



Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models

Saber Kouadri¹ · Chaitanya B. Pande^{2,3} · Balamurugan Panneerselvam⁴ · Kanak N. Moharir³ · Ahmed Elbeltagi^{5,6}

Received: 18 June 2021 / Accepted: 13 October 2021 / Published online: 8 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Forecasting the irrigation groundwater parameters helps plan irrigation water and crop, and it is commonly expensive because it needs various parameters, mainly in developing nations. Therefore, the present research's core objective is to create accurate and reliable machine learning models for irrigation parameters. To accomplish this determination, three machine learning (ML) models, viz. long short-term memory (LSTM), multi-linear regression (MLR), and artificial neural network (ANN), have been trained. It is validated with mean squared error (MSE) and correlation coefficients (r), root mean square error (RMSE), and mean absolute error (MAE). These machine learning models have been used and applied for predicating the six irrigation water quality parameters such as sodium absorption ratio (SAR), percentage of sodium (%Na), residual sodium carbonate (RSC), magnesium hazard (MH), Permeability Index (PI), and Kelly ratio (KR). Therefore, the two scenario performances of ANN, LSTM, and MLR have been developed for each model to predict irrigation water quality parameters. The first and second scenario performance was created based on all and second reduction input variables. The ANN, LSTM, and MLR models have discovered that excluding for ANN and MLR models shows high accuracy in first and second scenario models, respectively. These model's accuracy was checked based on the mean squared error (MSE), correlation coefficients (r), and root mean square error (RMSE) for training and testing processes serially. The RSC values are highly accurate predicated values using ANN and MLR models. As a result, machine learning models may improve irrigation water quality parameters, and such types of results are essential to farmers and crop planning in various irrigation processes.

Keywords ANN · LSTM · MLR · Irrigation water quality · Correlation

Introduction

Water is a general requirement to plant and living organisms on the earth's surface. It is essential for maintaining the balance of ecology, atmosphere, and natural resources. While natural resources are most important for whole natural life systems on the ground, harmless water must not encompass any harmful chemical materials or living bacteria in concentrations that affect impairment (WHO 2017). Growth and development in the world have led to extensive pollution from rainwater outlets like rivers (UNEP 2016). Many factors can affect the chemical, physical, and biotic substances of surface water, for example natural (i.e. rainfall, watershed geography, weather, geology) and anthropogenic activities (i.e. industrial activities, domestic, agricultural run-off) (Mishra et al. 2017; Ewaid et al. 2018; Su et al. 2018). The category and concentration of dissolved salts and solids have been identified by irrigation water quality (Mirabbasi et al. 2008; Ramakrishnaiah et al.

Responsible Editor: Marcus Schulz

✉ Chaitanya B. Pande
chaitanay45@gmail.com

¹ Laboratory of Water and Environment Engineering in Sahara Milieu (GEEMS), Department of Civil Engineering and Hydraulics Faculty of Applied Sciences, Kasdi Merbah University Ouargla, Ouargla, Algeria

² CAAST-CSAWM, MPKV Rahuri, Rahuri, India

³ Sant Gadge Baba Amravati University, Amravati, India

⁴ Department of Civil Engineering, M. Kumarasamy College of Engineering, Karur, Tamil Nadu, India

⁵ Agricultural Engineering Dept., Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

⁶ College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

2009; Shayannejad et al. 2020; and Golian et al. 2020). The groundwater quality monitoring system included plans and activities for assessing the quality of water resources and the performance of the pollution reduction system (Tavakol et al. 2017; Shayannejad et al. 2021). Agriculture is the most critical consumer of fresh water in the semi-arid region, as irrigation practices more than two-thirds of the world's accessible freshwater resources (Aliyu et al. 2017). The various climate change factors and manufacturing activities directly affected the water quality parameters, in which water quality is most affected due to activities. The regular monitoring of the drinking and irrigation water quality parameters is essential in sustainable development in groundwater and agricultural parts. Several approaches have been used to estimate the consequence of irrigation water on plants and soils. Various scientists and scholars have used different irrigation indices like SAR, Na%, KI, PI (Cieszynska et al. 2012; Fakhre 2014). The agricultural sector is also the major employer of water, accounting for 80% of the overall consumption and a source of water pollution. To maintain sustainable agriculture, strategic water planning using a reasonable cost for irrigation is needed. The irrigation water quality index (IWQI) has been created using the parameters defined by FAO guidelines 29 (Ayers and Westcot 1999). Machine learning (ML) models are significant and convenient accompaniments and replacements in water index and water quality forecast (Chang et al. 2017). In general, ML models are more concerned with the association of mapping between a system's inputs and outputs than with complex process mechanisms. The highly nonlinear relations can be reliably computed with or without prior information for the studied system by learning from a vast amount of historical information that contains the dynamic evolution mechanism. In this regard, various machine learning techniques, such as an artificial neural network, have been successfully developed for algal prediction (ANN) (Recknagel et al. 2002; Chang et al. 2017; Tian et al. 2017). The analysis of the statistical-based method is functional mainly to create the WQI. Utilized in Egypt, a multivariate analysis was used to establish the IWQI for surface water (Jahin et al. 2020). The findings show that water quality can be regulated quickly and cheaply using principal component analysis (PCA) and factor analysis (FA). The analysis of the PCA and Hierarchy of the Al-Gharaf River (CA) in Iraq examines potential pollution sources. Decision-makers may use these findings to reduce the number of samples analyzed and prioritize measures to enhance the river's efficiency (Ewaid and Abed 2017). While all these models were developed and considered effective instruments to evaluate IWQ, their application requires significant parameters and studies considering the cost and

time of the study, and their applicability. Therefore, the model based on the prediction can be used by farmers to optimize water quality assessments (Ewaid et al. 2019).

Long short-term memory (LSTM) is famous in deep learning models. The LSTM is a recurrent neural network (RNN) that gathers prolonged serial data in the hidden memory for processing, demonstration, and storage. It also upgrades over time to guarantee that information is appropriate. LSTM was used by Kratzert et al. (2018) to model the rainfall-runoff flow in 241 watersheds. Specific DL approaches have been applied to forecast water quality factors in particular. For example, Liu et al. (2019) predicted drinking-water quality through LSTM deep neural network in Yangzhou, China. Multiple neural network models estimate the level of water of an integrated sewage outflow architecture, namely LSTM and a gated recurrent network (GRU) used by Zhang et al. (2018). LSTM was used by Kratzert et al. (2018) to model the rainfall-runoff flow in 241 watersheds. Specific DL approaches have been applied to forecast water quality factors in particular. For example, Liu et al. (2019) predicted drinking-water quality through LSTM deep neural network in Yangzhou, China.

In addition, some models are currently developed to predict the availability of groundwater for agriculture irrigation purposes in a dry area. However, Ewaid et al. (2018) have created a model for quick prediction in the irrigation water and water quality indices and multiple linear regression (MLR) for agriculture, industry, and drinking purposes of the Tigris River in Iraq. The machine learning models have created five groundwater quality parameters. The regression analysis is the statistical technique used in the secondary relationship variable, i.e. the dependent variable, to establish independent variables. Dependent and fully independent variables are being referred to as responding and predictor variables. The regression model shall adequately define a dependent variable, predict it, and control it based on separate variables. Several authors investigated and applied regression analyses (Charulatha et al. 2017; Noori et al. 2015, 2017). The major focus of this paper is to (1) explore the capability of ML models, namely the ANN and MLR models. To compare LSTM models with predicting irrigation water quality parameters in two scenarios (sodium absorption ratio (SAR), percentage of sodium (%Na), residual sodium carbonate (RSC), magnesium hazard (MH), Permeability Index (PI), Kelly ratio (KR)). (2) To compare the performances of the LSTM models with ANN and MLR for short-range prediction of irrigation water quality factors, and (3) to develop two scenarios (ANN, LSTM, and MLR) that show which models are best in both scenarios. Also, to the greatest of the authors' knowledge, such a type of study, done for the first time, two scenarios have been utilized to predict irrigation water quality parameters. This research outcome is working to help the

farmers, crop system, and groundwater development in the semi-arid nations to the prediction of irrigation water quality within a quick time with low-cost results.

Study area

The Akot basin area is situated in the Akot Taluka of Akola district of Maharashtra, between 20°54'30" and 21°14'35" N latitudes and between 76°48' and 77°00'3'E longitudes with 450 sq. km. This study area's minimum and maximum temperatures are 12.6°C and 42.4°C (Fig. 1). The observed annual rainfall is 740 to 860 mm. The deep black is soil found in the southern part of the Akot basin. This soil has a deep, heavy colour with an angular block structure in the sub-surface horizon, medium drained and low to moderate water support. This basin is under the saline water zone because most the groundwater has very highly salted water found within the basin area. In this view, most of the farmers are suffering from so many groundwater quality issues.

Farmers who use the poor quality of irrigation groundwater have experienced direct decrease in crop production and soil fertility in the basin. The Purna River Alluvium, which covers Akot and Telhara talukas as well as the northern sections of Akola and Balapur talukas, is afflicted by inland saline issues as well as drought and water level drops. During exploratory drilling, operations in hard rock sections of the Akola district, encountered a wide range of topics, the most common of which were caving formations (red bole) and drilling medium loss (Khadri et al. 2013; Khadri and Pande 2015a, 2015b; Pande et al. 2019a).

Methodology

Machine learning models were currently used to estimate most groundwater quality variables precisely and show their effectiveness (Rahgoshay et al. 2018; Ho et al. 2019). One hundred groundwater samples are obtained from this prediction model. The dataset was collected from observation

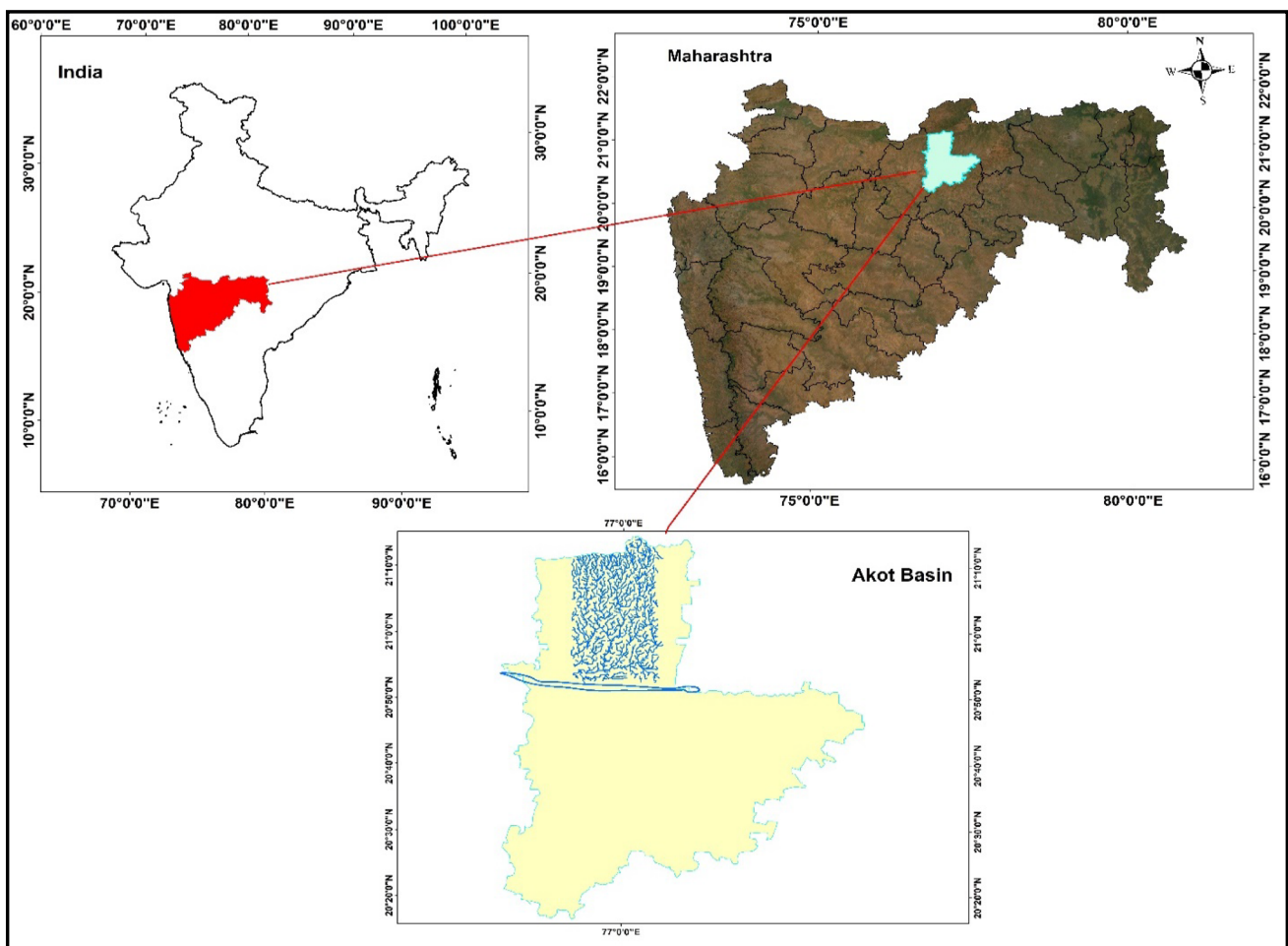


Fig. 1 Location map of study area

wells within the basin area. We have used 140 water samples for this model. We used 70% of data in training (98 samples) and 15 % of data used in ANN model validation. This research uses 15 % and 30 % data for ANN, MLR, and LSTM models' prediction purposes. This study has developed three machine learning models to predict irrigation water quality parameters, specifically ANN, LSTM, and MLR models. Therefore, three machine learning models, LSTM, MLR, and ANN, were selected for both prediction scenarios. The first and second scenarios of MLR, LSTM, and ANN models have been based on all input and reduction variables, respectively. These ML models of the first scenario show the very high accuracy of prediction of irrigation water quality parameters for the Akot basin (Saline tract) compared to the second scenario. ANN, LSTM, and MLR models were applied to the actual dataset in both scenarios. The compiled dataset was divided into two processes at this stage: training and testing. The developed model of ANN, LSTM, and MLR was validated with all the results and errors. After creating the above-stated ML models, model performance has been calculated based on matching with actual and predicted water quality data for every model. The accuracy of each developed model was checked by the mean squared error (MSE), correlation coefficients (r), root mean square error (RMSE), and mean absolute error (MAE) (Fiyadh et al. 2019).

Predicting modelling should be considered to model the past data collection, and the ever-changing dataset is not corrected. At the same time, it is permissible to use the knowledge for function fit (finding the internal relationships of the components). The neural network system for the two ANN models of prediction of water quality is created. Developed models have been framed by the hidden layers, the number of nodes in each layer, and the kind of transmission function. The various irrigation water quality variables, namely KI, MH, PI, RSC, SAR, and SSP, are widely used to classify irrigation water quality. PCA and correlation analysis was carried out using the computer-aided software package SPSS software, which is used to understand irrigation water quality status. Water quality data were collected from observation wells in the study area and divided into two scenarios by the fraction of training and testing datasets used in model development.

Artificial neural network

In recent years, scientists, researchers, and decision-makers in the field of water quality management are highly applying neural network modelling to identify the source of pollution, clear view of the quality of groundwater for specific purposes and watershed management (Ostad-Ali-Askari et al. 2017; Yıldız and Karakuş 2019; Vasanthi and Kumar 2019; El Baba et al. 2020; Ostad-Ali-Askari and Shayan

2021). The artificial neural network was first time developed (McCulloch and Pitts 1943). In the view prediction of water quality modelling part, physical and chemical characteristics of groundwater are the input of the model and predict the quality for future years. In previous studies, mathematical modelling was used to predict the KR, MH, PI, RSC, SAR, and SSP for irrigation use purposes in the semi-arid region. Among the various methods, ANN gives a more accurate and efficient manner for predicting and analyzing the vast dataset. The basic form of ANN consists of three layers: input, output, and hidden layer. The given dataset has been read in the input layer and allocated the correct number of receptors based on the independent variables (Fig.2).

In the hidden layer, the calculation operation is carried out to give the output of this layer by multiplying the input value with the corresponding weight (Othman et al. 2020). The output layer has taken the input value by multiplying factors with corresponding weights and gives the calculated value of each variable in the ANN model (Fig.5). Before running the ANN model, training the data is the primary and most significant process for the model for the most accurate outcome (Kim et al. 2020). For training the data, the

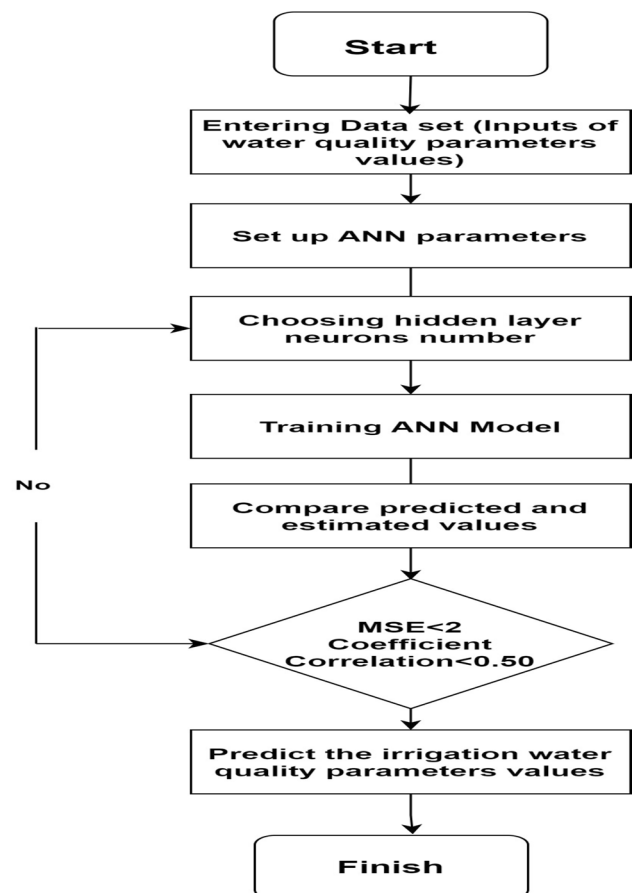


Fig. 2 Prediction methodology flowchart of irrigation water quality

feed-forward backpropagation algorithm was used to prepare the dataset. The present study assigns KR, MH, PI, RSC, SAR, and SSP as input variables. Based on these parameters, the irrigation water quality parameters for the study area were predicated. The quality of groundwater for irrigation purposes is present as one output parameter. The data usually refers to the Akot basin to estimate the prediction model’s feasibility and assess the proposed model’s ability in various climatic conditions. Figure 5 displays the prediction methodology flowchart of the stages followed.

Multi-linear regression

Multi-linear regression analysis is one of the most basic mathematical models. It is based on linear relationships with both inputs and outputs, or, to put it differently, it derives linear correlation between various variables by incorporating a regression steady into the formula. Results of MLR model are based on the equation below in 1:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots b_ix_i \tag{1}$$

where:

Y: the independent variable

B: the regression constant

X: the *i*th predictor

LSTM model

The LSTM model is a sophisticated recurrent neural system specifically developed to avoid the exploding/vanishing gradient difficulties common when learning long-term dependency, even when the relatively little time lags are extremely lengthy. Ouma et al. (2012) had introduced the LSTM method to overcome this problem. In the simulation of sequence-based issues with long-term dependencies, the LSTM is better suited (Chang et al. 2015). LSTM means a long short-term memory model compared with ANN and MLR models. The memory blocks of the LSTM-RNN model include the input, forget and output gates, which are used to reset the hidden units (Fig. 3). The gates are responsible for directing the network’s internal operations. Despite other LSTM types, a comparison study indicates that the conventional LSTM remains the most significant (Gref et al. 2017).

Model’s performance criteria

To compare and estimate the machine learning models, the subsequent statistical performance criteria measures were used in Eqs. 2 to 4:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_A^i - Q_P^i)^2} \tag{2}$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (Z_i - Z^i)^2}{\sum_{i=1}^n (Z_i)^2}} \tag{3}$$

$$MSE = \frac{\sum_{i=1}^n (Z_i - Z^i)^2}{n} \tag{4}$$

where *Z_i* and *Zⁱ* are the measures and estimated value; *n* is the number of values used in the model. Two models can be helpful for regression and sorting, learned by their particular methods and validated during the training process for hidden data. The comprehensive analysis of the artificial neural network modes should be marked as beyond the goals of this study. The performance of two models was estimated as per the old statistic performances such as coefficient of correlation (*r*) and mean square error (MSE).

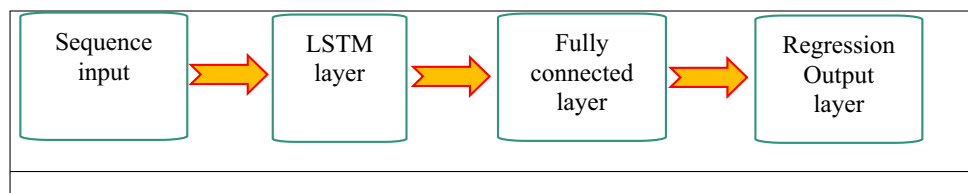
Agriculture water quality parameters

In this study, six parameters as KI, MH, PI, RSC, SAR, and SSP, were selected to predict the value of water quality factors and calculated to obtain the appropriate groundwater quality for agriculture uses in the study area. The prediction of irrigation water quality was estimated using ANN modelling.

Kelly ratio

Kelly ratio is an important index to evaluate the suitability of groundwater for irrigation uses. Kelly (1940) developed the equation to estimate water quality parameters such as calcium, magnesium, and sodium on groundwater quality for irrigation purposes. The classification of groundwater based on KR values 0 to 1 is suitable, and greater than one is unsuitable for

Fig. 3 LSTM model layers



irrigation uses. The following formula was used to estimate the KR of groundwater.

$$KI = \frac{Na}{Ca + Mg} \quad (5)$$

Magnesium hazards

Magnesium hazards is another primary index to assess the groundwater quality for irrigation use. The concentration of magnesium in groundwater plays a vital role in crop yield and growth. In general, calcium and magnesium maintain the state of groundwater equilibrium. The excessive concentration of Mg^{2+} causes soil structure deterioration, increasing the soil alkaline nature and reducing plant growth. The classification of groundwater based on MH is that the value less than 50 is suitable, and greater than 50 is unsuitable for irrigation purposes. The following formula has been used to calculate the MH value of groundwater.

$$MH = \frac{Mg}{Ca + Mg} * 100 \quad (6)$$

Permeability Index

Soil permeability plays a vital role in crop yield and the water circulation process on the field. The permeability of soil gets affected due to excessive concentrations of sodium, calcium, magnesium, and bicarbonate in groundwater for a long time. The classification of groundwater based on PI is class I (greater than 75%), class II (25–75 %), and class III (less than 25%). Doneen (1964a, 1964b) developed the formula for PI to estimate the water movement in the soil layer (Ghazaryan and Chen 2016).

$$PI = \frac{Na + \sqrt{HCO_3}}{Ca + Mg + Na} * 100 \quad (7)$$

Residual sodium carbonate

Residual sodium carbonate (RSC) is a significant index to assess the groundwater quality for irrigation uses. The concentration of bicarbonate and carbonate highly influenced the groundwater chemistry and its quality for irrigation use. The quality of groundwater diminishes when the concentration of carbonate and bicarbonate exceeds the total concentration of calcium and magnesium. Eaton (1950) developed the formula to estimate the RSC value of groundwater (Eq. 8).

$$RSC = (HCO_3^- + CO_3^{2-}) - (Ca^{2+} + Mg^{2+}) \quad (8)$$

Sodium absorption ratio

Sodium adsorption ratio (SAR) is a significant measure to evaluate the suitability of groundwater for irrigation purposes. The measured value of SAR divulged the relative concentration of sodium, calcium, and magnesium in groundwater. The excess sodium concentration in groundwater affects the quality of soil and deteriorates the groundwater equilibrium structure. The ratio of the sodium concentration and the sum of the concentration of calcium and magnesium gives the value of SAR of groundwater (Eq. 9)

$$SAR = \frac{Na}{\sqrt{(Ca + Mg)}/2} \quad (9)$$

Soluble sodium percentage (SSP)

The concentration of calcium, magnesium, and sodium plays a vital role in groundwater quality for irrigation uses. The soluble sodium percent (SSP) classification of groundwater less than 50% is suitable, and greater than 50% is unsuitable for irrigation uses.

$$SSP = \frac{Na * 100}{Ca + Mg + Na} \quad (10)$$

Principle component analysis

The PCA gives more accuracy, extracts correlation relationships, and reduces data into different components that describe a percentage of the total variance between chemical parameters. The varimax rotation methods have been adopted to identify the highest loading factors primarily related to groundwater's chemical composition. The estimated high loaded factor helps identify the process involved in deteriorating groundwater quality in the study region. PCA visualizes the variables in two- or three-dimensional space to determine homogeneous observer groups or, on the opposite, unusual observational groups (Tables 1 and 2). In addition, the number of variables decreases without losing information (Praus 2019; Nguyen et al. 2020). PCA is used to determine the current correlations between the chemical components of the irrigation water quality parameters. To choose the elements of a robust correlation with irrigation water quality parameters and use them in constructing the models (training step) as inputs, the PCA component and scree plots are presented in Fig. 4. PCA technique was an analysis based on SPSS 25.0 software. Therefore, we

Table 1 Principal component analysis of groundwater during pre-monsoon

Component	Initial eigenvalues		
	Total	% of variance	Cumulative %
EC	5.19	30.50	30.50
pH	2.50	14.69	45.19
TDS	2.23	13.13	58.31
Ca	1.89	11.11	69.42
Mg	1.49	8.79	78.21
Cl	1.30	7.67	85.88
CO3	0.97	5.71	91.59
HCO3	0.59	3.49	95.08
SO4	0.48	2.85	97.93
Na	0.29	1.71	99.64
K	0.03	0.17	99.81
RSC	0.02	0.09	99.90
MH	0.01	0.06	99.96
SAR	0.01	0.04	100.00
PI	0	0.001	100.00
SSP	4.16E−06	2.45E−05	100.00
KI	−4.61E−17	−2.71E−16	100
Extraction sums of squared loadings	5.19	30.50	30.50
	2.50	14.69	45.19
Rotation sums of squared loadings	4.02	23.64	23.64
	3.66	21.55	45.19

Table 2 Principal component analysis of groundwater during post-monsoon

Component	Initial eigenvalues		
	Total	% of variance	Cumulative %
EC	4.766	28.035	28.035
pH	2.804	16.492	44.527
TDS	2.207	12.981	57.508
Ca	1.915	11.266	68.773
Mg	1.599	9.408	78.181
Cl	1.276	7.507	85.688
CO3	0.906	5.332	91.02
HCO3	0.683	4.015	95.035
SO4	0.524	3.083	98.118
Na	0.268	1.578	99.696
K	0.023	0.138	99.833
RSC	0.014	0.082	99.915
MH	0.01	0.059	99.974
SAR	0.004	0.025	99.999
PI	0	0.001	100
SSP	1.66E−06	9.76E−06	100
KI	−1.72E−17	−1.01E−16	100
Extraction sums of squared loadings	4.766	28.035	28.035
	2.804	16.492	44.527
Rotation sums of squared loadings	4.363	25.667	25.667
	3.206	18.86	44.527

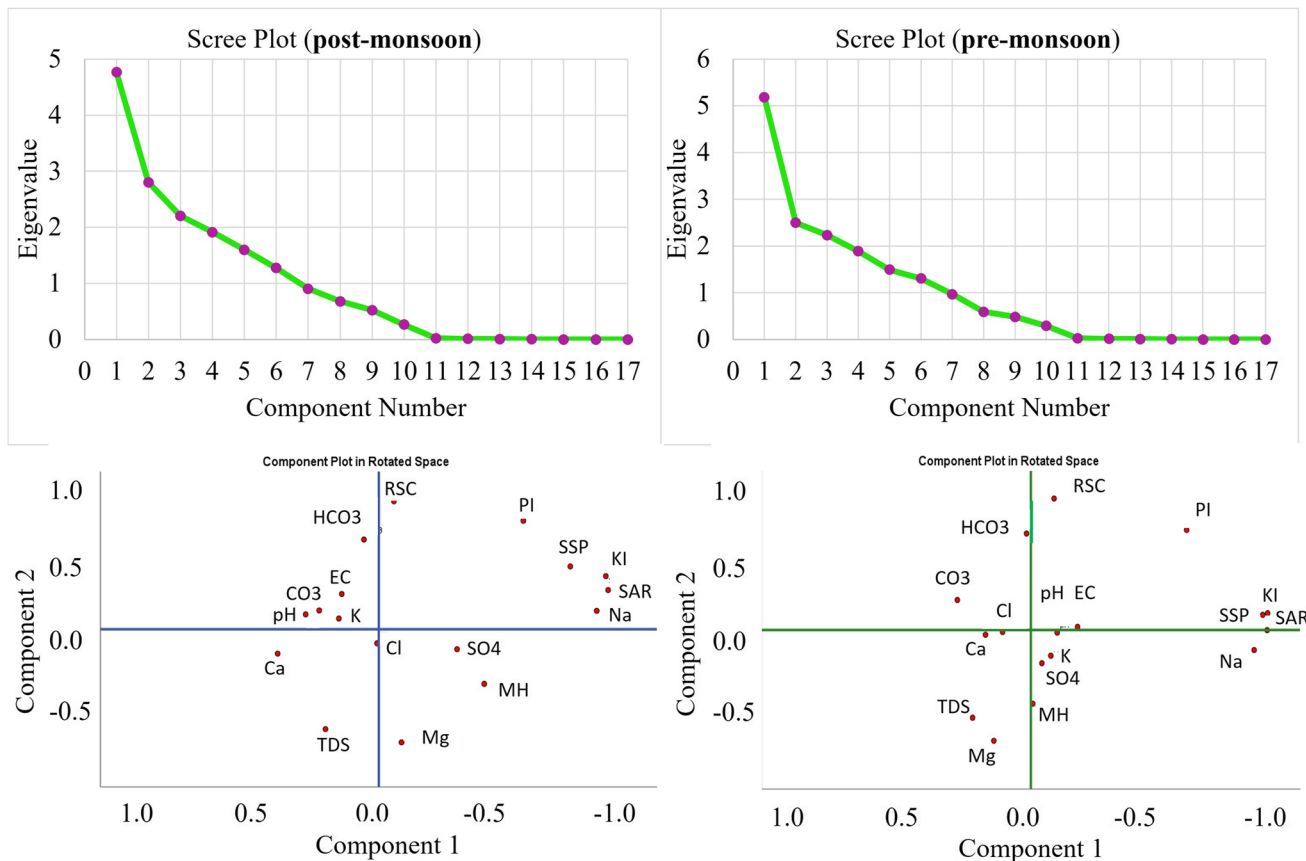


Fig. 4 PCA component and scree plots of pre-post monsoon in semi-arid region

Table 3 Correlation matrix between input and output variables of pre-monsoon

	EC	pH	TDS	Ca	Mg	Cl	CO3	HCO3	SO4	Na	K	RSC	MH	SAR	PI	SSP	KI
EC	1.000	.480	.003	.082	-.067	.295	.142	.046	.065	.116	.432	.047	-.123	.124	.096	.266	.115
pH	.480	1.000	-.091	-.109	-.025	.090	.022	-.107	-.393	.025	.195	-.038	.027	.040	.003	.128	.047
TDS	.003	-.091	1.000	.646	.751	-.008	.141	.036	.099	.097	.124	-.507	.074	-.118	-.518	-.329	-.309
Ca	.082	-.109	.646	1.000	-.019	-.100	.235	.197	-.071	.110	.139	-.172	-.700	-.038	-.240	-.176	-.177
Mg	-.067	-.025	.751	-.019	1.000	.077	-.019	-.123	.191	.033	.042	-.516	.703	-.122	-.471	-.279	-.251
Cl	.295	.090	-.008	-.100	.077	1.000	-.008	.046	-.024	-.149	.171	.043	.144	-.140	-.067	-.056	-.133
CO3	.142	.022	.141	.235	-.019	-.008	1.000	.244	.128	-.181	.096	.197	-.128	-.213	-.061	-.200	-.239
HCO3	.046	-.107	.036	.197	-.123	.046	.244	1.000	-.069	.030	-.069	.840	-.224	.037	.623	.021	.030
SO4	.065	-.393	.099	-.071	.191	-.024	.128	-.069	1.000	.029	.211	-.104	.165	.012	-.064	.075	-.005
Na	.116	.025	.097	.110	.033	-.149	-.181	.030	.029	1.000	.021	-.041	-.062	.973	.495	.831	.900
K	.432	.195	.124	.139	.042	.171	.096	-.069	.211	.021	1.000	-.119	-.048	-.022	-.118	.304	-.072
RSC	.047	-.038	-.507	-.172	-.516	.043	.197	.840	-.104	-.041	-.119	1.000	-.238	.081	.805	.184	.178
MH	-.123	.027	.074	-.700	.703	.144	-.128	-.224	.165	-.062	-.048	-.238	1.000	-.069	-.180	-.076	-.063
SAR	.124	.040	-.118	-.038	-.122	-.140	-.213	.037	.012	.973	-.022	.081	-.069	1.000	.619	.901	.976
PI	.096	.003	-.518	-.240	-.471	-.067	-.061	.623	-.064	.495	-.118	.805	-.180	.619	1.000	.665	.698
SSP	.266	.128	-.329	-.176	-.279	-.056	-.200	.021	.075	.831	.304	.184	-.076	.901	.665	1.000	.918
KI	.115	.047	-.309	-.177	-.251	-.133	-.239	.030	-.005	.900	-.072	.178	-.063	.976	.698	.918	1.000

Table 4 Correlation matrix between input and output variables of post-monsoon

	EC	pH	TDS	Ca	Mg	Cl	CO3	HCO3	SO4	Na	K	RSC	MH	SAR	PI	SSP	KI
EC	1.000	.430	-.059	.029	-.108	.284	.119	.080	.009	.014	.350	.103	-.112	.006	.039	.155	-.012
pH	.430	1.000	-.074	-.160	.050	-.008	.025	-.050	-.423	-.209	.195	.005	.112	-.196	-.108	-.075	-.180
TDS	-.059	-.074	1.000	.680	.723	.042	.037	-.064	-.050	-.019	.018	-.620	-.055	-.250	-.567	-.472	-.426
Ca	.029	-.160	.680	1.000	-.015	-.094	.034	.068	-.030	-.021	.251	-.334	-.758	-.207	-.363	-.269	-.355
Mg	-.108	.050	.723	-.015	1.000	.145	.019	-.151	-.040	-.005	-.212	-.530	.640	-.145	-.430	-.389	-.245
Cl	.284	-.008	.042	-.094	.145	1.000	-.015	-.002	.206	-.101	.183	-.026	.158	-.089	-.073	.017	-.067
CO3	.119	.025	.037	.034	.019	-.015	1.000	.117	-.105	-.098	.144	.133	-.028	-.128	-.057	-.074	-.153
HCO3	.080	-.050	-.064	.068	-.151	-.002	.117	1.000	-.161	.115	-.225	.821	-.158	.133	.646	.037	.137
SO4	.009	-.423	-.050	-.030	-.040	.206	-.105	-.161	1.000	.166	.120	-.103	.014	.158	.030	.210	.146
Na	.014	-.209	-.019	-.021	-.005	-.101	-.098	.115	.166	1.000	-.044	.094	.065	.964	.601	.781	.875
K	.350	.195	.018	.251	-.212	.183	.144	-.225	.120	-.044	1.000	-.176	-.323	-.067	-.177	.343	-.085
RSC	.103	.005	-.620	-.334	-.530	-.026	.133	.821	-.103	.094	-.176	1.000	-.094	.237	.822	.294	.340
MH	-.112	.112	-.055	-.758	.640	.158	-.028	-.158	.014	.065	-.323	-.094	1.000	.120	.025	-.002	.168
SAR	.006	-.196	-.250	-.207	-.145	-.089	-.128	.133	.158	.964	-.067	.237	.120	1.000	.727	.869	.972
PI	.039	-.108	-.567	-.363	-.430	-.073	-.057	.646	.030	.601	-.177	.822	.025	.727	1.000	.690	.790
SSP	.155	-.075	-.472	-.269	-.389	.017	-.074	.037	.210	.781	.343	.294	-.002	.869	.690	1.000	.891
KI	-.012	-.180	-.426	-.355	-.245	-.067	-.153	.137	.146	.875	-.085	.340	.168	.972	.790	.891	1.000

have expected to describe the Pearson coefficient correlation matrix (Tables 3 and 4).

Correlation analysis

Correlation analysis has been used in statistical techniques for measuring the strength of a linear relationship between two variables. Due to independence or dependence, the variables are not selected. In most research, an analysis of correlation was used to examine the linear relationship between two variables. The correlation matrix has been developed by calculating the coefficient of various sets of parameters to compute correlation coefficients. Evaluating *p* values tested the significance of the correlation. The variation is significant if *p* is less than 0.05, 0.01 (*p* < 0.05 and *p* < 0.01) (Tables 3 and 4). The change is not significant when *p* > 0.05 (Eq. 6). The significance level is measured between 0.01 and 0.05 (Malik and Hashmi 2017; Sar et al. 2017; Tiwary et al. 2018). Pearson correlation analysis between all variables (input/output) to investigate their relationships is conducted, and those results are enclosed in Tables 3 and 4.

$$r = \frac{\sum_{i=1}^n xi yi - \frac{\sum_{i=1}^n xi \sum_{i=1}^n yi}{n}}{\sqrt{\left[\sum_{i=1}^n xi^2 - \frac{(\sum_{i=1}^n xi)^2}{n}\right]} \sqrt{\left[\sum_{i=1}^n yi^2 - \frac{(\sum_{i=1}^n yi)^2}{n}\right]}} \quad (11)$$

Results

Evaluation of results

In the current period, machine learning models have been broadly used in various fields. They can be helpful to the prediction of future scenarios of conservational and natural processes. In this paper, we have studied LSTM, MLR, and artificial neural network models with two scenario results that are generally practised for measuring groundwater suitability in irrigation drives using the simply quantifiable input variables such as Mg²⁺, Ca²⁺, HCO₃⁻, CO₃, Na⁺, K⁺ for two scenario machine learning models. LSTM, ANN, and MLR models were carried out based on the number of neurons in the hidden layers. This study’s findings have shown machine learning models are very effective techniques for predicting water quality values (Elbeltagi et al. 2021). This paper describes the training and testing of LSTM, MLR, ANN models, as well as their validation and simplification outcomes for predicate values. The comparison between ANN, LSTM, and MLR models’ performance is enclosed in Tables 5 and 6. The ANN and MLR are more correct for predicting water quality values in the first and second scenario results. A detailed description is provided in the below sub-sections.

Comparison of training and testing datasets for scenario 1

We have included all irrigation water quality variable datasets that have been used for training and testing model developed in the scenario 1. The training and

Table 5 Comparison between different models' performance for scenario 1

1st scenario performance criteria results										
		ANN			LSTM			MLR		
		RMSE	R^2	MSE	RMSE	R^2	MSE	RMSE	R^2	MSE
Train	RSC	0.00064	1	4.09E-07	1.438569	0.923467	2.069482	2.32E-15	1	5.39E-30
	MH	0.002735	0.998001	7.48E-06	0.00164	0.999748	2.69E-06	0.016405	0.973859	0.000269
	SAR	0.023108	0.997801	0.000534	0.002883	0.999957	8.31E-06	0.049997	0.986722	0.0025
	PI	0.006181	0.996004	3.82E-05	0.001555	0.999753	2.42E-06	0.01737	0.970681	0.000302
	SSP	0.31305	0.998001	0.098	2.246052	0.93129	5.04475	1.158658	0.974383	1.342487
	KI	0.006116	0.996004	3.74E-05	0.0037	0.999931	1.37E-05	0.0261	0.94756	0.000681
Test	RSC	0.059414	0.9998	0.00353	1.457929	0.905571	2.125558	2.64E-15	1	6.98E-30
	MH	0.008781	0.994009	7.71E-05	0.043908	0.812187	0.001928	0.01934	0.956621	0.000374
	SAR	0.029631	0.993212	0.000878	0.251246	0.683688	0.063124	0.039245	0.988923	0.00154
	PI	0.011446	0.990025	0.000131	0.018754	0.967122	0.000352	0.020062	0.962027	0.000402
	SSP	0.363731	0.995006	0.1323	13.82245	0.066016	191.0601	1.170854	0.967353	1.370898
	KI	0.00634	0.996004	4.02E-05	0.049018	0.795875	0.002403	0.020295	0.959979	0.000412

Table 6 Comparison between different models' performance for scenario 2

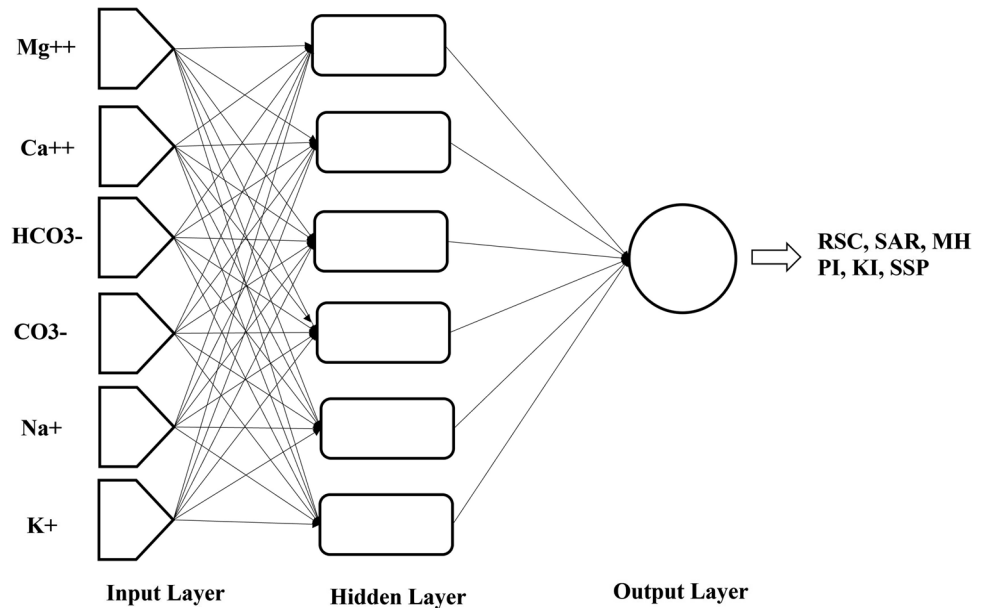
2nd scenario performance criteria results										
		ANN			LSTM			MLR		
		RMSE	R^2	MSE	RMSE	R^2	MSE	RMSE	R^2	MSE
Train	RSC	1.068082	0.893025	1.1408	0.042632	0.999834	0.001818	1.196797	0.867167	1.432323
	MH	0.057446	0.654481	0.0033	0.000356	0.999981	1.26E-07	0.069244	0.514359	0.004795
	SAR	0.059161	0.982081	0.0035	0.002488	0.999968	6.19E-06	0.092815	0.954267	0.008615
	PI	0.03873	0.844561	0.0015	0.001843	0.999896	3.4E-06	0.04616	0.784515	0.002131
	SSP	3.096224	0.826281	9.5866	2.198714	0.911968	4.834343	3.410414	0.778099	11.63092
	KI	0.037417	0.889249	0.0014	0.00165	0.999801	2.72E-06	0.046528	0.836761	0.002165
Test	RSC	1.123032	0.872356	1.2612	1.876118	0.703544	3.519817	1.177207	0.866449	1.385817
	MH	0.086023	0.235225	0.0074	0.144303	0.228355	0.020823	0.06877	0.465225	0.004729
	SAR	0.107238	0.906304	0.0115	0.136041	0.882314	0.018507	0.075667	0.956796	0.005726
	PI	0.05099	0.783225	0.0026	0.070351	0.621794	0.004949	0.045977	0.791098	0.002114
	SSP	3.487607	0.7396	12.1634	8.174895	0.141853	66.82891	3.399641	0.755534	11.55756
	KI	0.061644	0.777924	0.0038	0.066868	0.637924	0.004471	0.03783	0.848287	0.001431

testing results obtained by LSTM, MLR, and ANN are presented in Table 5. As depicted in Table 5, in training and testing, ANN models have shown the maximum RSC value of $R^2=1$ and $R^2=0.99$, respectively. Other variables of ANN models such as RSC, MH, SAR, PI, SSP, and KI value are above $R^2=0.99$, and the other values are (RMSE=0.00064, 0.002735, 0.023108, 0.006181, 0.31305, 0.006116 and RMSE=0.059414, 0.008781, 0.02963, 0.011446, 0.363731, 0.00634) in training and testing, respectively. ANN model has been given better performance as compared to other models for scenario 1. Similarly, MLR training and testing models developed based on the parameters such as RSC MH, SAR, PI, SSP and PI have been shown

coefficient correlation, RMSE, and MSE value of ($R^2=1$, 0.973859, 0.986722, 0.970681 and $R^2=0.974383$ 0.94756; 0.00064), (RMSE=2.32E-15, 0.016405, 0.049997, 0.01737, 1.158658, 0.0261 and RMSE=2.64E-15, 0.01934, 0.039245, 0.020062, 1.170854, 0.020295) and (MSE=5.39E-30, 0.000269, 0.0025, 0.000302, 1.342487, 0.000681 and MSE=6.98E-30, 0.000374, 0.00154, 0.000402, 1.370898, 0.000412), respectively.

Furthermore, the LSTM training and testing models of RSC and SSP have shown the lowest coefficient correlation and RMSE values of 0.92 and 0.066. Similarly, training and testing results of LSTM model of RSC MH, SAR, PI, SSP, and PI have shown RMSE and MSE values of

Fig. 5 Artificial neural network architecture



(RMSE=1.438569, 0.00164, 0.002883, 0.001555, 2.246052, 0.0037 and RMSE =1.457929, 0.043908, 0.251246, 0.018754, 13.82245, 0.049018) and (MSE=2.069482, 2.69E-06, 8.31E-06, 2.42E-06, 5.04475, 1.37E-05 and MSE=2.125558, 0.001928, 0.063124, 0.000352, 191.060, 0.002403), respectively (Table 5). As a result, it can be assumed that ANN has predicted the accurate values of irrigation of water quality elements are the most effectively (Figs. 5 to 7).

Comparison of training and testing datasets for scenario 2

The training and testing results obtained by LSTM, MLR, and ANN are presented in Table 6. In training and testing, ANN and LSTM models have shown the maximum RSC value of $R^2=0.89$, 0.872 and $R^2=0.99$, 0.88, respectively. Furthermore, other irrigation water quality parameters such as MH, SAR, PI, SSP, and KI value of ($R^2=0.654481$, 0.982081, 0.844561, 0.826281, 0.889249 and $R^2=0.235225$, 0.906304, 0.783225, 0.7396, 0.777924), (RMSE=1.068082, 0.057446, 0.059161, 0.03873, 3.096224, 0.037417 and RMSE=1.196797, 0.069244, 0.092815, 0.04616, 3.410414, 0.046528) and (MSE=1.1408, 0.0033, 0.0035, 0.0015, 9.5866, 0.0014 and MSE=1.2612, 0.0074, 0.0115, 0.0026, 12.1634, 0.0038), and ($R^2=0.999981$, 0.999968, 0.999896, 0.911968, 0.999801 and $R^2=0.228355$, 0.882314,

0.621794, 0.141853, 0.637924), (RMSE=0.042632, 0.000356, 0.002488, 0.001843, 2.198714, 0.00165 and RMSE= 1.876118, 0.144303, 0.136041, 0.070351, 8.174895, 0.066868) and (MSE=0.001818, 1.26E-07, 6.19E-06, 3.4E-06, 4.834343, 2.72E-06 and MSE=3.519817, 0.020823, 0.018507, 0.004949, 66.82891, 0.004471) in training and testing models of ANN and LSTM, respectively (Figs. 6 to 8). In training and testing, MLR models have shown the maximum SAR value of $R^2=0.95$, respectively Furthermore, MLR models of training and testing have shown well performance as compare to ANN and LSTM developed models in scenario 2. Similarly, MLR training and testing models developed based on water quality parameters such as RSC MH, SAR, PI, SSP, and KI have shown coefficient correlation, RMSE, and MSE values of ($R^2=0.867167$, 0.514359, 0.954267, 0.784515, 0.778099, 0.836761 and $R^2=0.866449$, 0.465225, 0.956796, 0.791098, 0.755534, 0.848287), (RMSE=1.196797, 0.069244, 0.092815, 0.04616, 3.410414, 0.046528 and RMSE=1.177207, 0.06877, 0.075667, 0.045977, 3.399641, 0.03783), and (MSE=1.432323, 0.004795, 0.008615, 0.002131, 11.63092, 0.002165, 1.385817 and MSE=1.385817, 0.004729, 0.005726, 0.002114, 11.55756, 0.001431), respectively. As a result, it can be concluded that of all the machine learning models designed for training and testing, MLR has good prediction values of irrigation of water quality parameters as compare other models.

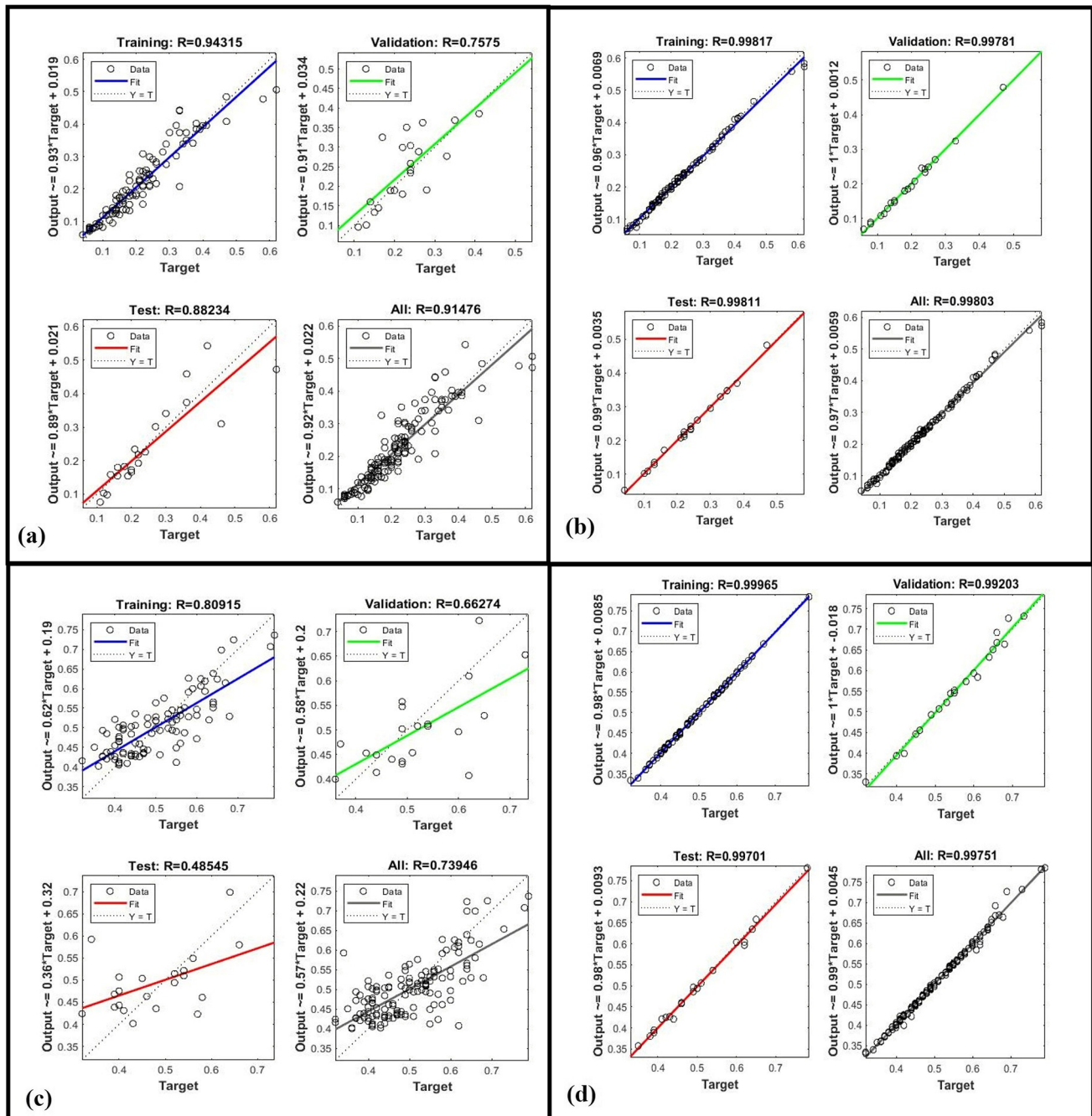


Fig. 6 Graphical plots of the observed values and ANN model predicted values using the training, validation, test, and all dataset

Discussion

Artificial neural networks (ANN) use the neural network's modelling structure, a robust method of modelling complex non-linear relations, mainly when the relationship between variables are unclear (Smith 1994). Each layer consists of one or more essential elements termed as a neuron or node. Each neuron represents an algebraic function which is assigned a parameter with limit values (Dryfus et al. 2002). In this investigation, ANN models were used to predict irrigation

parameter values by using RSC, MH, SAR, PI, Na%, and KI in between irrigation water quality parameters (Figs. 12 and 13). The most acceptable number of neurons and training iterations of the hidden layer are essential indicators in ANN modelling. No specific algorithm is available to determine the adequate number of neurons in the hidden layer, and these values were obtained on a trial-and-error basis (Alizadeh and Kavianpour 2015). The complexity of the problem determines the number of hidden layers, and in most cases, a single hidden layer suffices to model a problem (Rezvan et al. 2016).

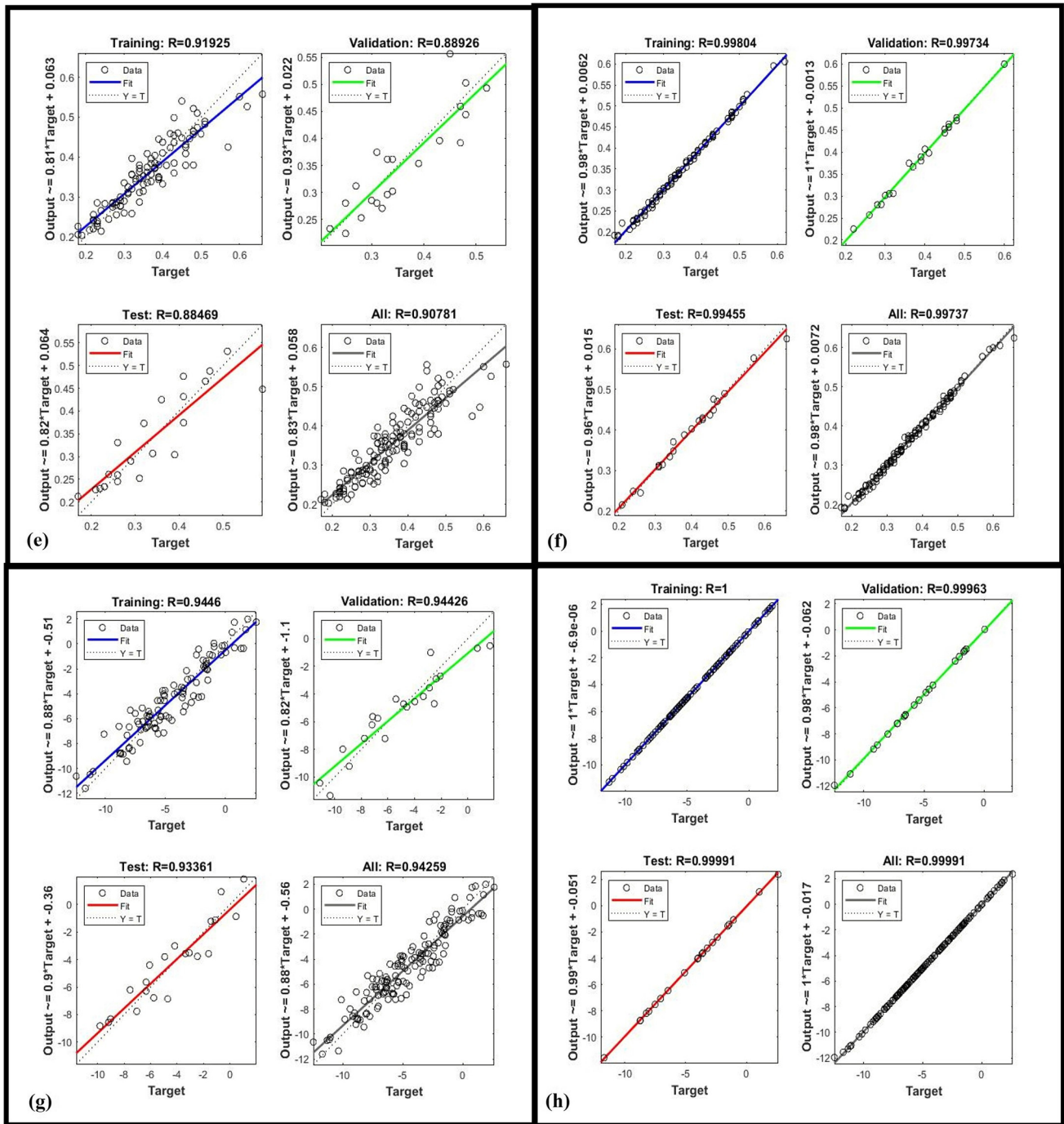


Fig. 7 Graphical plots of the observed values and ANN model predicted values using the training, validation, test, and all dataset

However, the outcomes of a high correct prediction of artificial intelligence methods for irrigation water quality are approved (Castrillo and García 2020; Ahmed et al. 2019; Liu et al. 2019; Lu and Ma 2020). The exact prediction is strongly dependent on the number and impact of the input variables, but all data must be available and

cost-effective. A few studies currently use the parameter that can be used as input variables in situ and in real time (Castrillo and García (2020)). In contrast, numerous variables have significant effects on groundwater quality, such as the hydrologic regime, land use, geomorphologic, and geologic conditions, as well as on anthropogenic activities,

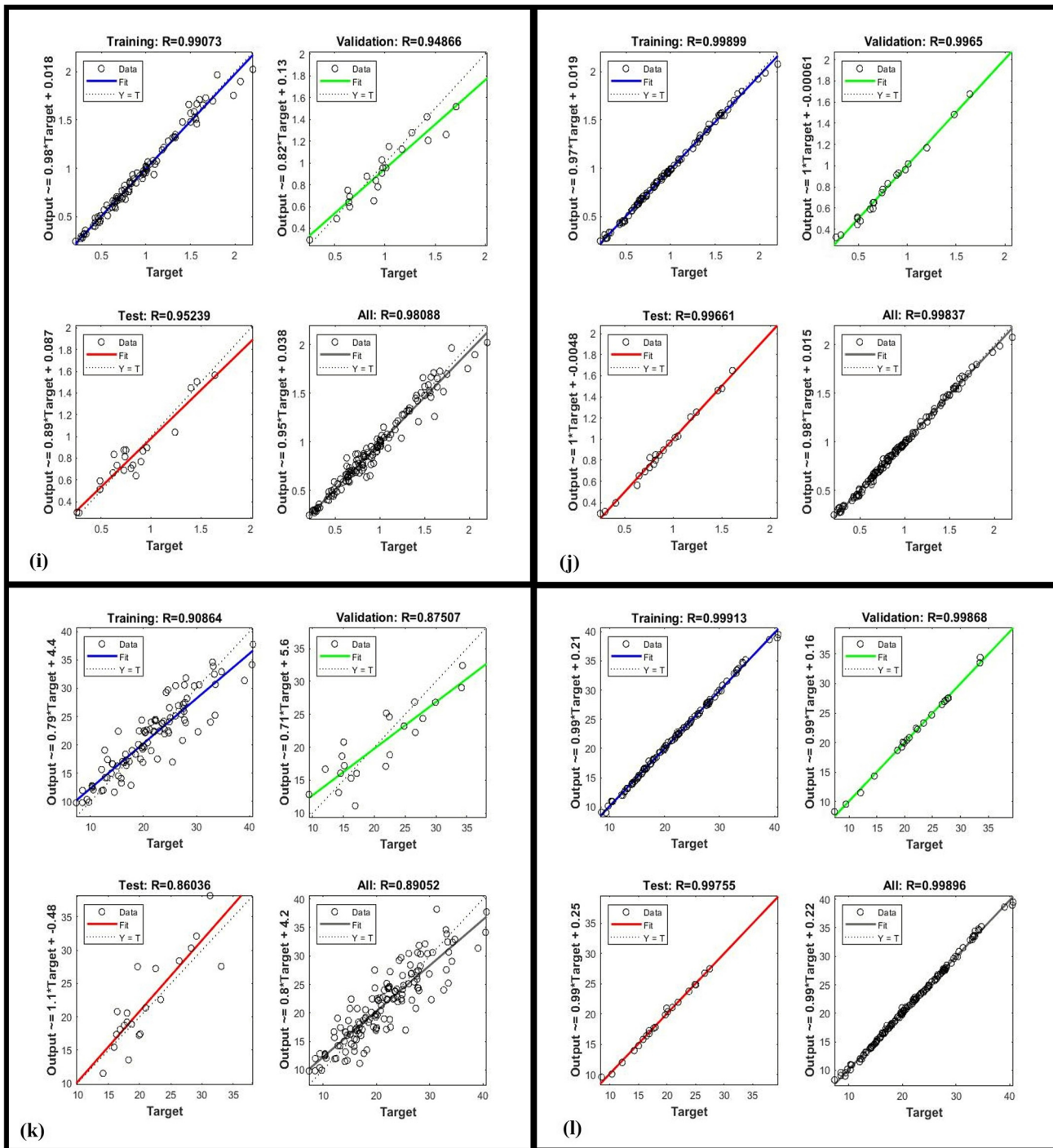


Fig. 8 Graphical plots of the observed values and ANN model predicted values using the training, validation, test, and all dataset

that need to be addressed in widespread use in areas that are different from those used in development (Pande and Moharir 2018). These situations play an essential role in the prediction mixtures of the input parameters used.

A hybrid deep learning model, long short-term memory (LSTM), was used to predict the irrigation water quality, namely total nitrogen, phosphorous, and organic carbon (Liu et al. 2019).

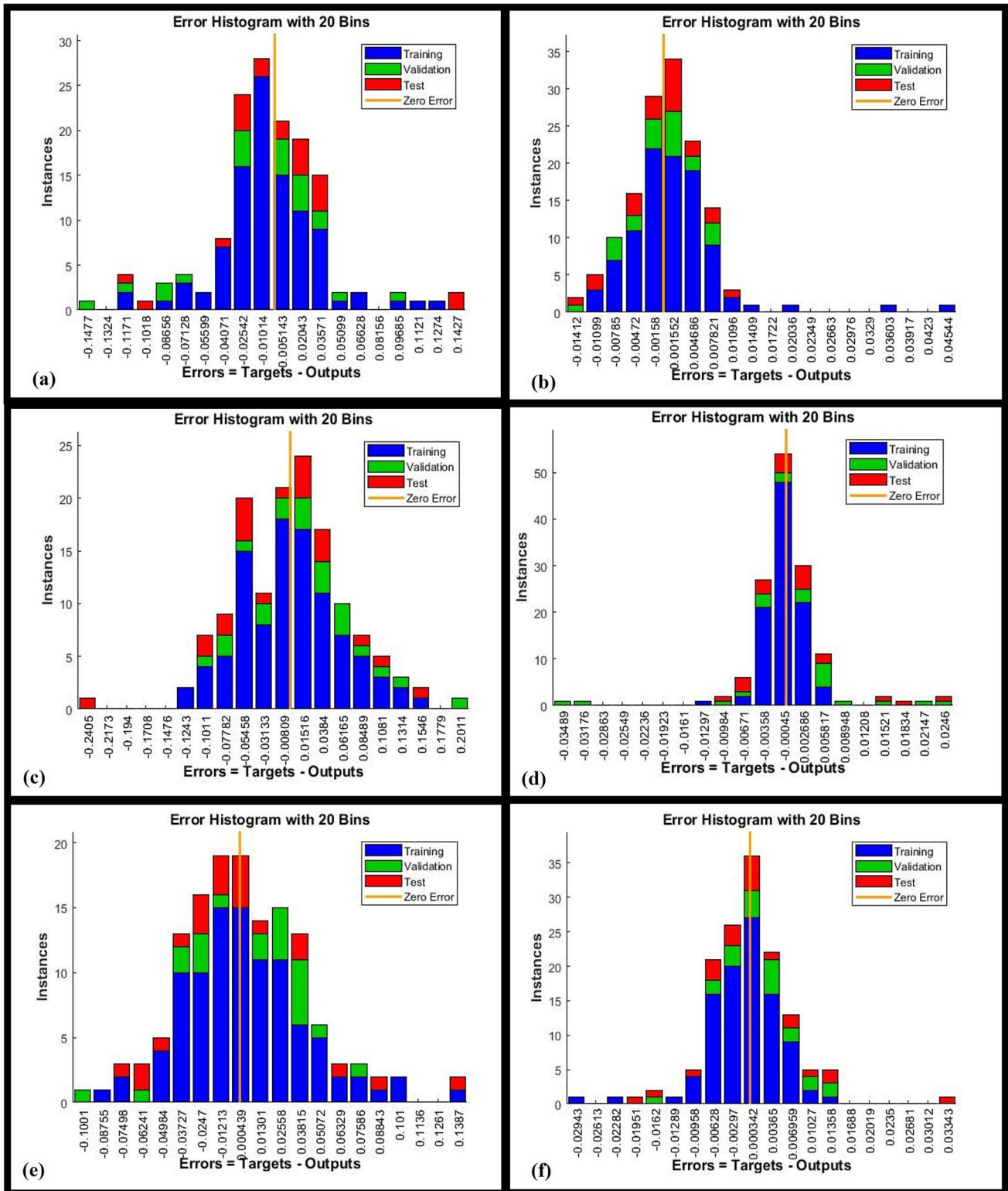


Fig. 9 Error histogram of the both ANN models

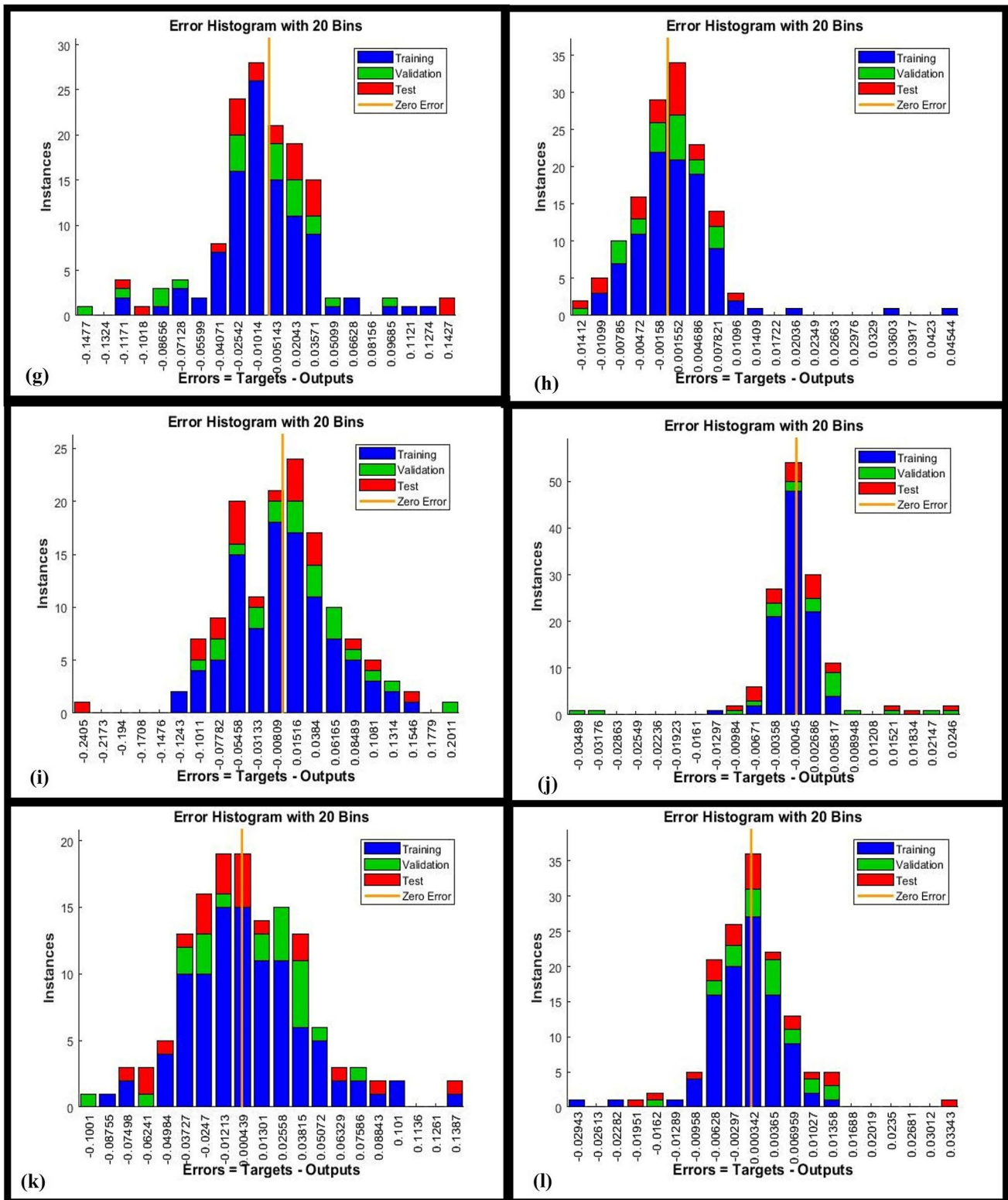


Fig. 9 (continued)

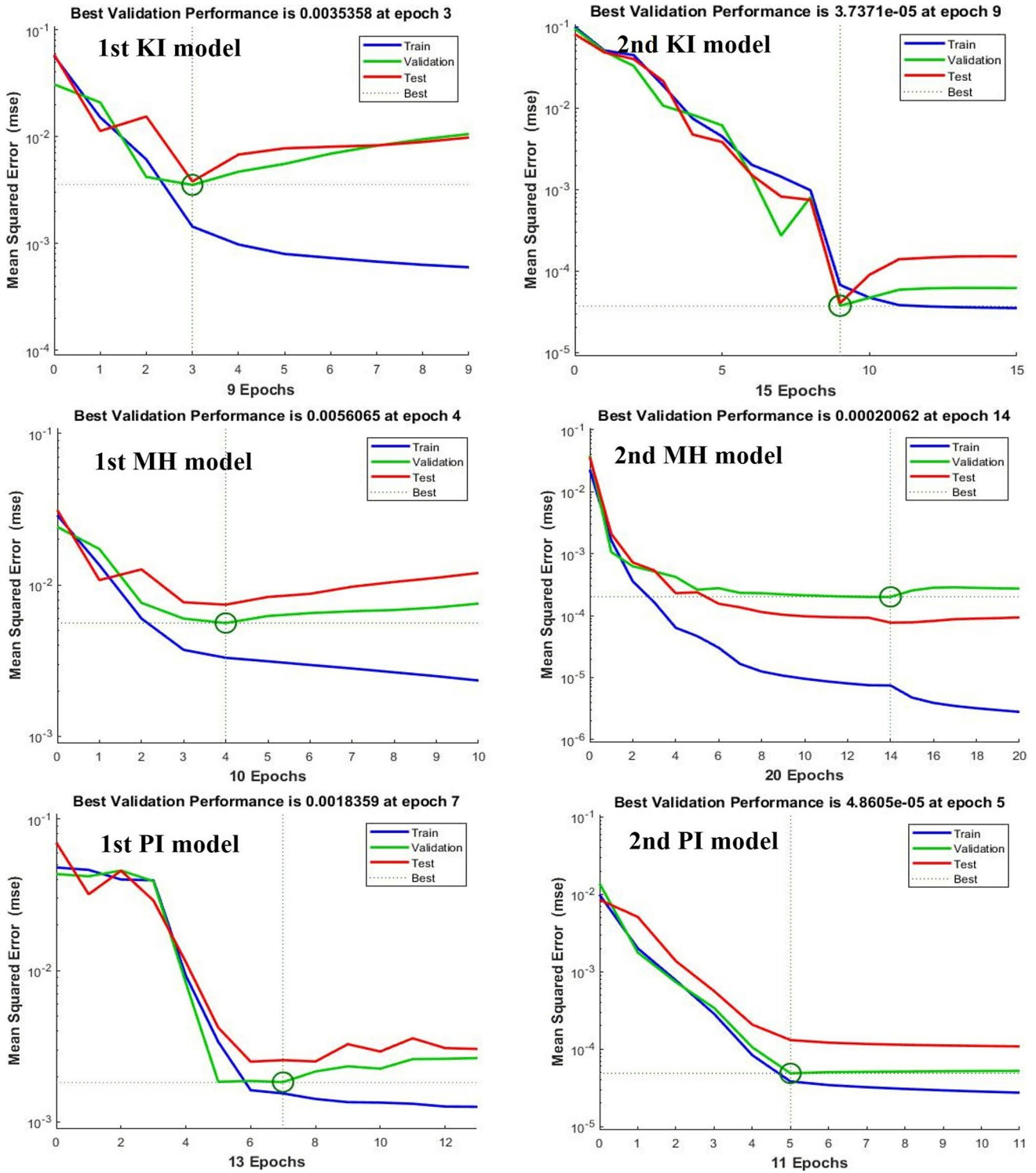


Fig. 10 MSE vs epoch variation of the deviation for ANN models

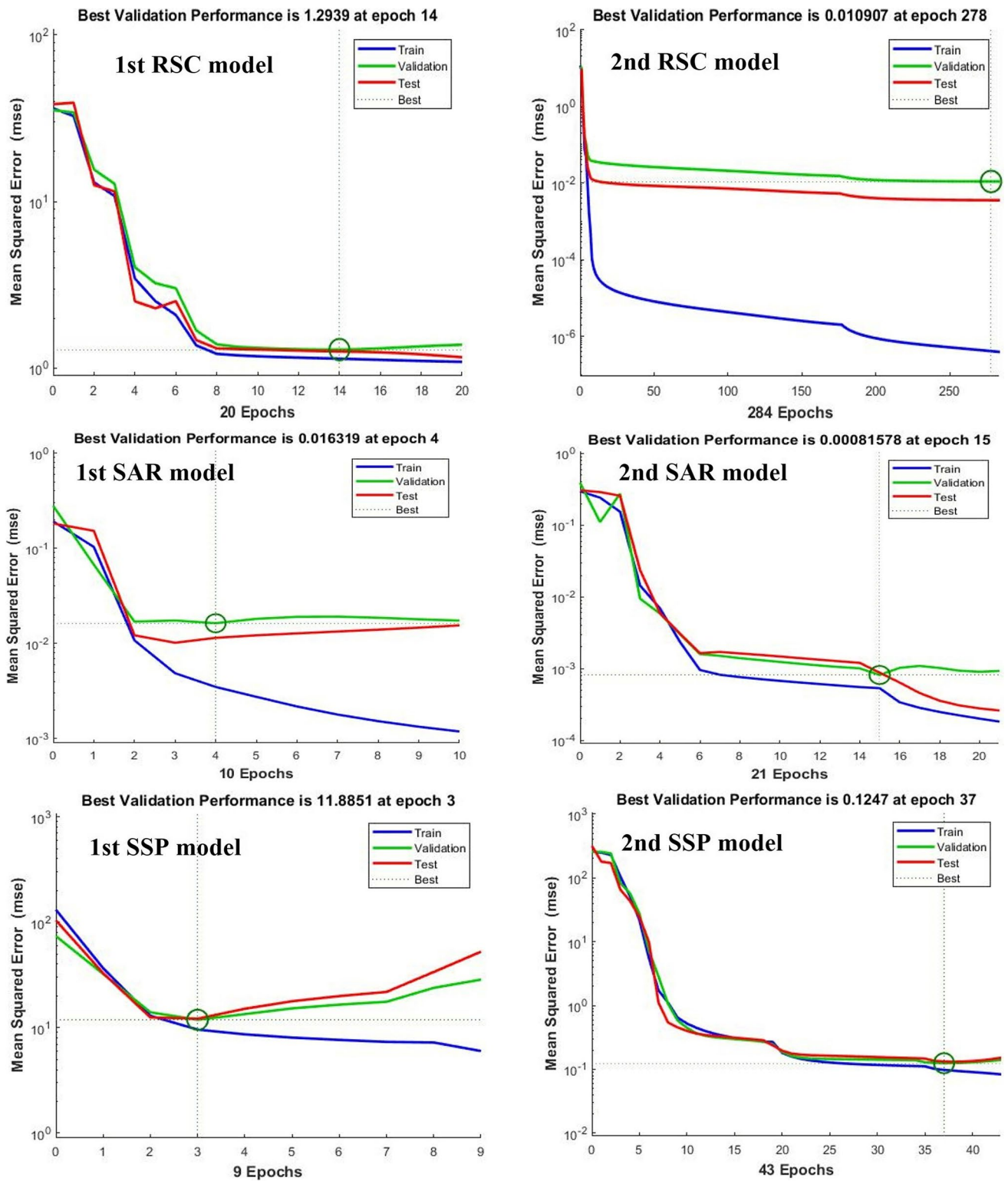


Fig. 11 MSE vs epoch variation of the deviation for ANN models

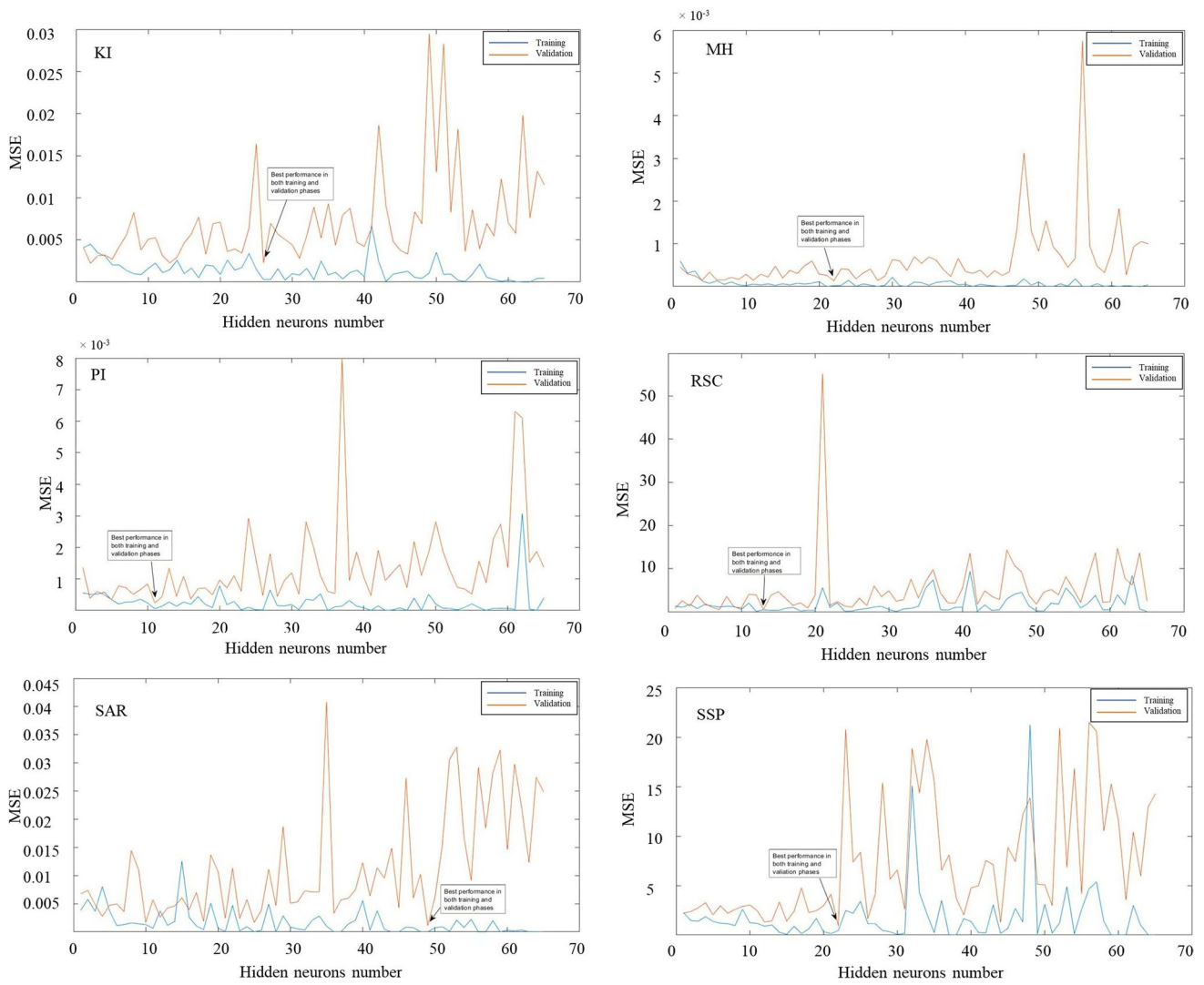


Fig. 12 Show of hidden neurons number vs MSE in ANN model 1

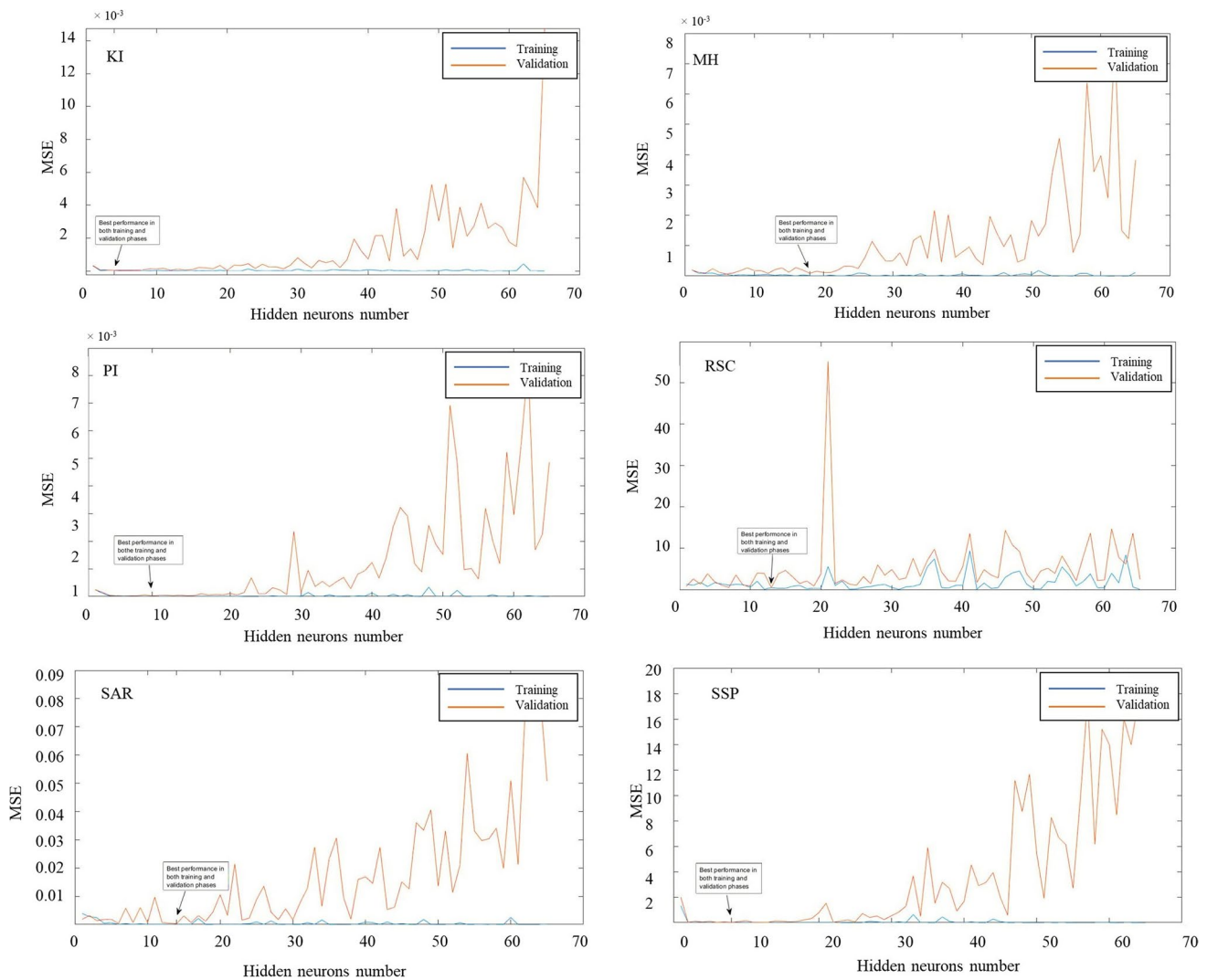


Fig. 13 Show of hidden neurons number vs MSE in ANN model 2

The size of the training dataset has a significant impact on LSTM network training. It is widely assumed that network training necessitates a large amount of training sample data. However, the dataset size is determined by the catchment characteristics and flows of concern, which determine the complexity of the input-output relationships represented by the LSTM (Kratzert et al. 2019). The MLR models of ANN 4-6-6-1 and 1.028, and 1.106 were 0.836 and 0.882 during the training and testing period. Their findings showed that ANN's evaporation estimation was superior to MLR's, matching the current investigation findings (Alizamir et al. 2020).

The irrigation water quality is most important from an agricultural perspective and sustainable crop production. Currently, climate change factors directly impact surface and groundwater water quality (Ostad-Ali-Askari et al. 2018; Ostad-Ali-Askari et al. 2019; Pande et al. 2019a; Derakhshannia et al. 2020; Fattahi Nafchi et al. 2021a, 2021b).

In this context irrigation, groundwater quality prediction can be helpful to maintain the excellent quality of groundwater under various climate change factors. Groundwater is one of the significant sources during the absence of rainwater, while anthropogenic activities affect the groundwater quality parameters, particularly irrigation water quality parameters (Moharir et al. 2019; Javadinejad et al. 2019; Talebmorad et al. 2021). They are mainly determined by analyzing an important number of water quality parameters to quantify the dissolved substance. However, in developing countries, the measurement of all groundwater parameters has often been unsatisfactory and costly. Therefore, opinion reduction and asset maintenance for water quality assessment are significant challenges (Pande et al. 2019b; Salehi-Hafshejani et al. 2019). This research outcome can be beneficial to the established dams for farming purposes, where evaporation degrades the chemical quality of the water considerably, especially in summer times. This

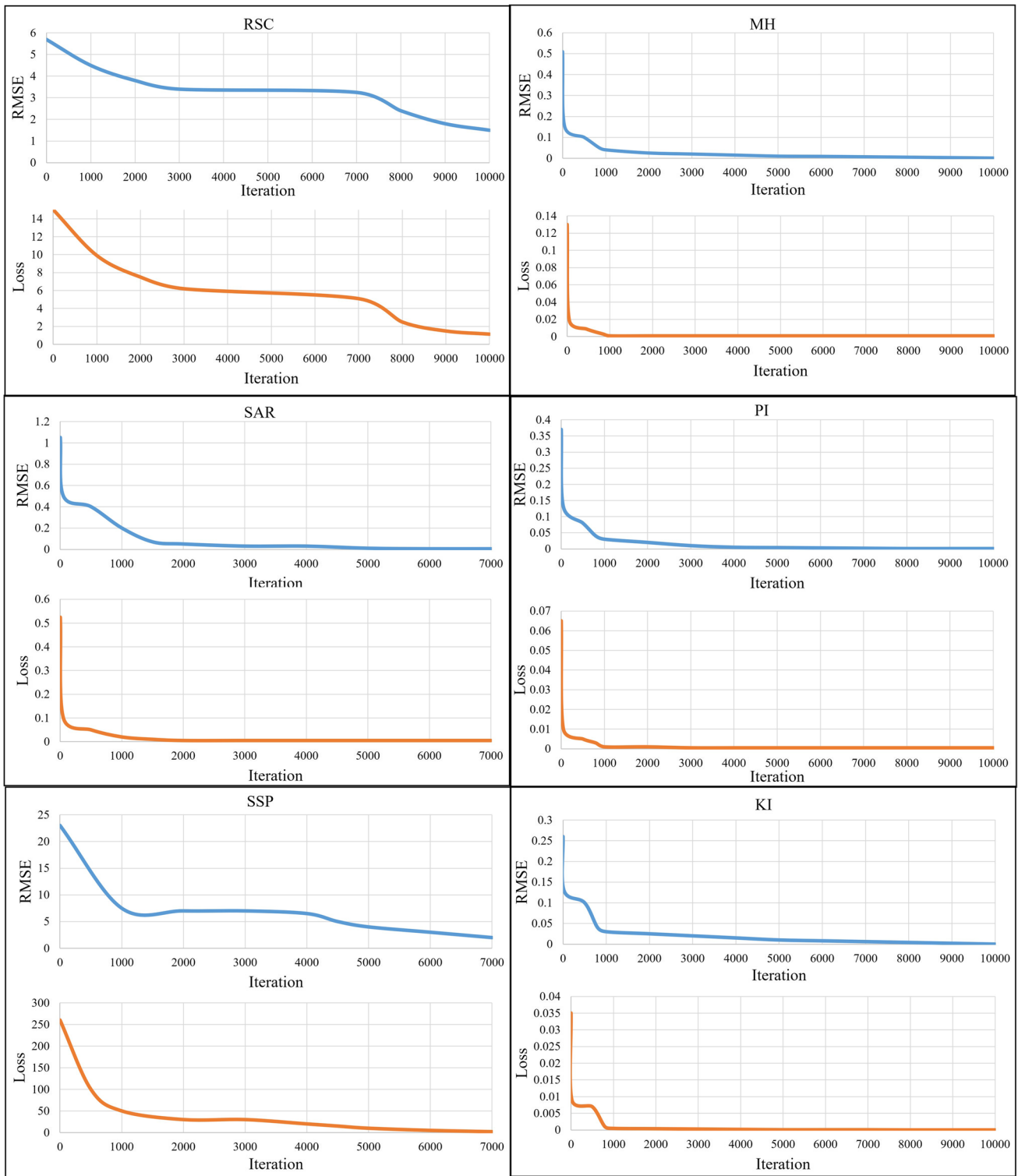


Fig. 14 LSTM model loss plots during scenrio 1

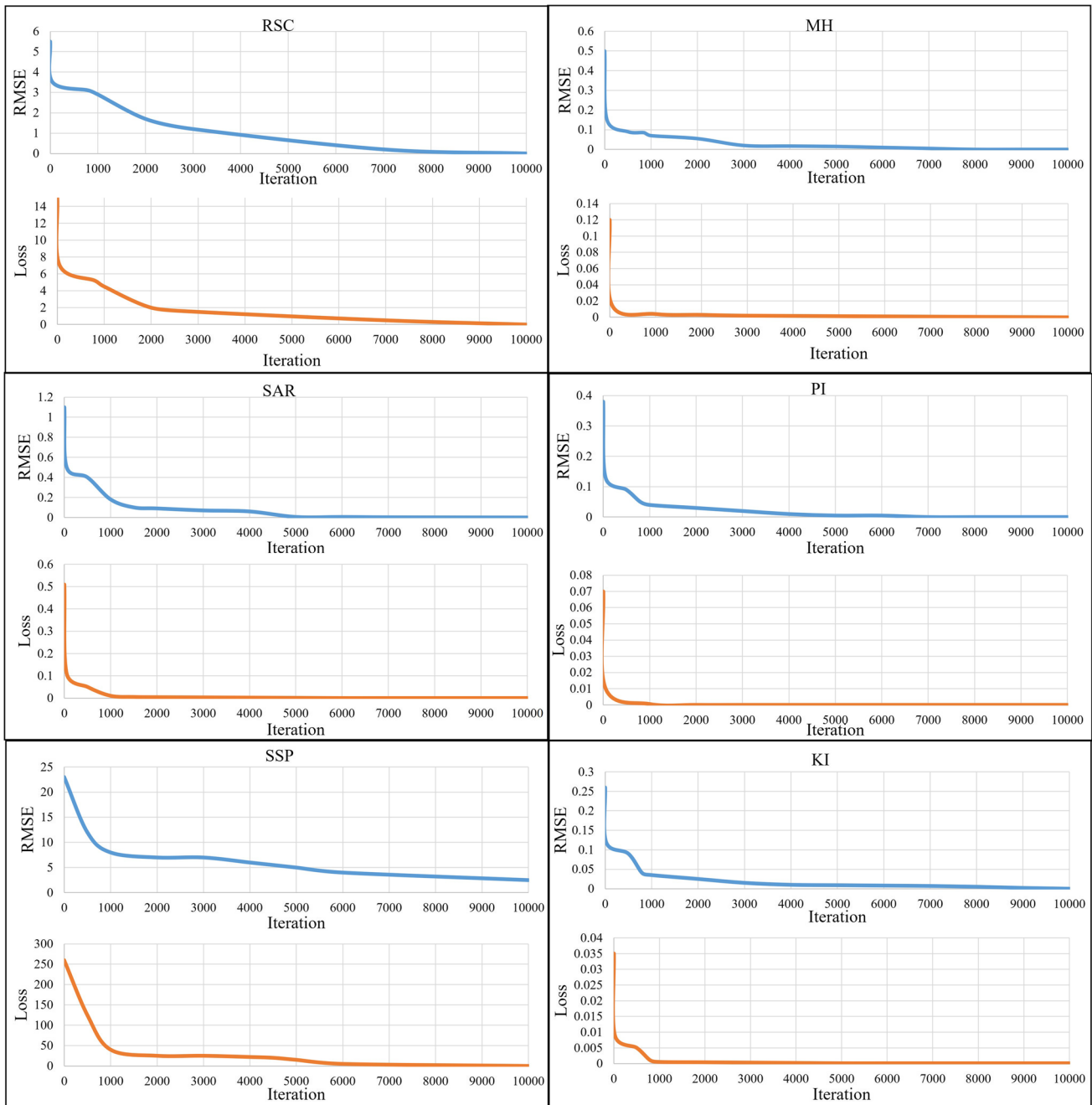


Fig. 15 LSTM model loss plots during scenrio 1

work will thus assist farmers in managing water quality at an efficient cost. Over a short time, since the water assessment depends on the type of soil cultivation according to the water quality class for irrigation purposes, the machine learning model classification, including decreases saline water, is proposed for future studies. The error histogram of ANN models of scenarios 1 and 2 is presented in Fig. 9. The outcomes of MSE vs epoch variation of the deviation for ANN training and testing models are shown in Figs. 10 and 11. Figures 12 and 13 have demonstrated hidden neurons number

vs MSE in ANN models 1 and 2. Figures 14 and 15 show LSTM model loss plots in scenarios 1 and 2 (Table 7).

Table 7 Sizes of training, test and validation sets

	ANN	LSTM	MLR
Training	98	98	98
Test	21	/	/
validation	21	42	42

Conclusion

Performance shows the machine learning models were given to forecasting values of the irrigation water quality parameters in the Akot basin (India). In the current research, water quality parameters were calculated by using MLR, LSTM, and ANN. The input water quality variables for forecasting water quality factors values were Mg^{2+} , HCO_3^- , Ca^{2+} , CO_3 , K^+ , and Na^+ . The water quality parameters were collected from observation wells in the Akot basin, processed, and analyzed in water laboratory, which is a significant concern and limitation of this study. The suggested models were trained and tested in two separate scenarios, i.e. scenario 1 and scenario 2, using different percentages of wells data. The ML models were assessed using statistical tools, including R², RMSE, and MSE, through visual assessment used scatter plots, and line and bar diagrams. Investigation results showed the ANN and MLR models' capability to predicate water quality parameters, integrating all six water quality parameters like RSC, MH, SAR, PI, SSP, and KI. The ANN and MLR models have been highest accuracy in scenario 1 and scenario 2, respectively. It is also clear that the testing data will show a highly precise modelled result for the water quality values forecasting with a large and small sample set for training. The analysis of PCA displays those 17 principal components described based on the results of the data. PCA is used to obtain the current correlations between the chemical components of the irrigation water quality parameters. However, it would also be helpful to conduct studies on benchmarking of different prediction models. It is also suggested that ANN and MLR models be applied under climatic conditions and water quality parameters. In future work, we will use these techniques in different areas. We shall also discover possible enhancements to the method, such as the assertion of lost values and the study of diverse global landscapes. Additionally, we shall encompass these approaches to the cooperative prediction of multiple parameters.

Data availability and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contribution Saber Kouadri: development of ML models, validation, and formal analysis. Chaitanya B. Pande: whole original draft writing and finalized the paper, methodology, supervision, formal analysis, data collection and analysis for modelling purpose, and investigation. Balamurugan Panneerselvam: writing, and review and editing. Kanak N. Moharir: writing, and review and editing. Ahmed Elbeltagi: writing, and review and editing.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publish Not applicable.

Competing interests The authors declare no competing interests.

References

- Ahmed AN, Binti Othman F, Abdulmohsin Afan H, Khaleel Ibrahim R, Ming Fai C, Shabbir Hossain MD, Ehteram M, Elshafie A (2019) Machine learning methods for better water quality prediction. *J Hydrol* 578:124084. <https://doi.org/10.1016/j.jhydrol.2019.124084>
- Aliyu T, Balogun O, Namani C, Olatinwo L, Aliyu A (2017) Assessment of the presence of metals and quality of water used for irrigation in Kwara State, Nigeria. *Pollution* 3(3):461–470
- Alizadeh MJ, Kavianpour MR (2015) Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Mar Pollut Bull* 98(1–2):171–178
- Alizamir M, Kisi O, Muhammad Adnan R, Kuriqi A (2020) Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophys* 68:1113–1126
- Ayers RS, Westcot DW (1999) Water quality for agriculture, FAO Irrigation and Drainage Paper No, 29, Rev. 1, U. N. Food and Agriculture Organization, Rome
- Castrillo M, García AL (2020) Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. *Water Res* 172:115490. <https://doi.org/10.1016/j.watres.2020.115490>
- Chang FJ, Lo YC, Chen PA, Chang LC (2015) Shieh MC multistep-ahead reservoir inflow forecasting by artificial intelligence techniques. Springer International Publishing, Cham, pp 235–249
- Chang NB, Bai K, Chen CF (2017) Integrating multi-sensor satellite data merging and image reconstruction in support of machine learning for better water quality management. *J Environ Manag* 201:227–240
- Charulatha G, Srinivasalu S, Maheswari OU, Venugopal T, Giridharan L (2017) Evaluation of ground water quality contaminants using linear regression and artificial neural network models. *Arab J Geosci* 10(6):128
- Cieszynska M et al (2012) Application of physicochemical data for water-quality assessment of watercourses in the Gdansk Municipality (South Baltic coast). *Environ Monit Assess* 184:2017–2029
- Derakhshannia et al (2020) Corrosion and deposition in Karoon River, Iran, based on hydrometric stations. *International Journal of Hydrology Science and Technology Inderscience Publishers* 10(4):334–345. <https://doi.org/10.1504/IJHST.2020.108264>
- Doneen, L.D., (1964a). Water Quality for Agriculture. Dep. Irrig. Univ. California, Davis, p. 48
- Doneen LD (1964b) Water Quality for Agriculture. Dep. Irrig. Univ. California, Davis, p 48
- Dryfus G, Martinez JM, Samuelides M, Gordon MB, Badran F, Thiria S et al (2002) *Reseaux de Neurones: Methodologie et Applications*. Editions Eyrolles, Paris
- Eaton FM (1950) Significance of Carbonates in Irrigation Waters. *Soil Sci* 69:123–134
- El Baba M, Kayastha P, Huysmans M, De Smedt F (2020) Evaluation of the groundwater quality using the water quality index

- and geostatistical analysis in the Dier al-Balah Governorate, Gaza Strip, Palestine. *Water* 12:262
- Elbeltagi A, Pande CB, Kouadri S et al (2021) Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India. *Environ Sci Pollut Res* (2021). <https://doi.org/10.1007/s11356-021-17064-7>
- El Bilali A, Taleb A (2020) Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *J Saudi Soc Agric Sci* 19(7):439–451
- Ewaid SH, Abed SA (2017) Water quality assessment of Al-Gharraf River, South of Iraq using multivariate statistical techniques. *J Al-Nahrain Univ* 20:114–122. <https://doi.org/10.22401/juns.20.2.16>
- Ewaid SH, Abed SA, Kadhum SA (2018) Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. *Environ Technol Innov* 11:390–398. <https://doi.org/10.1016/j.eti.2018.06.013>
- Ewaid SH, Kadhum SA, Abed SA, Salih RM (2019) Development and evaluation of irrigation water quality guide using IWQG vol I software: a case study of Al-Gharraf Canal, Southern Iraq. *Environ Technol Innov* 13:224–232. <https://doi.org/10.1016/j.eti.2018.12.001>
- Fakhre A (2014) Evaluation of hydrogeochemical parameters of groundwater for suitability of domestic and irrigational purposes: a case study from central Ganga Plain, India. *Arab J Geosci* 7:4121–4131. <https://doi.org/10.1007/s12517-013-1055-6>
- Fattahi Nafchi R, Raiesi Vanani H, Noori Pashae K et al (2021a) Investigation on the effect of inclined crest step pool on scouring protection in erodible river beds. *Nat Hazards*. <https://doi.org/10.1007/s11069-021-04999-w>
- Fattahi Nafchi R., Yaghoobi, P., Raiesi Vanani, H. et al. (2021b) Eco-hydrologic stability zonation of dams and power plants using the combined models of SMCE and CEQUALW2. *Appl Water Sci* 11, 109 Volume 11, Issue 7. [10.1007/s13201-021-01427-z](https://doi.org/10.1007/s13201-021-01427-z)
- Fiyadh SS, AlOmar MK, Binti Jaafar WZ et al (2019) Artificial neural network approach for modelling of mercury ions removal from water using functionalized CNTs with deep eutectic solvent. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20174206>
- Ghazaryan K, Chen Y (2016) Hydro-chemical assessment of surface water for irrigation purposes and its influence on soil salinity in Tikanlik oasis, China. *Environ Earth Sci* 75(5):383
- Golian et al (2020) Prediction of tunnelling impact on flow rates of adjacent extraction water wells. *Q J Eng Geol Hydrogeol* 53(2):236. <https://doi.org/10.1144/qjgegh2019-055>
- Gref K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Ho JY, Afan HA, El-Shafe AH et al (2019) Towards a time and cost effective approach to water quality index class prediction. *J Hydrol* 575:148–165. <https://doi.org/10.1016/j.jhydrol.2019.05.016>
- Jahin HS, Abuzaid AS, Abdellatif AD (2020) Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate. *Egypt Environ Technol Innov* 17. <https://doi.org/10.1016/j.eti.2019.100532>
- Javadinejad et al (2019) Using simulation model to determine the regulation and to optimize the quantity of chlorine injection in water distribution networks. *Modeling Earth Systems and Environment*. Springer International Publishing 5(3):1015–1023. <https://doi.org/10.1007/s40808-019-00587-x>
- Khadri S.F.R and Chaitanya Pande (2015a) Ground water quality mapping for Mahesh River Basin in Akola and Buldhana Districts of (MS) India using interpolation methods, *International Journal on Recent and Innovation Trends in Computing and Communication*, Volume: 3 Issue: 2, page no. 113 – 117
- Khadri S.F.R. and Chaitanya Pande (2015b), Analysis of hydro-geochemical characteristics of groundwater quality parameters in hard rocks of Mahesh River Basin, Akola, and Buldhana Dist. Maharashtra, India using geo-informatics techniques, *American Journal of Geophysics, Geochemistry and Geosystems*, Vol. 1, No. 3, 2015, pp. 105-114
- Khadri, S.F.R, Chaitanya Pande and Kanak Moharir (2013), Groundwater quality mapping of PTU-1 Watershed in Akola district of Maharashtra India using geographic information system techniques, *International Journal of Scientific & Engineering Research*, Volume 4, Issue 9, September-2013
- Kim DE, Liang S-Y, Gourbesville P, Andres L, Liu J (2020) Simple-yet-effective SRTM DEM improvement scheme for dense urban cities using ANN and remote sensing data: application to flood modeling. *Water* 12:816
- Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrol Earth Syst Sci* 22(11):6005–6022
- Kratzert F, Klotz D, Shalev G, Günter K, Nearing G (2019) Benchmarking a catchment-aware long short-term memory network (LSTM) for large-scale hydrological modeling. *Hydrol Earth Syst Sci* 23(12):1–32
- Liu P, Wang J, Sangaiah A, Xie Y, Yin X (2019) Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* 11:2058. <https://doi.org/10.3390/su11072058>
- Lu H, Ma X (2020) Hybrid decision tree-based machine learning models for short term water quality prediction. *Chemosphere* 249:126169. <https://doi.org/10.1016/j.chemosphere.2020.126169>
- Malik RN, Hashmi MZ (2017) Multivariate statistical techniques for the evaluation of surface water quality of the Himalayan foot hills streams, Pakistan. *Appl Water Sci* 7:2817–2830. <https://doi.org/10.1007/s13201-017-0532-6>
- McCulloch WS, Pitts WH (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115–133
- Mirabbasi R, Mazlounzadeh SM, Rahnema MB (2008) Evaluation of irrigation water quality using fuzzy logic. *Res J Environ Sci* 2(5):340–352
- Mishra, Binaya K., Ram K. Regmi, Yoshifumi Masago, Kensuke Fukushima, Pankaj Kumar, and Chitresh Saraswat. (2017). Assessment of Bagmati river pollution in Kathmandu Valley: scenario-based modeling and analysis for sustainable urban development, sustainability of water quality and ecology
- Moharir K, Pande C, Singh S, Choudhari P, Rawat K, Jeyakumar L (2019) Spatial interpolation approach-based appraisal of groundwater quality of arid regions in. *Aqua Journal* 68(6):431–447
- Nguyen TH, Helm B, Hettiarachchi H, Caucci S, Krebs P (2020) Quantifying the information content of a water quality monitoring network using principal component analysis: a case study of the Freiburger Mulde River Basin, Germany. *Water* 12:420
- Noori R, Yeh H-D, Abbasi M, Kachooosangi FT, Moazami S (2015) Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. *J Hydrol* 527:833–843
- Noori R, Ghiasi B, Sheikhan H, Adamowski JF (2017) Estimation of the dispersion coefficient in natural rivers using a granular computing model. *J Hydraul Eng* 143(5). [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001276](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001276)
- Ostad-Ali-Askari K, Shayan M (2021) Subsurface drain spacing in the unsteady conditions by HYDRUS-3D and artificial neural networks. *Arab J Geosci* 14:1936. <https://doi.org/10.1007/s12517-021-08336-0>
- Ostad-Ali-Askari et al (2017) Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-Rood River, Isfahan, Iran. *KSCE Journal of Civil Engineering*, Korean Society of Civil Engineers 21(1):134–140. <https://doi.org/10.1007/s12205-016-0572-8>
- Ostad-Ali-Askari et al (2018) Comparison of solutions of Saint-Venant equations by characteristics and finite difference methods for

- unsteady flow analysis in open channel. *International Journal of Hydrology Science and Technology* 8(3):229–243. <https://doi.org/10.1504/IJHST.2018.093569>
- Ostad-Ali-Askari et al (2019) Effect of management strategies on reducing negative impacts of climate change on water resources of the Isfahan-Borkhar aquifer using MODFLOW. *River Research and Applications*, John Wiley & Sons Ltd 35(6):611–631. <https://doi.org/10.1002/rra.3463>
- Othman F, Alaaeldin ME, Seyam M, Ahmed AN, Teo FY, Ming Fai C, El-Shafie A (2020) Efficient river water quality index prediction considering minimal number of inputs variables. *Engineering Applications of Computational Fluid Mechanics* 14(1):751–763
- Ouma YO, Owiti T, Kipkorir E, Kibiyi J, Tateishi R (2012) Multitemporal comparative analysis of TRMM-3B42 satellite estimated rainfall with surface gauge data at basin scales: daily, decadal and monthly evaluations. *Int J Remote Sens* 33(12):7662–7684. <https://doi.org/10.1080/01431161.2012.701347>
- Pande CB, Moharir K (2018) Spatial analysis of groundwater quality mapping in hard rock area in the Akola and Buldhana districts of Maharashtra, India, *Applied Water Science*. Springer Journal 8(4):1–17
- Chaitanya B. Pande, Kanak N. Moharir, Sudhir Kumar Singh & Bloodless Dzwairo (2019a), Groundwater evaluation for drinking purposes using statistical index: study of Akola and Buldhana districts of Maharashtra, India, *Environment, Development and Sustainability (A Multidisciplinary Approach to the Theory and Practice of Sustainable Development)* Springer Journal, <https://doi.org/10.1007/s10668-019-00531-0>
- Chaitanya B. Pande, Kanak N. Moharir, Sudhir Kumar Singh, Abhay M. Varade (2019b), An integrated approach to delineate the groundwater potential zones in Devdari watershed area of Akola district, Maharashtra, Central India in *Environment, Development, and Sustainability* Springer Journal, <https://doi.org/10.1007/s10668-019-00409-1>
- Balamurugan Panneerselvam, Kirubakaran Muniraj, Chaitanya Pande and Nagavinothini Ravichandran (2021), Prediction and evaluation of groundwater characteristics using the radial basic model in Semi-arid region, India, *International Journal of Environmental Analytical Chemistry*, <https://doi.org/10.1080/03067319.2021.1873316>.
- Praus P (2019) Principal component weighted index for wastewater quality monitoring. *Water* 11:2376
- Rahgoshay M, Feiznia S, Arian M, Hashemi SAA (2018) Modeling daily suspended sediment load using improved support vector machine model and genetic algorithm. *Environ Sci Pollut Res* 25:35693–35706. <https://doi.org/10.1007/s11356-018-3533-6>
- Ramakrishnaiah CR, Adashiv C, Ranganna G (2009) Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *Egypt J Chem* 6:523–530
- Recknagel F, Bobbin J, Whigham P, Wilson H (2002) Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. *J Hydroinf* 4(2):125–133
- Rezvan K, Fakhri Y, Mehrorang G, & Kheibar D (2016). Back propagation artificial neural network and central composite design modeling of operational parameter impact for sunset yellow and azul (II) adsorption onto MWCNT and MWCNT-Pd-NPs: Isotherm and kinetic study. *Chemometrics and Intelligent*
- Salehi-Hafshejani et al (2019) Determination of the height of the vertical filter for heterogeneous earth dams with vertical clay core. *International Journal of Hydrology Science and Technology* 9(3):221–235. <https://doi.org/10.1504/IJHST.2019.102315>
- Sar SK, Sahu M, Singh S, Diwan V, Jindal M, Arora A (2017) Assessment of uranium in ground water from Durg District of Chhattisgarh state and its correlation with other quality parameters. *J Radioanal Nucl Chem* 314:2339–2348. <https://doi.org/10.1007/s10967-017-5587-1>
- Shayannejad et al (2020) Impermanent changes investigation of shape factors of the volumetric balance model for water development in surface irrigation. *Modeling Earth Systems and Environment*, Springer Nature Switzerland AG 6(3):1573–1580. <https://doi.org/10.1007/s40808-020-00771-4>
- Shayannejad, et al., (2021), Quantity and quality modelling of groundwater to manage water resources in Isfahan-Borkhar Aquifer. *Environment, Development and Sustainability*, Springer Nature Switzerland AG. 23(3). <https://doi.org/10.1007/s10668-021-01323-1>
- Smith M (1994) Neural networks for statistical modelling (p. 235). Van Nostrand Reinhold, NY
- Su R et al (2018) Water resources and climate change. *Journal of Water and Climate Change* 9(2):239. <https://doi.org/10.2166/wcc.2018.999>
- Talebmorad H et al (2021) Evaluation of the impact of climate change on reference crop evapotranspiration in Hamedan-Bahar plain. *International Journal of Hydrology Science and Technology* 11(3):333–347. <https://doi.org/10.1504/IJHST.2021.114554>
- Tavakol M, Arjmandi R, Shayeghi M, Monavari SM, Karbassi A (2017) Application of multivariate statistical methods to optimize water quality monitoring network with emphasis on the pollution caused by fish farms. *Iran J Public Health* 46(1):83
- Tian W, Liao Z, Zhang J (2017) An optimization of artificial neural network model for predicting chlorophyll dynamics. *Ecol Model* 364:42–52
- Tiwary RK, Kumari B, Singh DB (2018) Water quality assessment and correlation study of physico-chemical parameters of Sukinda chromite mining area, Odisha, India. In: Singh VP, Yadav S, Yadava RN (eds) Singapore, 2018. *Environmental pollution*. Springer Singapore, pp 357–370
- United Nations Environment Programme (2016) A snapshot of the world's water quality: towards a global assessment. Nairobi
- Vanani et al (2017) Development of a new method for determination of infiltration coefficients in furrow irrigation with natural non-uniformity of slope, *Sustainable Water Resources Management*. 3:163–169. <https://doi.org/10.1007/s40899-017-0091-x>
- Vasanthi SS, Kumar AS (2019) Application of artificial neural network techniques for predicting the water quality index in the parakai lake, Tamil Nadu, India. *Appl Ecol Environ Res* 17(2):1947–1958
- Viotti P, Liuti G., Di Genova P., (2002) "Atmospheric urban pollution: application of an artificial neural network (ANN) to the city of Perugia" *Ecological Modelling*, Vol. 148, issue 1, 2002, pag. 27-46
- World Health Organization (2017) "Guidelines for drinking-water quality: first addendum to the fourth edition."
- Yıldız S, Karakuş CB (2019) Estimation of irrigation water quality index with development of an optimum model: a case study. *Environment, Development and Sustainability*, pp 1–16
- Yıldız S, Karakuş CB (2020) Estimation of irrigation water quality index with development of an optimum model: a case study. *Environ Dev Sustain* 22:4771–4786
- Zhang D, Lindholm G, Ratnaweera H (2018) Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *J Hydrol* 556:409–418

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