005). Fig. 4 demonstrates the internal memory cell and functioning mechanism of LSTM.

Under three gates in a cell state, information is processed through successive calculations by following equations (Hochreiter & Schmidhuber, 1997):

$$f_t = \sigma(W_f xt + U_f h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$C_t^{'} = tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes C_t^{'} \quad (5)$$

$$ot = softsign(W_o x_t + U_o h_{t-1} + b_o) \quad (6)$$

$$ht = o_t \otimes ReLU(C_t) \quad (7)$$

where x_t indicates the input variable at time (t), C_{t-1} is the memory unit of network in the last timestep (t-1) which makes available the prior information, C_t is the state of memory at time (t), h_t is the output of hidden state at time t, h_{t-1} is a vector which contains values of the output by each memory in the hidden layer at the previous time step (t-1) or in the primary hidden state. The forget gate, input gate, cell gate, and output gate units are symbolized by f_t, i_t, C'_t, and o_t, respectively. The memory cell learns at the time t on the basis of the input at the time t and the output at the time (t-1) in the network. W_f, W_i, W_c, and W_o are weight matrices of the forget gate, input gate, cell gate, and output gate units, respectively. b_f, b_i, b_c, and b_o are the bias vectors of the forget gate, input gate, cell gate, and output gate units, respectively. σ is the classical

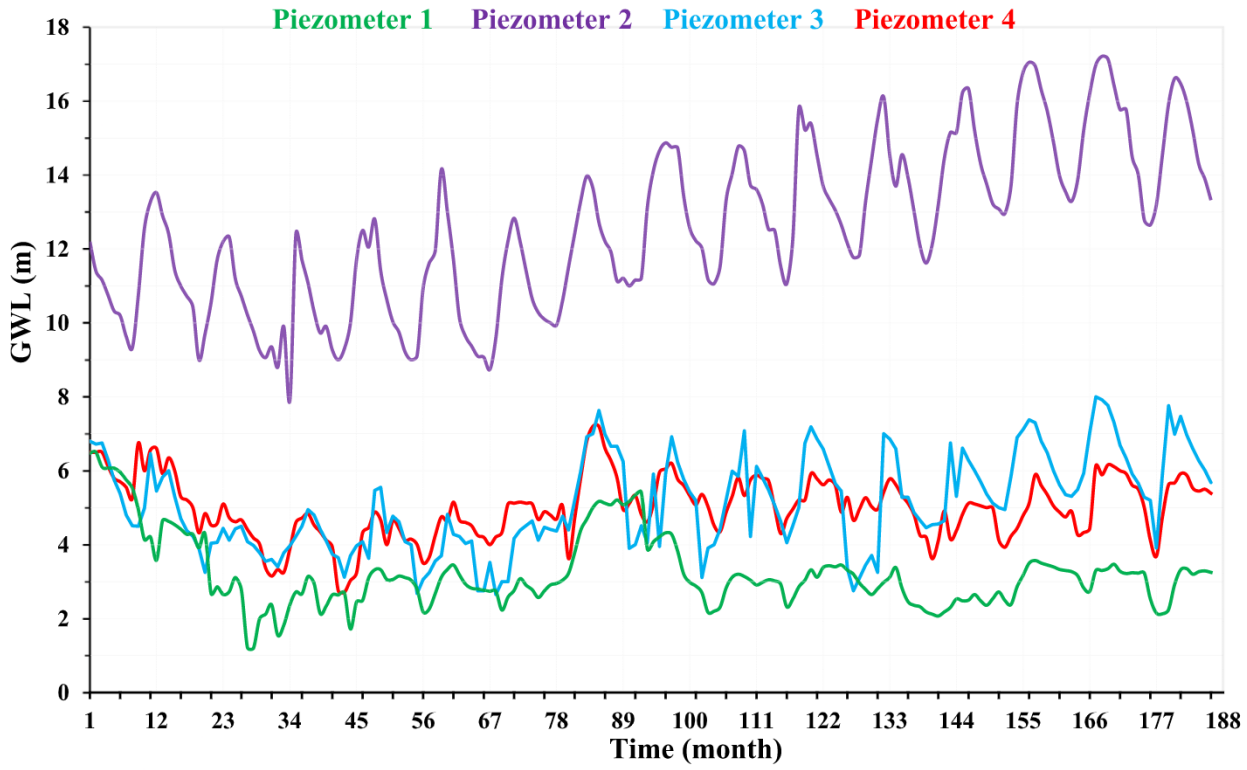

Fig. 3. Plot of time series for the observed *GWL* in applied piezometers.

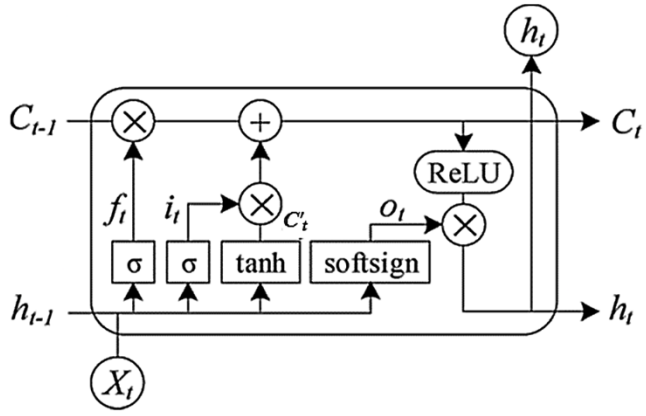

Fig. 4. Internal structure and components of LSTM memory cell.

sigmoid activation function of the forget gate and input gate units. The *tanh*, *softsign,* and *ReLU* indicate activation functions of the cell gate unit, output gate unit, and output, respectively. The element-wise multiplication of two vectors is indicated with ⊗ . All weights and bias vectors compute under the learning process correspond to the training data through optimization algorithms.

3.2. Bidirectional LSTM (BiLSTM) neural network

Graves (2013) introduced BiLSTM on the basis of different from the general one-directional LSTM which has only the backward propagating for gaining prior information in the historical variables when processing data. The main difference between BiLSTM neural network and other RNN network is that it employs several gates for performing the memory cells to control the flow of inputs and output inside the network rather than the conventional hidden layer architecture with nonlinear activation functions (Schuster and Paliwal, 1997; Graves and Schmidhuber, 2005). BiLSTM has also chain-like constructions, but repeating modules with the intricate hidden layer are more complicated compared to the conventional RNN. It comprises numerous parameters which allow it to display dynamic temporal behaviors. In BiLSTM the layer of backward LSTM was implanted to converse data and the hidden layer by generating the forward and backward information in order to the cells in the network can concurrently obtain completely the previous and next contextual information. The structure of BiLSTM is shown in Fig. 5.

There is no information about the flow between the forward and backward layers to certify the non-cycle expansion graph (Yin et al., 2020). The hidden layer of BiLSTM model requires to save two values of h_f and h_b which engage in the forward and backward computations, respectively. The ultimate output value (o_t) attains by combination the outputs of forward and backward layers which can be formulated mathematically as follows (Yin et al., 2020):

$$h_{ft} = f(w_{f1}x_t + w_{f2}h_{t-1}) \tag{8}$$

$$h_{bt} = f(w_{b1}x_t + w_{b2}h_{t+1}) \tag{9}$$

where h_f is the forward LSTM network output, h_b is the backward LSTM network output. The final output of the hidden layer is:

$$o_t = g(w_{o1} \otimes h_f + w_{o2} \otimes h_b) \tag{10}$$

In the equations (8–10), w_i is the corresponding weight coefficient matrix which is applied recurrently at each time step.

3.3. Models development

In this modelling, an algorithm-tuning process is operated to achieve a suitable configuration and improve the ability of designed models by decreasing the overfitting impact. In this direction, the type of state activation functions (*SAF*), value of dropout rate (*P-rate*), and network topology pattern (*NTP*) as hyperparameters are tuned. Since there is not a functioning instruction to set suitable hyperparameters for a model with a certain dataset, this process ponders as a difficult and time-

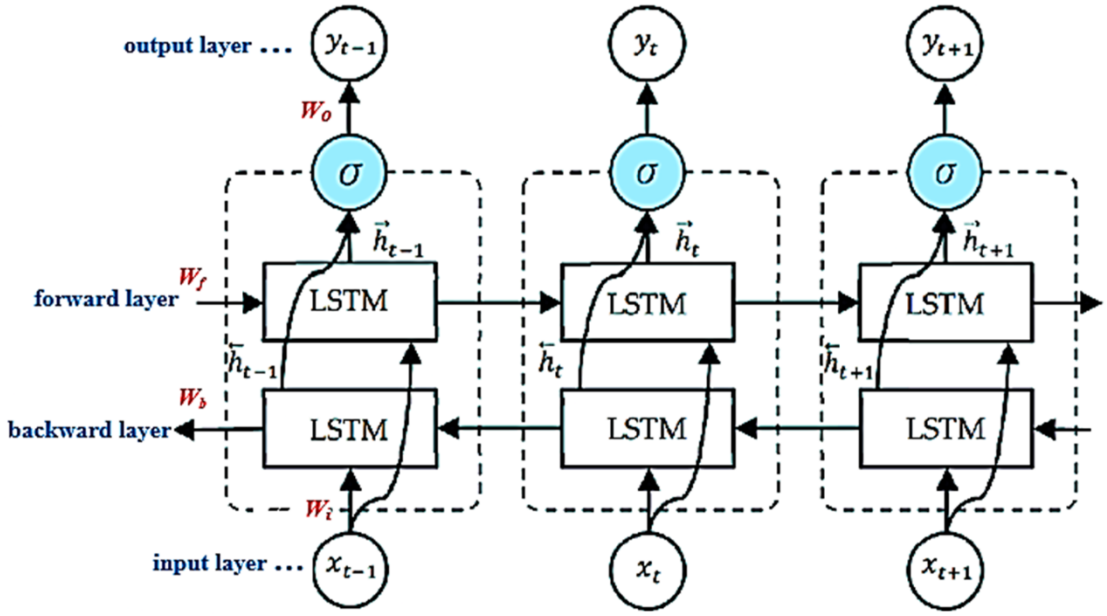

Fig. 5. BiLSTM structure diagram.

consuming task. Therefore, various scenarios must be categorized to determine an appropriate value of hyperparameters by using a trial-and-error procedure.

As advised by Maier et al. (2010), first of all, the architectural structure of DNNs models should be designed on the basis of the intent of research. Concerning, in the beginning, the general Single-LSTM neural network model (1) with 5 layers is developed by using MATLAB 2021a as the baseline model. Next, the discussers develop a Single-BiLSTM network model (2) with 5 layers. In addition, for tuning *NTP* hyperparameter in the model (2), three different layer structures of BiLSTM based models (i.e., deeper models (3), (4), and (5)) are developed. Fig. 6 (A-E) the depicts layer structure of designed models.

Model (3) is a simple Double-BiLSTM model with 6 layers, while, novel suggested models (4) and (5) are Double-BiLSTM with inserted Addition (+) and Multiplication (×) layers with 7 layers, respectively. To the best knowledge of the authors, the current study deliberates as a unique investigation that delves the application impact of Addition and Multiplication layers with the output mode of sequence on the performance of prediction using the module of sequence-to-one regression.

In the construction of designed models, Input layer inputs the time series *GWL* in the layers network models. In this direction, its size in all models is set to 1.

Addition layer adds the inputs from multiple neural network layers element-wise. All inputs to this layer should have an equal dimension that is specified automatically by software, yet it has only a single output (The MathWorks, Inc., 2021). A Multiplication layer multiplies the inputs from multiple neural network layers element-wise. The size of inputs to this layer must be either the same across all dimensions or the same across at least one dimension with other dimensions that are specified automatically by software, still, it has only a single output (The MathWorks, Inc., 2021).

DNNs are compatible for evaluating big datasets, nevertheless, very high datasets bring on overfitting. Concerning, since Dropout layer presents an operational regularization scheme, it has been operated to impede overfitting (Hinton et al., 2012). Its chief attribute is that in each training reiteration that the system is updating, a certain layer in where it has functioned, some neurons automatically discount with a given probability rate of *P* (Srivastava et al., 2014). Hence, Dropout layer help thwart the network on too much counting to given neurons in layers and diminish the neurons' co-compliance (Zhang et al., 2018). In the current modelling, for tuning the amount of *P-rate* hyperparameter, different values as 0.2, 0.5, and 0.8 are tested. By using a good tuned *P-rate*, the model learns to estimate functions via certain existing resources i.e., the number of layers and nodes per layer.

BiLSTM neural network has a high ability to learn long-period time series datasets, still, their fitting ability can be insufficient (Zhang et al., 2018), consequently a Fully Connected layer should be employed. It multiplies inputs by a weight matrix and then adds a bias vector to improve the fitting capability of model (The MathWorks, Inc., 2021). In this modelling, its input size in all designed models is set as "auto" to regulate mechanically by software and its output size is also set to positive integer 1.

The finale layer of all designed structures is selected as a Regression Output layer by which computes the "*half-mean-squared-error loss*" for regression aim through loss function as follow,

$$Loss \sum_{i=1}^{N} (y_m - y_p)^2 \tag{12}$$

where y_m and y_p are measured and predicted value of *GWL* at time *i*, respectively.

To tune *SAF*, different combinations of *softsign* and *tanh* in BiLSTM layers are applied, separately, and for tuning *NHU*, various amounts in all designed models are checked along with adopting a trial-and-error process.

To update the network weights and bias in all designed models, training options are set to "*Adam*" optimization algorithm with maximum epochs of 1000, an "*Initial Learning Rate*" of 0.001, a "*Learn Rate Drop Period*" of 125, and a "*Learn Rate Drop Factor*" of 0.2. In addition, in order to avert the gradients exploding problem, "*Gradient Threshold*" and "*Initial Batch Size*" are set to 1 and 128, respectively. Technically, these factors consider norms of all designed models which have a perceptible impact on their performance.

4. Performance evaluation metrics

In this paper, R^2 (coefficient of determination) and *RMSE* (root means square error) are employed to assess the ability of all designed models on predicting *GWL* in the study area. R^2 is applied to evaluate how deviations in a variable can be explicated with variance in a second variable when forecasting an event, which ranges between [0,1]. *RMSE* measures the precision of models and goodness-of-fit related to high

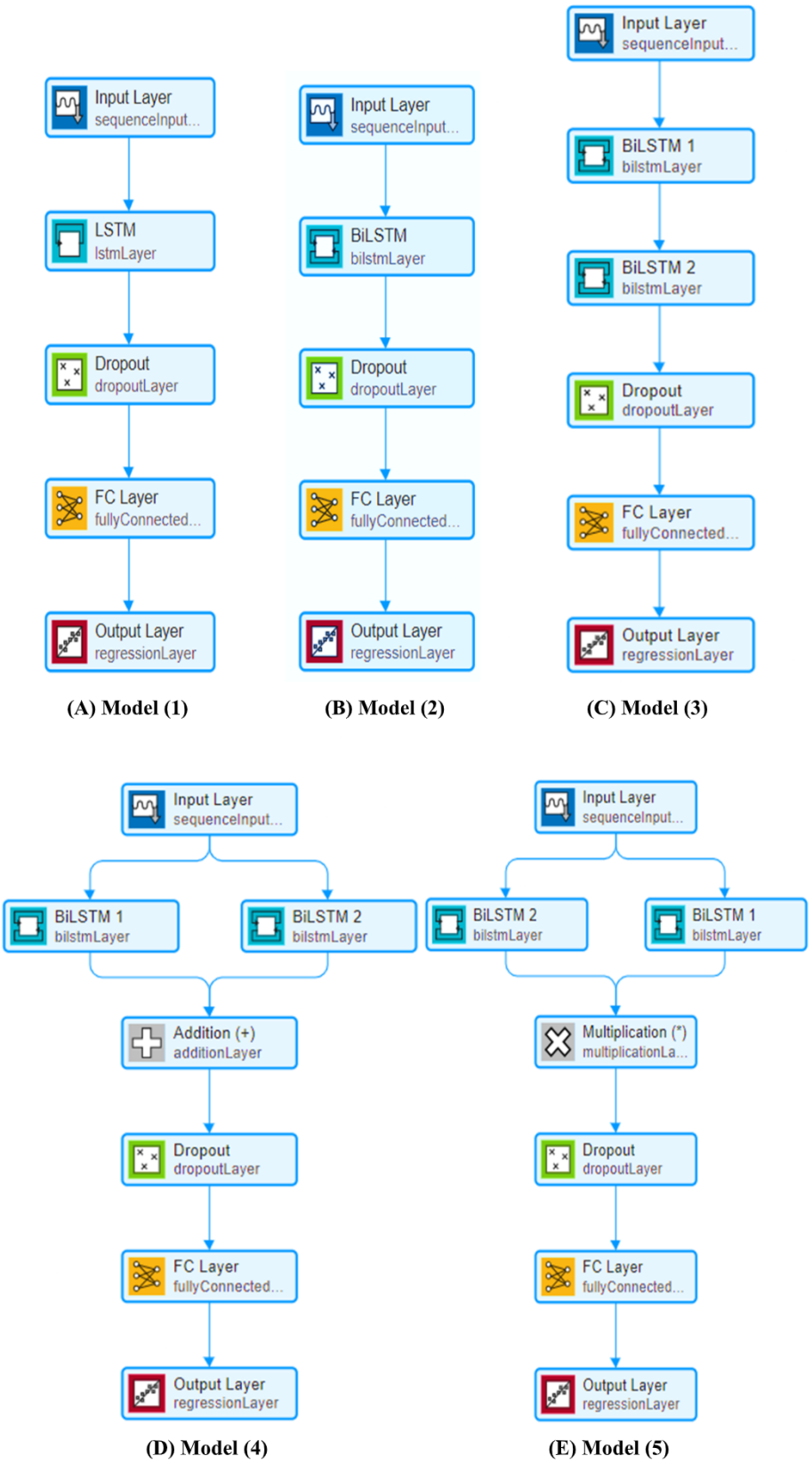

Fig. 6. Architect of LSTM and BiLSTM based layer network models: (A) Structure of the general Single-LSTM network model (1), (B) Structure of Single-BiLSTM layer network model (2), (C) Structure of Double-BiLSTM layer network model (3), (D) Structure of Double-BiLSTM layer network model combined with Addition layer (+) (model 4), (E) Structure of Double-BiLSTM layer network model combined with Multiplication layer (×) (model 5).

estimated values which generate a positive amount by squaring errors. Its scores range amongst [0, ∞]; if *RMSE* is 0, the model regards as the perfect one. R^2 and *RMSE* equations are defined as follows,

$$R^2 = \left(\frac{\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)}{N \sigma_x \sigma_y} \right)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}} \quad (4)$$

where N is the number of datasets, x_i and y_i are observed and predicted *GWL* at time i, σ_x and σ_y are standard deviation (*SD*) of observed and predicted *GWL*, respectively, μ_x and μ_y are the mean of observed and predicted *GWL*, respectively. The best values of the above-mentioned metrics are 0 and 1, respectively.

5. Results and discussion

5.1. Validation of general Single-LSTM and Single-BiLSTM models

Since determining an optimal type of *SAF*, *P-rate*, and *NHU* is a difficult task, numerous scenarios are characterized for all piezometers. After implementing, appropriate amounts of *P-rate* and *NHU* for the model (1) in all piezometers are obtained 0.5 and 100, respectively and for the model (2) are obtained 0.5 and 175, respectively. Besides, a suitable type of *SAF* for both models is achieved *tanh*. In all piezometers, final values of *Loss*, *RMSE,* and *Learning rate* for the general Single-LSTM model (1) and Single-BiLSTM model (2) in the training phase under optimal hyperparameters after 1000 reiterations are presented in Table 2.

After training, each model is tested using the validation dataset. Results of both models under optimal hyperparameters in the testing phase are presented in Table 3.

By comparison of Tables 2 and 3, it can be understood that both models in the training phase are more accurate than the testing phase, moreover, the model (2) is more precise than the model (1) in the testing phase. All in all, because both models show weak performance, they are not capable to operate for modelling *GWL* in the study area.

5.2. Results of Double-BiLSTM models

Under numerous experiments, in all piezometers, appropriate amounts of *P-rate* and *NHU* for the model (3) are determined 0.5 and 150, respectively, and for both newly proposed models (4) and (5) are determined 0.5 and 130, respectively.

As aforesaid, different combinations of *AF* including *tanh-softsign*, *tanh-tanh*, *softsign-softsign*, and *softsign-tanh* were employed in *SAF* of BiLSTM layers. In this respect, after several experiments, *tanh-softsign* combination is determined as a suitable choice in all piezometers. The phrase "*tanh-softsign*" denotes that the type of *SAF* in BiLSTM layers 1 (SAF_1) and 2 (SAF_2) are *tanh* and *softsign*, respectively. This combination brings on learning more intricate functions and not to be more prone to vanishing gradients problem during training. Technically, *softsign* activation function can intensify the speediness of model training, meanwhile, *tanh* can capture suitably complicated correlations of long-term time series. Besides, the type of data plays a significant role in deciding which combination is the best to choose.

Table 2
Final values of *Loss*, *RMSE,* and *Learning rate* in the training phase of models (1) and (2) under optimal hyperparameters.

Model	*Loss*	*RMSE(m)*	*Learning rate*
Model 1	1E−3	1E−3	1.28E−8
Model 2	1E−4	1E−4	1.28E−8

Table 3
Results of models (1) and (2) under optimal hyperparameters in the testing phase.

Piezometer	*NHU*		*RMSE(m)*		R^2	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
1	100	175	0.44	0.36	0.61	0.71
2	100	175	0.65	0.51	0.46	0.59
3	100	175	0.58	0.42	0.55	0.65
4	100	175	0.39	0.28	0.69	0.76

In all piezometers, final values of *Loss*, *RMSE*, and *Learning rate* for Double-BiLSTM models (3, 4, and 5) in the training phase after 1000 reiterations under optimal hyperparameters are presented in Table 4.

After training, the ability of these models is tested using the validation dataset. Results of optimal Double-BiLSTM models in the testing phase under optimal hyperparameters are presented in Table 5.

As said in Table 5, it can be discerned that all models in the training phase are more accurate than the testing phase, moreover, the model (5) is the most accurate one.

5.3. Performance comparison of models used

In the present modelling, to assess the performance of all designed models, discussers accentuate their substructure and capacity. Regarding, MATLAB 2021a presents a remarkable factor namely, *TLP* (Total Learnable Parameters), as a crucial criterion to specify the actual capacity of models. Explicitly, a model with proper *NTP* and an optimal *NHU* yield in a well-proportioned *TLP* and a capable system. In this direction, Table 6 is provided to represent characteristics of designed models under optimal hyperparameters for all piezometers in the testing phase.

Based on Table 6, it can be concluded that due to the model (1) holds least *TLP*, it has too little capacity and has underfitted in all piezometers. Therefore, it is not capable to forecast *GWL* fluctuations in the study area.

Besides, as can be seen from Table 6, increasing the number of BiLSTM layer in the model (2) with the same *P-rate*, caused to improve noticeably the accuracy of forecasting in all piezometers. As such, the model (3) in comparison with model (2) under optimal hyperparameters causes to decrease in the value of *RMSE* by approximately 7.2% in all piezometers.

The model (3) owing to the highest *TLP*, weights up undoubtedly as the most capacity model and is also expected spontaneously as the best and accurate one. But, consistent with two-evaluation metrics, it is observed that newly proposed models (4) and (5) outperform and predict more accurately than the model (3) in the same *P-rate* and *SAF*. The main reason for this nonconformity can be justified based on the value of *TLP*. In practice, due to the over extra *TLP* value in the model (3), it has memorized the training dataset, meaning it has slightly overfitted or has gotten stuck and lost during the optimization process. On the other hand, according to the value of *TLP* in newly proposed models (4) and (5), it can be construed that inserting Addition and Multiplication layers in the model (3) induces to create a good-adjusted *TLP* value. Hence, they have a greater representational capability and learning ability to approximate mapping functions.

Table 4
Final values of *Loss*, *RMSE,* and *learning rate* of Double-BiLSTM models in the training phase under optimal hyperparameters.

Model	*Loss*	*RMSE(m)*	*Learning rate*
Model 3	1E−4	1E−4	1.28E−8
Model 4	1E−5	1E−5	1.28E−8
Model 5	1E−5	1E−5	1.28E−8

Table 5
Results of Double-BiLSTM models under optimal hyperparameters in the testing phase.

Piezometer	*RMSE(m)*			R^2		
	Model (3)	Model (4)	Model (5)	Model (3)	Model (4)	Model (5)
1	0.28	0.21	0.19	0.74	0.82	0.84
2	0.34	0.29	0.27	0.61	0.72	0.75
3	0.30	0.23	0.20	0.70	0.80	0.83
4	0.25	0.19	0.17	0.77	0.86	0.89

Table 6
Statistical indices values for designed models under optimal hyperparameters in all piezometers.

Piezo meter	Model	*P-rate*	*NHU*	*RMSE(m)*	R^2	*TLP*
1	1	0.5	100	0.44	0.61	40,901
1	2	0.5	175	0.36	0.71	248,151
1	3	0.5	150	0.28	0.74	723,901
1	4	0.5	130	0.21	0.82	274,821
1	**5**	0.5	**130**	**0.19**	**0.84**	274,821
2	1	0.5	100	0.65	0.46	40,901
2	2	0.5	175	0.51	0.59	248,151
2	3	0.5	150	0.34	0.61	723,901
2	4	0.5	130	0.29	0.72	274,821
2	**5**	0.5	**130**	**0.27**	**0.75**	274,821
3	1	0.5	100	0.58	0.55	40,901
3	2	0.5	175	0.42	0.65	248,151
3	3	0.5	150	0.30	0.7	723,901
3	4	0.5	130	0.23	0.8	274,821
3	**5**	0.5	**130**	**0.20**	**0.83**	274,821
4	1	0.5	100	0.39	0.69	40,901
4	2	0.5	175	0.28	0.76	248,151
4	3	0.5	150	0.25	0.78	723,901
4	4	0.5	130	0.19	0.86	274,821
4	**5**	0.5	**130**	**0.17**	**0.89**	274,821

Above and beyond, on the basis of Table 6, the optimal performance of prediction is achieved in the piezometer 4 on account of better statistical indices. We justify it owing to the least range of *GWL* fluctuations in the piezometer 4 compared to other ones, which connotes the high stability and steadiness of water table depth.

Results of models under optimal hyperparameters with respect to the running time in the piezometer 4 are presented in Fig. 7.

Based on Fig. 7, it can be inferred that on account of the highest *TLP* in the model (3), it has permitted to learn more and even globally optimum weight sets, and accordingly has taken longer outstandingly to train. Whilst, due to the lowest *TLP* in the model (1), it has brought forth to learn faster sub-optimal weight sets. As can be observed from Fig. 7, by increasing *P-rate* in the same model, the running time decreased, moreover, because the lowest running time in the model (2) among Double-BiLSTM models, it considers the fastest model.

Fig. 8 displays measured and predicted *GWL* by models (5) and (3) under optimal hyperparameters in the piezometer 4, respectively. For concision, only results of the piezometer 4 are presented.

As seen from Fig. 8, owing to the high ability of model (5), it captures fittingly variations trend of the measured *GWL*, nonetheless, the model (3) has gone belly in up and down, especially in the peak and bottom points, that indicates further deviations from the measured *GWL*.

By more attention in Tables 5 and 6 and Fig. 8, it can be distinguished that there is not a substantial statistical difference between newly proposed models (4) and (5) under optimal hyperparameters. Scilicet, these Tables and Figure are not able to provide any useful information such as full distribution of predicted dataset and relative amplitude in the testing stage. Accordingly, a violin plot is employed to determine a superior model as is shown in Fig. 9.

According to Fig. 9, based on the mean and maximum *GWL* in models (4) and (5), it can be deduced that there is no significant deviation among these models in comparison with the observed model. Although both models (4) and (5) overestimated the minimum *GWL* compared to the observed one, the model (5) outperformed to some extent the model (4).

6. Summary and conclusion

In this study, the discussers were interested in forecasting *GWL* fluctuations of four different piezometers with different statistical features during the time period (Sep 2001–Feb 2017) in Miandoab plain. Regarding, different layer structure of BiLSTM based models for accurate predicting *GWL* fluctuations with time series characteristics were designed. For reducing overfitting effects, the algorithm tuning along with a trial-and-error process were adopted. The major results of modelling are as follows:

1. All models in the training phase were more accurate than the testing phase. Moreover, under numerous experiments, in all piezometers, appropriate amounts of *P-rate* were determined 0.5. Besides, the

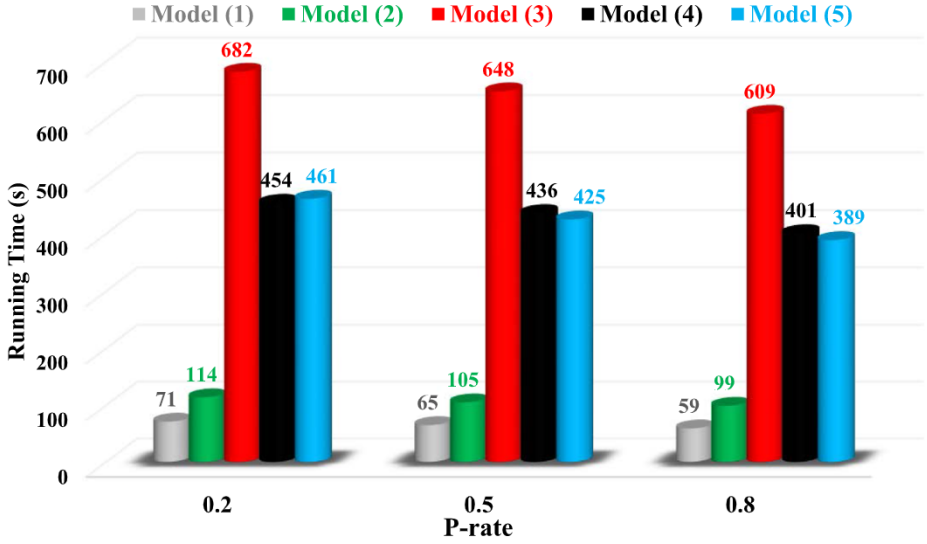

Fig. 7. Results of models under optimal hyperparameters with respect to the running time in the piezometer 4.

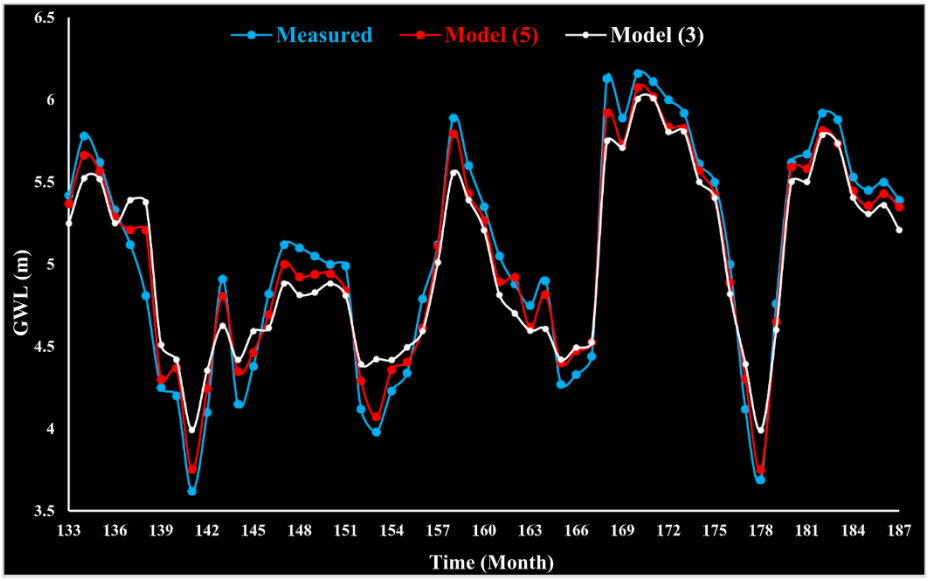

Fig. 8. Comparison between measured and predicted *GWL* by models (5) and (3) under optimal hyperparameters in the piezometer 4.

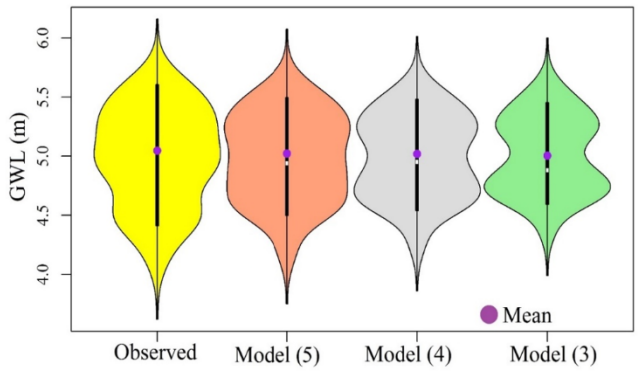

Fig. 9. Violin plot of newly proposed models (4) and (5) under optimal hyperparameters in the piezometer 4.

optimal performance of prediction is achieved in the piezometer 4 owing to the least range of *GWL* fluctuations compared to other ones.

2. On account of severe vacillations of *GWL* in the applied observation piezometers during the study period, the general Single-LSTM and the simple Single-BiLSTM layer network showed weak performance in comparison with other models. However, by deepening simple Single-BiLSTM model via effective relevant hidden layers instead of many numbers of nodes, its performance strikingly improved. Thus, it can be concluded that boosting the capacity of simple Single-BiLSTM based model via expanding depth was confirmed as a successful scheme.
3. In spite of the highest *TLP* in the model (3), the newly proposed models (4) and (5) outperformed and predicted more accurately than the model (3) under optimal hyperparameters. Accordingly, it can be concluded that an unbalanced capacity or *TLP* caused to reduce the performance of model due to the overfitting condition.
4. Using nonidentical nonlinear *AF* in the hidden layers of designed Double-BiLSTM structures especially *tanh-softsign* combination, caused models not to be more vulnerable to the vanishing gradients problem. Likewise, this scheme engendered to learn more complicated functions, and consequently, better results in high ranges and standard deviations value of *GWL* in the training stage.
5. Inserting Multiplication and Addition layers in the simple Double-BiLSTM based model yielded a model with a good-balanced *TLP* value, and consequently, a better capability and learning ability.
6. By increasing *P-rate* in the same model, the running time decreased, moreover, because the lowest *TLP* in the model (2) among Double-BiLSTM models, it considers the fastest model.
7. Since the operated framework and structure of BiLSTM based models were developed in missing conditions of meteorological parameters records, so these models can be directly utilized for diverse climatic districts in the world. Thereby, the employed methodology and data management could be transferable to other regions.

The suggested unique structure models (4) and (5) by discussers are different from conventional neural network models, since they have not been frequently operated in the field of hydrology. Ultimately, the discussers prefer the newly proposed model (5) as the most superior and reliable model for forecasting time-based parameters with sufficient accuracy. It presents an encouraging novel model with reasonable performance to predict *GWL* fluctuations under different circumstances of water resources management. It is suitable for districts wherein complicated meteorological and hydrogeological parameters are missing or difficult to acquire, and accordingly, physical interactions of the surface–groundwater are indefinite. Besides, it could be employed for any type of time series forecasting task.

In the end, the discussers introduce *TLP* value as a key measure and decisive norm to judge on predicting the ability and performance of designed DL-based models, as it shows the actual capacity and occurrence possibility of the overfitting condition.

Data availability statement

Computations and questionnaires details for the applied models are available from the corresponding author by reasonable request. In addition, data and information related to the experimental operations and management used in the current paper are not proprietary or confidential and may be provided.

CRediT authorship contribution statement

Redvan Ghasemlounia: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Amin Gharehbaghi:** Methodology, Software, Writing – original draft. **Farshad Ahmadi:** Writing –

review & editing. **Hamid Saadatnejadgharahassanlou:** Writing – review & editing.

Declaration of Competing Interest

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

References

Abdelhameed, A.M., Daoud, H.G., Bayoumi, M., 2018. Deep convolutional bidirectional LSTM recurrent neural network for epileptic seizure detection. June. In: 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS). IEEE. pp. 139–143.

Bai, P., Liu, X., Xie, J., 2021. Simulating runoff under changing climatic conditions: a comparison of the long short-term memory network with two conceptual hydrologic models. J. Hydrol. 592, 125779. https://doi.org/10.1016/j.jhydrol.2020.125779.

Boithias, L., Sauvage, S., Lenica, A., Roux, H., Abbaspour, K., Larnier, K., Dartus, D., Sánchez-Pérez, J., 2017. Simulating flash floods at hourly time-step using the SWAT model. Water 9 (12), 929. https://doi.org/10.3390/w9120929.

Bowes, B.D., Sadler, J.M., Morsy, M.M., Behl, M., Goodall, J.L., 2019. Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks. Water 11 (5), 1098.

Byeon, W., Breuel, T.M., Raue, F., Liwicki, M., 2015. Scene labeling with LSTM recurrent neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3547–3555.

Chang, F.-J., Chang, L.-C., Huang, C.-W., Kao, I.-F., 2016. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. J. Hydrol. 541, 965–976.

Dikshit, A., Pradhan, B., Alamri, A.M., 2021. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. Sci. Total Environ. 755, 142638. https://doi.org/10.1016/j.scitotenv.2020.142638.

EARWO (East Azerbaijan Regional Water Organization), 2020. Preparation of water balance and water cycle in the Malekan region. 56p.

Fallah-Mehdipour, E., Bozorg Haddad, O., Mariño, M.A., 2013. Prediction and simulation of monthly groundwater levels by genetic programming. J. Hydro-Environ. Res. 7 (4), 253–260.

Feng, Z., Wang, Y., Peng, L., Hong, H., 2020. Predicting flood susceptibility using long short-term memory (LSTM) neural network model. J. Hydrol. https://doi.org/10.1016/j. jhydrol.2020.125734.

Feng, S., Kang, S., Huo, Z., Chen, S., Mao, X., 2008. Neural networks to simulate regional groundwater levels affected by human activities. Ground Water 46 (1), 80–90.

Gharehbaghi, A., 2016. Explicit and implicit forms of differential quadrature method for advection– diffusion equation with variable coefficients in semi-infinite domain. J. Hydrol. 541 (B), 935–940. https://doi.org/10.1016/j.jhydrol.2016.08.002.

Gharehbaghi, A., 2017. Third- and fifth-order finite volume schemes for advection–diffusion equation with variable coefficients in semi-infinite domain. Water Environ. J. 31 (2), 184–193. https://doi.org/10.1111/wej.2017.31.issue-210.1111/wej.12233.

Gharehbaghi, A., Kaya, B., Saadatnejadgharahassanlou, H., 2017. Two-dimensional bed variation models under non-equilibrium conditions in turbulent streams. Arab. J. Sci. Eng. 42, 999–1001. https://doi.org/10.1007/s13369-016-2258-4.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 18 (5-6), 602–610.

Guzman, S.M., Paz, J.O., Tagert, M.L.M., 2017. The use of NARX neural networks to forecast daily groundwater levels. Water Resour. Manag. 31, 1591–1603.

Hochreiter, S., Schmidhuber, J., 1997. LSTM can solve hard long time lag problems. Adv. Neural Inform. Process. Syst. 473–479.

Jalilvand, E., Tajrishy, M., Ghazi Zadeh Hashemi, S.A., Brocca, L., 2019. Quantification of irrigation water using remote sensing of soil moisture in a semi-arid region. Remote Sens. Environ. 231, 111226. https://doi.org/10.1016/j.rse.2019.111226.

Jeong, J., Park, E., 2019. Comparative applications of data-driven models representing water table fluctuations. J. Hydro. 572, 261–273.

Jeong, J., Park, E., Chen, H., Kim, K.Y., Han, W.S., Suk, H., 2020. Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. J. Hydrol. 582, 124512.

Kim, N.W., Chung, I.M., Won, Y.S., Arnold, J.G., 2008. Development and application of the integrated SWAT–MODFLOW model. J. Hydrol. 356 (1-2), 1–16.

Lee, T., Shin, J.-Y., Kim, J.-S., Singh, V.P., 2020. Stochastic simulation on reproducing long-term memory of hydro-climatological variables using deep learning model. J. Hydrol. 582, 124540. https://doi.org/10.1016/j.jhydrol.2019.124540.

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environ. Model. Softw. 25 (8), 891–909.

MATLAB User's Guide, 2021a. The MathWorks Inc. (Deep Learning Toolbox). Natick, Massachusetts, United State; Computer Software. www.mathworks.com/.

Mohanty, S., Jha, M.K., Kumar, A., Panda, D.K., 2013. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India. J. Hydrol. 495, 38–51.

Norouzi, H., Moghaddam, A.A., 2020. Groundwater quality assessment using random forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). Arab. J. Geosci. 13 (18), 1–13.

Pelosi, A., Medina, H., Villani, P., D'Urso, G., Chirico, G.B., 2016. Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system. Agric. Water Manag. 178, 106–118.

Raghavendra, N.S., Deka, P.C., 2016. Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS. In: Advanced Computing and Systems for Security, Volume 396 of the Series Advances in Intelligent Systems and Computing, pp. 289–302.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.

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

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn Res. 15, 1929–1958.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inform. Process. Syst. 3104–3112.

Todd, D.K., Mays, L.W., 2005. Groundwater Hydrology. Third Revision John Wiley and Sons Inc., p. 636

Lawrence, S., Back, A.D., Tsoi, A.C., Giles, C.L., 1997. On the distribution of performance from multiple neural network trials. IEEE Trans. Neural Net. 8 (6), 1507–1517.

Bengio, Y., 2007. Learning deep architectures for Al. Found. Trends Mach. Learn. 2 (1), 1–127.

Vu, M.T., Jardani, A., Massei, N., Fournier, M., 2021. Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. J. Hydrol. 597, 125776. https://doi.org/10.1016/j.jhydrol.2020.125776.

Yin, J., Deng, Z., Ines, A.V.M., Wu, J., Rasu, E., 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (BiLSTM). Agric. Water Manag. 242, 106386. https://doi.org/10.1016/j.agwat.2020.106386.

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol. 396 (1-2), 128–138.

Zare, M., Koch, M., 2018. Groundwater level fluctuations simulation and prediction by ANFIS-and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: application to the Miandarband plain. J. Hydro-environ. Res. 18, 63–76.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929.

Zhou, X.-L., Huang, K.-Y., Wang, J.-H., 2017. Numerical simulation of groundwater flow and land deformation due to groundwater pumping in cross-anisotropic layered aquifer system. J. Hydro-environ. Res. 14, 19–33.

Graves, A., 2013. Generating Sequences with Recurrent Neural Networks. arXiv preprint arXiv 1308.0850.

Jalali, S.M.J., Ahmadian, S., Khosravi, A., Shafie-khah, M., Nahavandi, S., Catalao, J.P., 2021a. A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting. IEEE Trans. Ind. Inform.

Jalali, S.M.J., Ahmadian, S., Kavousi-Fard, A., Khosravi, A., Nahavandi, S., 2021. Automated deep cnn-lstm architecture design for solar irradiance forecasting. IEEE Trans. Syst. Man Cybernet.: Syst.

Jalali, S.M.J., Ahmadian, S., Khodayar, M., Khosravi, A., Ghasemi, V., Shafie-khah, M., Nahavandi, S., Catalão, J.P., 2021b. Towards novel deep neuroevolution models: chaotic levy grasshopper optimization for short-term wind speed forecasting. Eng. Comput. 1–25.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.