AQUA
Water Infrastructure, Ecosystems and Society

 AQUA — Water Infrastructure, Ecosystems and Society Vol 72 No 8, 1404 doi: 10.2166/aqua.2023.009

Artificial intelligence-based approach to study the impact of climate change and human interventions on groundwater fluctuations

Chetan Singla [a,*], Rajan Aggarwal[a], Samanpreet Kaur[a] and Rohit Sharma[b]

[a] Department of Soil and Water Engineering, Punjab Agricultural University, Ludhiana 141004, India
[b] Department of Processing and Food Engineering, Punjab Agricultural University, Ludhiana 141004, India
*Corresponding author. E-mail: chetan_singla@pau.edu

CS, 0000-0001-9983-7967

ABSTRACT

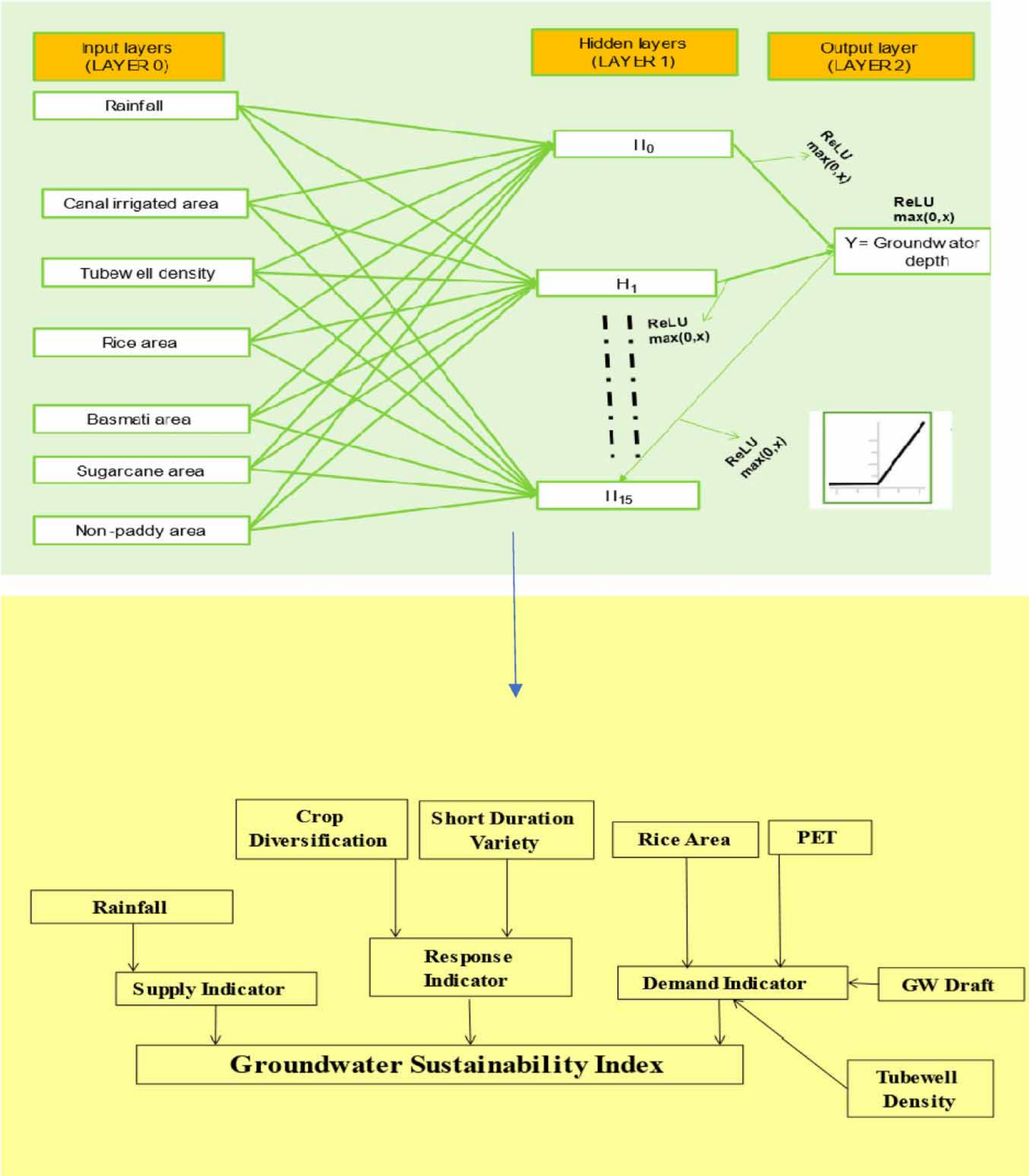
Water resource management is highly impacted by variations in rainfall, maximum and minimum temperature, and potential evapotranspiration. The rice area is also a key aspect for groundwater declination due to high-water consuming crop. Groundwater in Central Punjab has declined at an alarming rate over the last two decades. The decisions regarding water resource management need accurate information for the groundwater level. Therefore, to explore the main reason for the depletion of groundwater, it is essential that the most influential factors responsible for groundwater depletion should be addressed. A study was conducted in Central Punjab by using artificial neural network (ANN) and multiple linear regression (MLR) models during 1998–2018 to forecast the groundwater depth. ANN performed better than MLR. The sensitivity analysis showed that tubewell density, rice area, and rainfall are highly responsible for groundwater fluctuation.

Key words: climate, Mann–Kendall, rainfall, Sen's slope, temperature

HIGHLIGHTS

- In the present study, both climatic and human-induced factors were taken for groundwater modeling.
- Artificial neural network, a complex phenomenon was used to forecast groundwater depth.
- Python was used for groundwater modeling.
- ANN was found to be more accurate than MLR.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Groundwater resources accomplish one-third of Earth's freshwater demand (Brar *et al.* 2016). India uses an estimated amount of 230 km^3 of groundwater per year, making the country the leading consumer of groundwater in the world (Das & Burke 2013). Among all states of India, Punjab covers around 1.53% of India's geographical area. Punjab's climate is semi-arid (Dhillon *et al.* 2019). Agriculture in Punjab is highly intensive with heavy requirement for water which cannot be met by rainfall. In addition, surface water resources are fully utilized through canal irrigation systems and existing surface water resources are unable to meet the requirements of agriculture. Therefore, groundwater is used as the primary source of irrigation water to grow crops, mainly paddy and wheat.

The average water table depth was found to be declining at a rate of 47.6 cm/year during 1998–2013 (Brar *et al.* 2016). Analysis of 138 blocks revealed that 109 are overexploited, 2 critical, 5 semi-critical, and only 22 are safe blocks (Anon 2018).

The decisions regarding water resource management need accurate information for groundwater level. This useful information can be provided by groundwater level modeling to policy-makers to get suitable outcomes. Multiple linear regression (MLR) is a linear approach for modeling groundwater and in recent studies this method has been applied to model and analyze groundwater recharge. MLR typically estimates the level of correlation between two or more predictors (independent variable) and one response variable (dependent variable). Mogaji *et al.* (2015) evaluated and forecasted groundwater recharge rates using the MLR technique in the southern part of Perak, Malaysia. It was concluded that the developed MLR model can be used in any place with the same geology. The simulation of significant amounts of data in hydrological and groundwater systems contain non-linear relationships between variables (Osman *et al.* 2021). The limitations of non-linearity can be overcome by artificial neural network (ANN) models, which have been used by many researchers (Lai *et al.* 2019). ANNs can offer a more active strategy for forecasting groundwater amounts in a vigorous and highly unreliable system. ANN consists of three layers, namely input, hidden, and output. The total number of input variables represent the total number of neurons in the input layer. The hidden layer is connected to the input layer and the output layer is connected to the hidden layer.

Mohanty *et al.* (2015) simulated groundwater level fluctuations weekly in multiple wells located over a river basin in the Mahanadi Delta of Odisha, India. Groundwater level was forecasted 1 week in advance with a reasonable accuracy by the developed ANN model.

There is a need to develop a prediction model to study groundwater fluctuation so that the effect of sustainable use of water resources on groundwater can be studied. The main objective of this research is to develop a groundwater model in Central Punjab during 1998–2019. The aims of the present research are: (1) to evaluate and confer the effect of climatic factors and human interventions on groundwater depth and (2) to observe the main factors responsible for groundwater depth fluctuations.

2. STUDY AREA, DATASETS, AND METHODOLOGY

2.1. Study area

Broadly, Punjab can be classified into three zones, North-East, South-West, and Central. The Central Punjab comprises 10 districts as mentioned in Table 1. Central Punjab represents an area of 18,000 km^2 which covers more than 36% of the total area of Punjab. Their geographic description is given in Table 1. Keeping in view the problems of groundwater depth decline in Central Punjab, a trend analysis of 10 districts of Central Punjab was carried out.

Table 1 | Topographic features of selected districts

District name	Longitude (E)	Latitude (N)	Elevation (m.a.s.l.)	Geographical area (km^2)
Amritsar	74°52′	31°38′	219	2,647
Tarantaran	74°55′	31°27′	227	2,449
Kapurthala	75°22′	31°22′	225	1,632
Ludhiana	75°51′	30°54′	244	3,767
Sangrur	75°49’	30°13’	237	3,610
Moga	75°10′	30°48′	217	2,216
Barnala	75°32′	30°22′	229	1,410
Fatehgarh Sahib	76°22’	30°38’	246	1,180
Jalandhar	75°34′	31°19′	228	2,632
Patiala	76°22’	30°20’	257	3,218

2.2. Datasets

In the present study, important factors responsible for groundwater fluctuations in Central Punjab were used. Series of annual rainfall, paddy area, basmati area, sugarcane area, non-paddy area, canal-irrigated area, tubewell density, and groundwater depth were examined. The gridded data of rainfall from 1998 to 2019 for different districts of Punjab were collected from the India Meteorological Department, New Delhi and analyzed on an annual basis (Pai *et al.* 2014). The paddy area, basmati area, sugarcane area, non-paddy area, canal-irrigated area, and tubewell density were gathered from statistical abstracts (Statistical Abstract of Punjab 2017, 2019). Initially, the water table for various districts was computed using the Kriging technique described by Kaur *et al.* (2012). To study the behavior of groundwater, data on groundwater levels during the pre-monsoon month (June) was obtained from the Directorate of Water Resource & Environment and Directorate of Agriculture for the study period (Singla *et al.* 2022).

2.3. Methodology

ANN models can provide a more practical approach for estimating groundwater levels. Input, hidden, and output layers form an ANN model. The overall number of input variables corresponds to the overall number of input layer neurons. The input layer is connected to the hidden layer, which is connected to the output layer. There is no fixed process for the selection of the appropriate network architecture for ANN before training. 70% observation data were used for training and 30% for testing (Afzaal *et al.* 2020). The purpose of training was to adjust hidden layers such that the close approximation between the observed and predicted values is achieved. In the present study, the units of hidden neurons were figured out by trial-and-error method. The trial-and-error protocol showed that there are 16 neurons in the hidden layer. The most appropriate model is the ANN7, 16–1, which signifies that the model consists of 16 hidden layer neurons, 7 input variables, and 1 output layer (Figure 1).

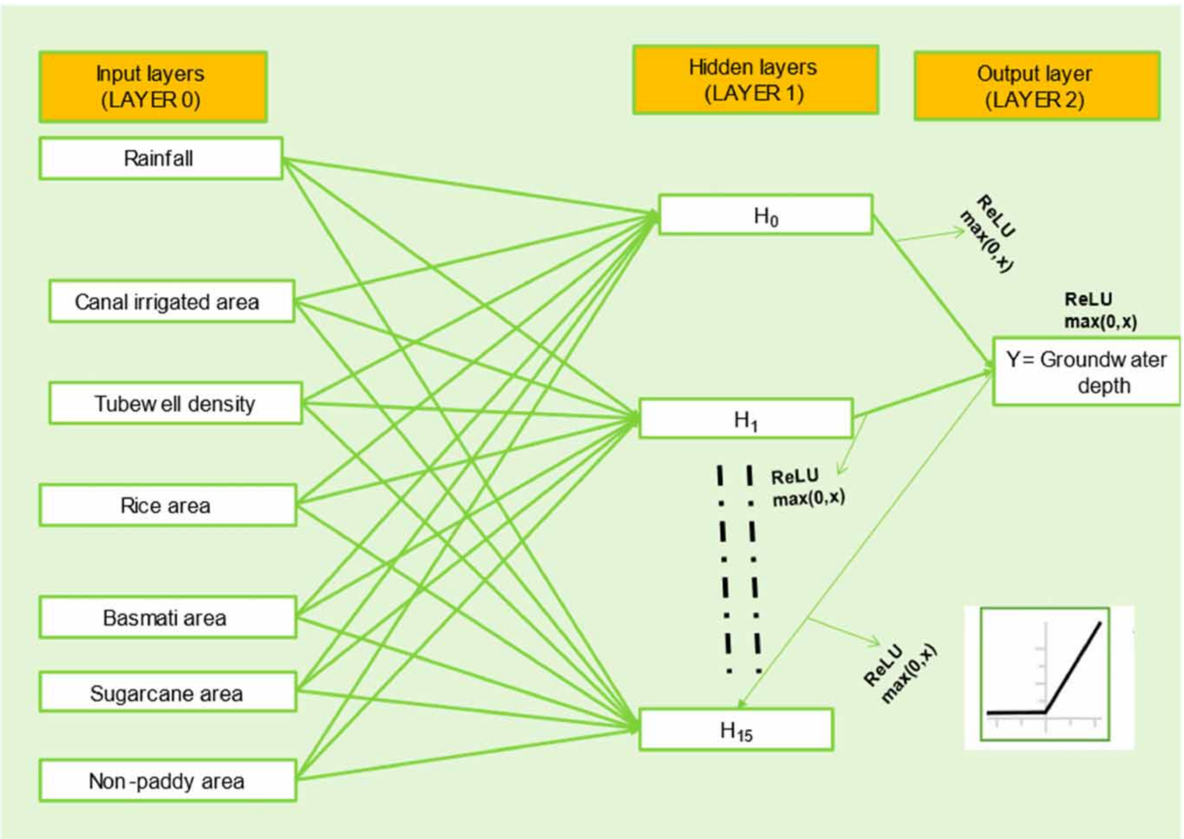


Figure 1 | Structure of the ANN used in the study.

2.4. Activation function

A Rectified Linear Unit (ReLU) function was used for the activation function in all neurons using Equation (1). Compared with logistic sigmoid function and hyperbolic tan function, the ReLU has better performance in groundwater prediction (Afzaal *et al.* 2020).

$$ReLU(x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x \leq 0 \end{cases} \quad (1)$$

2.5. Development of ANN architecture

At the beginning of the training, network weights and biases were assigned randomly. The solver for weight optimization is a stochastic gradient-based optimizer (Adam). The algorithm then projected these forwards from the input layer to the hidden layer. The results obtained from the hidden layer were propagated to the output layer. Then the error was calculated between the value of the output layer and the observed groundwater recharge in training data. The connection weights and biases were automatically adjusted until the network error reached a predetermined value by iteratively propagating the error back to the network.

The datasets of rainfall, sugarcane area, rice area, basmati area, non-paddy area, canal-irrigated area, and tubewell density were used as input and the values of groundwater depth was the output target variable. The input variables and the target output variable were standardized before the training and testing.

2.6. Multiple linear regression

MLR is the most common type of linear regression analysis, and each value of the independent variable corresponds to a value of the dependent variable. The general shape of an MLR model is depicted in the following.

$$Y = a_o + \left(\sum_{j=1}^{m} a_j\, Xj \right) \quad (2)$$

where Y is the model's output, Xj values are the independent input variables to the model, and a_o, a_1, a_2, …., a_m are partial regression coefficients. The *t*-test was used to determine any significant difference between observed groundwater depth and predicted depth.

2.7. Python scripting for ANN and MLR

The MLR model and ANN model were built using the Python packages StatsModels and scikit-learn, respectively. Tensorflow, a Python-based end-to-end machine learning platform, was used to create the model. All of the modeling, data analysis, and visualization were done in the Python 3.5 environment. Python was chosen for the work environment because of its cross-platform processing capability, open-source code, presence of third-party modules, and extensive support libraries.

2.8. Sensitivity analysis of ANN

The sensitivity analysis is very necessary to identify the most important input parameters. In ANN, the sensitivity analysis is executed after completion of the network training process. The network error (Error) for 7 input variable is noted. After that each input variable is eliminated and the network error is noted again after re-applying the training process. Now, the new network error ($Error_i$) will be noted for six input variables. There is great expectation of error to be increased after removing any variable. The new error will increase if an important variable is removed. Otherwise, the error will reduce if an unimportant input variable is removed. After that the ratio of new error and original error is noted as shown in Equation (3).

$$W = \frac{Error_i}{Error} \quad (3)$$

The larger the network error after the input variable is excluded compared to the primary error (for a network with all the input variables), the more sensitive the network is to the lack of this variable.

3. RESULTS

3.1. ANN model for Central Punjab

The observed and predicted groundwater depth during the ANN model is presented in Figure 2. Blue lines exhibited the recorded groundwater levels. Green lines presented the forecasted values predicted by ANN models. The scatter plot of recorded and forecasted groundwater depth is shown in Figure 3. The predicted groundwater depth values are uniformly distributed about the 1:1 line indicating reasonably accurate prediction. The predicted values are in close approximation with the corresponding observed values. As seen in Table 2, a low value of root mean square error (RMSE)(0.50), normalized root mean square error (NRMSE) (0.034) was found. A high value of coefficient of determination (0.98) and Nash Sutcliffe model efficiency (98.6%) was obtained. The *t*-test for simulation of groundwater depth revealed that there is no significant difference in means of observed and predicted groundwater depth at 95% confidence level as *t*-calculated is less than *t*-critical (Table 2). Mohanty *et al.* (2015) forecasted weekly groundwater levels by using the ANN approach in multiple wells located over a river basin and the statistical indicators RMSE, coefficient of determination, and NSE values were 0.4118 m, 0.9715, and 0.9288, respectively.

3.2. Multiple linear regression model for Central Punjab

The observed and predicted groundwater depth values during MLR model are given in Figure 4. The scatter plot of observed and predicted groundwater depth is shown in Figure 5. The predicted groundwater depth values are uniformly distributed about the 1:1 line indicating reasonably accurate prediction. The predicted values are in close approximation with the corresponding observed values (Table 3). The MLR model predicted groundwater depth with reasonable accuracy as is obvious from low values of RMSE (0.55), NRMSE (0.037), high values of coefficient of determination (0.98) and Nash Sutcliffe model efficiency (98.4%). Goodness-of-fit statistics for simulation of groundwater depth revealed that there is no significant difference between means of observed and predicted groundwater depth at 95% confidence level because t-critical is higher than t-calculated.

3.3. Comparison of the performance of ANN and MLR models

The ANN model has greater generalization performance as ANN used 30% data for testing. ANN showed better performance than MLR (Figure 6). Goodness-of-fit statistics for simulation of groundwater depth revealed that means of observed and predicted groundwater depth are not significantly different at 95% confidence level as t-calculated is less than t-critical for both

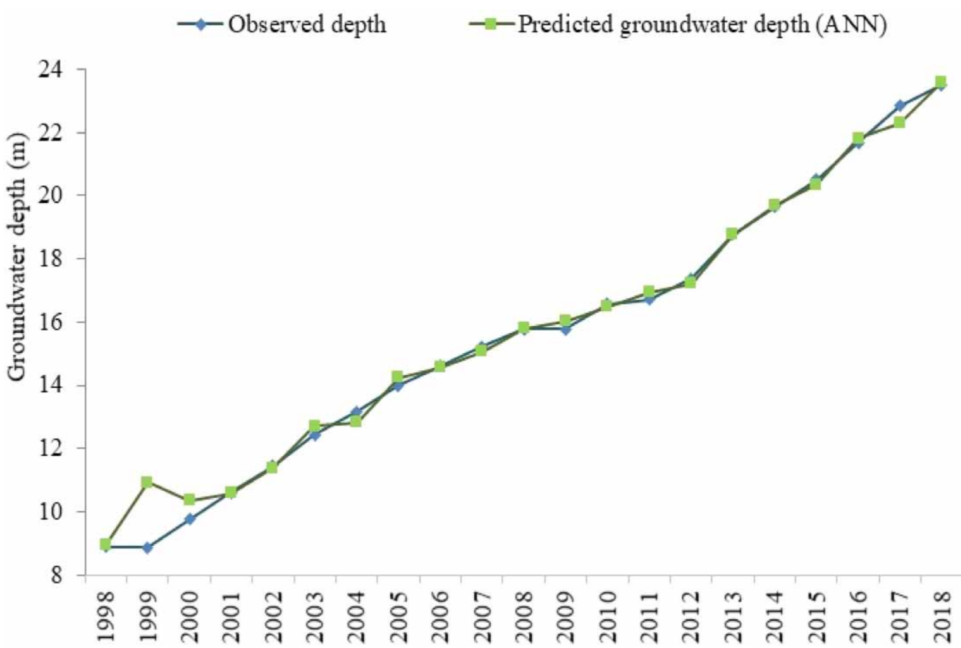


Figure 2 | Observed and predicted groundwater depth using the ANN model in Central Punjab. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/aqua.2023.009.

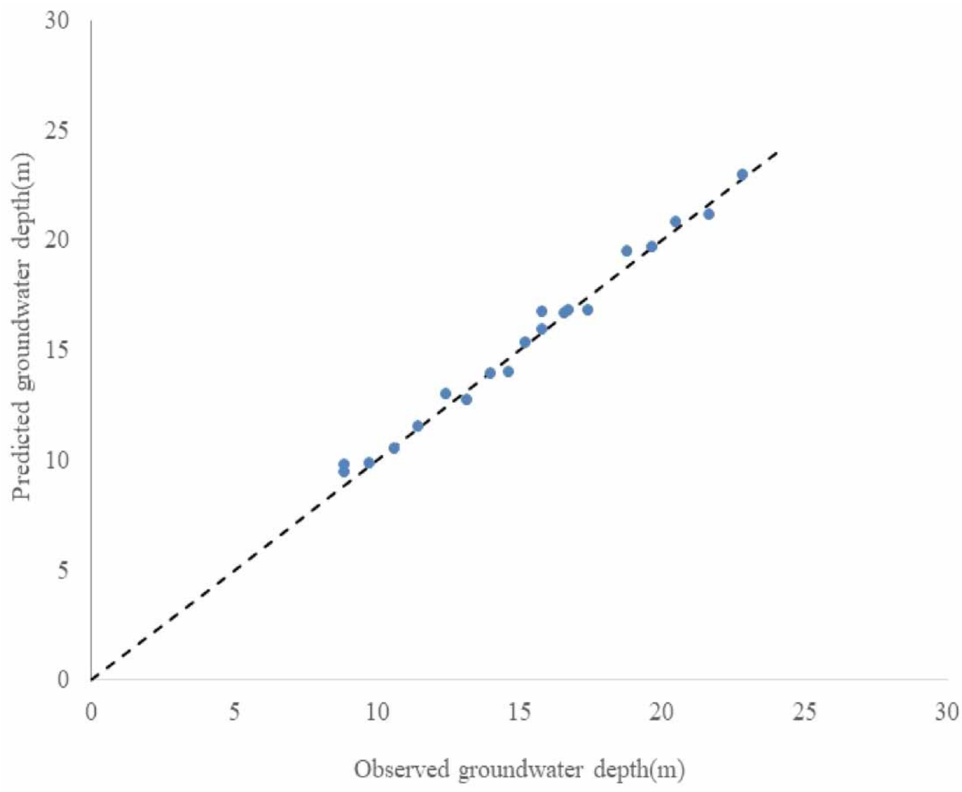


Figure 3 | Comparison between observed and predicted groundwater depth during the ANN model in Central Punjab.

Table 2 | Statistics of observed and predicted groundwater depth during the ANN model in Central Punjab

Parameter	Groundwater depth
RMSE (m)	0.50
NRMSE	0.034
Coefficient of determination	0.98
Model efficiency (%)	98.6
t-stat	-1.32
$P(T \leq t)$ two-tail	0.25
t-critical two-tail	2.09

the periods (Table 4). Pasandi *et al.* (2017) used two multivariate methods, ANN and MLR, to calculate water table depth in an unconfined aquifer situated in Shibkooh, Iran and it was revealed that ANN is efficient in the modeling of non-linear relationships.

3.4. Sensitivity analysis of ANN

Sensitivity analysis of ANN for groundwater depth was exhibited as discussed in Section 2.8 (Table 5). After eliminating each indicator one by one and running ANN with six indicators and new error is determined for each eliminated variable. Then ratios (W) of new error and original error are calculated.

The groundwater depth is most affected by tubewell density followed by rainfall, area under rice, and sugarcane respectively (Table 5). Canal-irrigated area has the lowest effect on the output. In general, it can be revealed that the importance of input parameters of the model can be differentiated with the use of sensitivity analysis for the ANN model. In addition to this, it

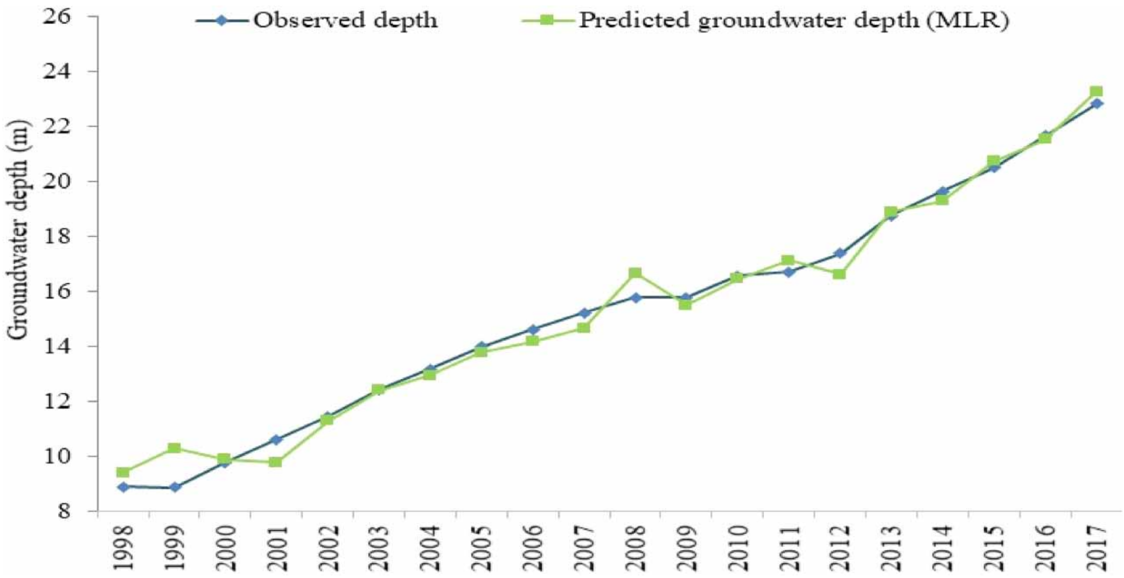


Figure 4 | Observed and predicted groundwater depth using the MLR model in Central Punjab.

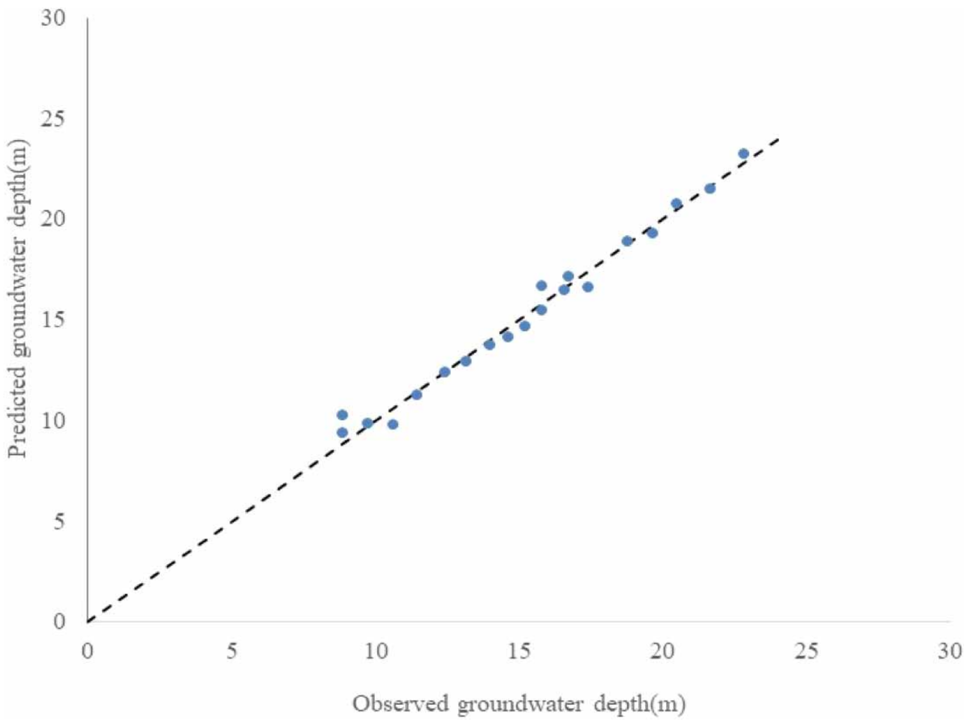


Figure 5 | Comparison between observed and predicted groundwater depth during the MLR model in Central Punjab.

should be observed that this is immensely helpful when there are various input variables for the ANN model. From the sensitivity analysis, it is noticed that tubewell density, rainfall and rice area had a major influence on groundwater depth.

4. DISCUSSION

Groundwater declination is one of the major concerns for Central Punjab, as groundwater has fallen around >1 m depth annually over the last 20 years (Singla *et al.* 2022). The ANN and MLR models were an approach to making a forecasting

Table 3 | Statistics of observed and predicted groundwater depth during the MLR model in Central Punjab

Parameter	Groundwater depth
RMSE (m)	0.55
NRMSE	0.037
Coefficient of determination	0.98
Model efficiency (%)	98.4
t-stat	-1.6×10^{-15}
$P(T \leq t)$ one-tail	0.5
t-critical one-tail	1.72
$P(T \leq t)$ two-tail	1
t-critical two-tail	2.09

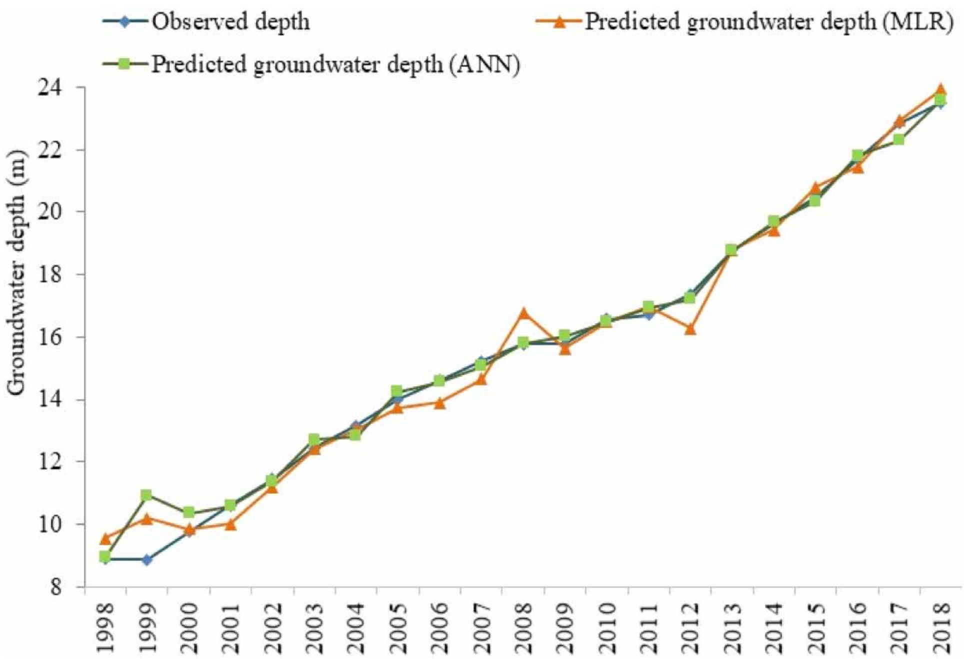


Figure 6 | Predicted groundwater depth using ANN and MLR models in Central Punjab.

model for measuring groundwater depth, as groundwater depth is influenced by both climatic factors and human interventions. So, models were developed keeping in view the influence of both factors. ANN performed better than MLR in the present study.

The sensitivity analysis showed that groundwater is greatly influenced by pumping and rainfall. This is reasonable, since more pumping is projected to increase groundwater withdrawal. The rice area is also found to be a significant influencer. The area under is approximately 40 times more than the sugarcane area (Statistical Abstract of Punjab 2017, 2019). The deficiency in rainfall over the years consequently impact evapotranspiration rates and subsequently the water requirements. Thus, the crops will need to take up more water from the groundwater. Rice area was found to be another major influencer on groundwater depth. Water productivity of rice (quantity of water required to produce 1-kg rice) in the state in the triennium (TE) ending 2013–14 was 5,337 l, whereas the all-India average was 3,875 l (Central Ground Water Board 2018). This is also due to applying a higher number of irrigations than the recommended doses. The paddy area in Central Punjab increased by 0.23 million ha during 1998–2017, indicating water requirement of (3.43 BCM (billion cubic metres)), considering an average of 1,500 mm irrigation water (Singla *et al.* 2022).

Table 4 | Statistics for predicted groundwater depth using ANN and MLR models

	Groundwater depth	
Parameter	**ANN**	**MLR**
Mean	15.75	15.63
Variance	17.88	19.39
t-stat	1.07	
$P(T \leq t)$ one-tail	0.15	
t-critical one-tail	1.72	
$P(T \leq t)$ two-tail	0.29	
t-critical two-tail	2.08	

Table 5 | Sensitivity analysis of groundwater depth using the ANN model

ANN (sensitivity analysis)	
Indicators	***W***
Tubewell density	2.84
Rainfall	2.05
Rice area	2.05
Sugarcane area	2.04
Non-paddy area	1.78
Basmati	1.65
Canal-irrigated area	1.60

The decline in water table depth can only be addressed by reducing demand on the groundwater table. It is only possible with crop diversification (Kaur & Vatta 2015). Central Punjab needs to shift a huge area from under the paddy. It is of utmost importance for ensuring sustainable agriculture, ensuring livelihoods, saving water for future generations and saving Punjab from the looming desertification.

5. CONCLUSION

The present study was done to develop a prediction model using different techniques to study groundwater fluctuations in Central Punjab. Climatic factors like rainfall and other factors like area under rice, basmati, non-paddy, sugarcane, tubewell density, and canal-irrigated area were considered as input. Both ANN and MLR techniques were used to develop the model and ANN performed better. Furthermore, sensitivity analysis revealed that rainfall, rice area, and tubewell density are the main factors that cause groundwater depletion. Furthermore, the improvements in the model can be made by taking into consideration micro scale/district level analysis. Central Punjab needs huge diversification in rice for addressing the groundwater issue. The main function of the model is to study groundwater fluctuation, when diversified rice is considered.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Afzaal, H., Farooque, A. A., Abbas, F., Acharya, B. & Esau, T. 2020 Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning. *Water* **12** (1), 5.

Anonymous. 2018 *Statistical Abstract of Punjab*. Directorate of Economics and Statistics, Government of Punjab, Chandigarh.

Brar, M. S., Aggarwal, R. & Kaur, S. 2016 GIS-based investigations on ground water behavior in Indian Punjab. *Agricultural Research Journal* **53** (4), 519–523.

Central Ground Water Board. 2018 *Ground Water Resources of Punjab State – 31st March, 2017*. Water Resources & Environment Directorate, Water Resources Department, Mohali, Punjab.

Das, S. V. G. & Burke, J. 2013 *Small Holders and Sustainable Wells. A Retrospect: Participatory Groundwater Management in Andhra Pradesh (India)*. Food and Agriculture Organization (FAO), Rome.

Dhillon, M. S., Kaur, S. & Aggarwal, R. 2019 Delineation of critical regions for mitigation of carbon emissions due to groundwater pumping in central Punjab. *Groundwater for Sustainable Development* **8**, 302–308.

Kaur, S. & Vatta, K. 2015 Groundwater depletion in Central Punjab: patterns, access and adaptations. *Current Science* **108**, 485–490.

Kaur, S., Lubana, P. P. S. & Aggarwal, R. 2012 Groundwater management for adaptation under changing climate conditions in Indian Punjab. *Journal of Water and Climate Change* **4** (1), 38–50.

Lai, V., Ahmed, A. N., Malek, M. A., Abdulmohsin, A., Ibrahim, H., K, R., El-Shafie, A. & El-Shafie, A. 2019 Modeling the nonlinearity of sea level oscillations in the Malaysian coastal areas using machine learning algorithms. *Sustainability* **11** (17), 4643.

Mogaji, K. A., Lim, H. S. & Abdullah, K. 2015 Modeling of groundwater recharge using a multiple linear regression (MLR) recharge model developed from geophysical parameters: a case of groundwater resources management. *Environmental Earth Sciences* **73**, 1217–1230.

Mohanty, S., Jha, M. K., Raul, S. K., Panda, R. K. & Sudheer, K. P. 2015 Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. *Water Resources Management* **29**, 5521–5532.

Osman, A. I., Ahmed, A. N., Chow, M. F., Huang, Y. F. & El-Shafie, A. 2021 Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal* **12** (2), 1545–1556.

Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S. & Mukhopadyay, B. 2014 Development of a new high spatial resolution ($0.25° \times 0.25°$) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *MAUSAM* **65**, 1–18.

Pasandi, M., Salmani, N. & Samani, N. 2017 Spatial estimation of water-table depth by artificial neural networks in light of ancillary data. *Hydrological Sciences Journal* **62** (12), 2012–2024.

Singla, C., Aggarwal, R. & Kaur, S. 2022 Groundwater decline in Central Punjab – is it a warning? *Groundwater for Sustainable Development* **16**, 100718. ISSN 2352-801X. https://doi.org/10.1016/j.gsd.2021.100718.

Statistical Abstract of Punjab, Anonymous. 2017 *The Economic Advisor to Govt of Punjab, Chandigarh, India*. Available from: http://www.esopb.gov.in/static/PDF/Abstract2017.pdf. (accessed 31 March 2021).

Statistical Abstract of Punjab, Anonymous. 2019 *The Economic Advisor to Govt of Punjab, Chandigarh, India*. Available from: http://www.esopb.gov.in/static/PDF/Abstract2019.pdf. (accessed 31 March 2021).

First received 23 January 2023; accepted in revised form 7 July 2023. Available online 17 July 2023