MAUSAM, **75**, 3 (July 2024), 729-746

MAUSAM

DOI : https://doi.org/10.54302/mausam.v75i3.6272
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 556.12 : 004.383.8 (540.53)

Prediction of rainfall and groundwater using machine learning Algorithms for Nagpur division

TULSHIDAS M. JIBHAKATE and YASHWANT B. KATPATAL*

Dept. of Civil Engineering, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra – 440 010, India

**Dept. of Civil Engineering, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra – 440 010, India*

(*Received 30 May 2023, Accepted 8 April 2024*)

e mails : tulshimj@gmail.com; ybkatpatal@rediffmail.com

सार – वर्षा आर्द्र भूजल का पूर्वानुमान जल संसाधन योजना के लिए आर्द्र सूखे आर्द्र बाढ़ जैसी आपदाओं के परिणामों को कम करने के लिए भी महत्वपूर्ण है। वर्तमान अध्ययन में, सकारात्मक आर्द्र नकारात्मक विसंगतियों की गणना के लिए 20 साल की अवधि (2001-2020) के लिए वर्षा विसंगति सूचकांक (RAI) का आकलनकिया गया।भूजल स्तर में उतार-चढ़ाव पर वर्षा के प्रभाव की तुलना करने के लिए अनुमानित न्यूनतम आर्द्र उच्चतम RAI वर्षों का उपयोग किया गया । मशीन लर्निंग एल्गोरिदम का उपयोग करके वर्षा आर्द्र भूजल का अनुमान लगाया गया। नाइव (N), एक्सपोनेंशियल स्मूथिंग (ES), डिसीजन ट्री रिग्रेसर (DT), रैंडम फॉरेस्टरिग्रेसर (RF), ऑटोएरिमा (AA), के-नेबर रेजिस्टर (KN), आर्द्र ग्रेडिएंट बूस्टर रेजिस्टर (GB)जैसे मशीन लर्निंग एल्गोरिदम का उपयोग करके अध्ययन क्षेत्र में वर्षा आर्द्र भूजल स्तर का अनुमानलगाने के लिए Sktime आर्द्र scikit - लर्न लाइब्रेरी का उपयोग किया गया।नागपुर भूखंड के लिए 2001 से 2015 तक प्रेक्षित मौसमी वर्षा आर्द्र भूजल डेटा के आधार पर, वर्तमान अध्ययन 2016-2020 की अवधि के लिए मूल्यों का अनुमान लगाता है।फिर, 2016-2020 के लिए प्रेक्षितआर्द्र अनुमानित मूल्यों का उपयोग करते हुए, सहसंबंध गुणांक (r), माध्य निरपेक्ष त्रुटि (MAE), नैश-सटक्लिफ दक्षता (NSE), मूल माध्य वर्ग त्रुटि (RMSE), आर्द्र टेलर आरेख जैसे सटीकता मूल्यांकन प्राचलों का सत्यापन आर्द्र सबसे अच्छे आर्द्र सबसे खराब मॉडल पूर्वानुमानकर्ताओं की जांच करने के लिए मूल्यांकन किया गया। वर्तमान अध्ययन में पाया गया है कि वर्षा के मामले में, Auto ARIMA पूर्वानुमान सबसे उपयुक्त मॉडल है, आर्द्र भूजल के मामले में, नैवेपूर्वानुमानमॉडल सबसे उपयुक्त है। डिसीजन ट्री पूर्वानुमान मॉडल वर्षा आर्द्र भूजल मानों दोनों में सबसे खराब-फिट है। फिर, वर्ष 2021-2025 के लिए क्रमशः वर्षा आर्द्र भूजल मूल्यों कापूर्वानुमान करने के लिए Auto ARIMA आर्द्र नैवे मॉडल का उपयोग किया गया। ENSO घटना 2001-2010 के दौरान अधिक प्रमुख थी, आर्द्र 2011-2020 के दौरान, दोनों ISMR को प्रभावित करने वाले प्रेरक कारक हो सकते हैं।

ABSTRACT. Rainfall and groundwater predictions are important for water resource planning and also to reduce the consequences of catastrophes like drought and floods. In the present study, Rainfall Anomaly Index (RAI) was estimated for 20 years period (2001-2020) to calculate the positive and negative anomalies. The estimated lowest and highest RAI years were used to compare the effect of rainfall on groundwater level fluctuations. Predictions of rainfall and groundwater were performed using machine learning algorithms. Sktime and scikit-learn libraries were used to predict the rainfall and groundwater levels in the study area using machine learning algorithms such as Naive (N), Exponential Smoothing (ES), Decision Tree Regressor (DT), Random Forest Regressor (RF), Auto ARIMA (AA), K-Neighbour regressor (KN), and Gradient Booster Regressor (GB). Based on the observed seasonal rainfall and groundwater data from 2001 to 2015 for Nagpur division, present study predicts values for the 2016-2020 period. Then, using observed and predicted values for 2016-2020, accuracy assessment parameters like correlation coefficient (r), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE), and Taylor diagram were assessed for validation and to investigate the best and worst model forecasters. The present study observes that in the case of rainfall, the Auto ARIMA forecaster is the best-fitted model, and in the case of groundwater, the naïve forecaster is the best-fitted model. The decision tree forecaster is the worst-fitted model in both rainfall and groundwater data. Then, the Auto ARIMA and Naïve models were used to predict rainfall and groundwater values, respectively, for the years 2021-2025. Impact of ENSO and IOD on ISMR has been assessed. The ENSO phenomenon was more prominent during 2001-2010, and during 2011-2020, both may be the driving factors impacting ISMR.

Key words – Rainfall, Rainfall anomaly index (RAI), Groundwater, Machine learning algorithms, Accuracy assessment parameters, Taylor diagram, ENSO, IOD and ISMR.

1. Introduction

Rainfall is one of the most important climatic factors or hydrological parameters. If it occurs in excess, it causes flooding, and a lack of it causes drought (Srivastava *et al*., 2015). In both situations, it causes disasters, causing loss of human life, crop destruction, livestock loss, economic impacts, water supply problems, limited growing seasons, and so on. Indian Summer Monsoon Rainfall (ISMR) is affecting the agriculture sectors which ultimately influencing the Indian economy (Mooley and Parthasarathy 1984; Webster *et al*., 1998; Kripalani *et al*., 2003; Gadgil and Gadgil 2006; Mertz *et al.*, 2009; Chaudhari *et al*., 2010; Preethi and Revadekar 2012; Prasanna 2014). For a healthy agriculture, a regular rain pattern is normally essential, but too much or too little rain can be damaging or even destroy crops. El Niño-Southern Oscillation (ENSO) is influencing the Indian summer Monsoon rainfall (Ashok *et al*., 2001 and Shukla *et al*., 2011). Variations in rainfall amounts and frequency have a direct impact on the stream flow, the spatiotemporal distribution of runoff, groundwater storage, and soil moisture (Srivastava *et al*., 2014 and Islam *et al*., 2012). In the past, various attempts have been made in India to identify regional and statewide trends in rainfall (Goswami *et al*., 2006; Guhathakurta *et al.*, 2008; Bisht *et al*., 2018; Nikumbh *et al*., 2019; and Varikoden *et al.*, 2019). Sanikhani *et al.* (2018) studied the trend analysis of rainfall patterns over Central India for the period of 1901-2010.

The term "groundwater" describes the water that exists below the earth's surface. Rain and snowmelt that seeps or infiltrates into the ground are where it all begins. Due to the different types of land surfaces present, the amount of water that seeps into the ground varies greatly from place to place. Groundwater is the purest form of water on earth, available in abundant quantity at the cheapest cost. In India, groundwater is extracted at a rate that is higher than that of any other nation in the world: 89% for irrigation, 9% for domestic consumption, and 2% for industrial uses (Mahammad and Islam 2021; Margat and Van der Gun 2013; Ahada and Suthar 2018).

The classification of the positive and negative severities of rainfall anomalies is done using the Rainfall Anomaly Index (RAI), which was introduced by Van Rooy (1965). With only one input (precipitation), RAI is simple to compute and may be examined on a monthly, seasonal and annual timeline. A statistically based model examines the characteristics of past time series of rainfall and then forecasts evolution based on characteristics.

Machine Learning is typically referred to as the most well-liked newest technologies in the fourth industrial revolution (4IR or Industry 4.0), as it gives systems the capacity to learn and improve from experience automatically without being specifically programmed. The availability of data is typically seen as being essential when developing a machine learning model or data-driven real-world systems (Sarker and Kayes 2020; Sarker *et al*., 2019). Numerous machine learning techniques have been employed in recent decades with the development of soft computational data-driven models to anticipate the dynamics of groundwater level (GWL) (Pham *et al.*, 2022). In order to predict hydrological events, the autoregressive model (AR), autoregressive moving average (ARMA) and autoregressive moving integrable average (ARIMA) models have been commonly used (Carlson *et al*., 1970; Hipel and McLeod, 1994; Box *et al*., 2015; Valipour *et al*., 2013; Zhang and Moore, 2014). One of the more successful methods among them, ARIMA is a modified version of the ARMA model and has seen a lot of use in recent years (Chattopadhyay and Chattopadhyay, 2010; Bari *et al.*, 2015; Valipour, 2015; Geetha and Nasira, 2016; Rahman *et al.*, 2017; Wanders *et al.*, 2017). Several researchers have used traditional statistical models to predict the GWL dynamics, including multivariate linear regression models (MLR) and time series modes like the autoregressive integrated moving average (ARIMA) model and seasonal autoregressive moving average (SARIMA) models (Rahaman *et al.*,2019). A few researchers used the extreme gradient boost (XGB), decision tree (DT) and random forest (RF) models to forecast GWL (Kenda *et al*., 2018, Koch *et al.*, 2019, Hikouei *et al.*, 2023). The results are greatly improved by larger temporal aggregations (monthly or annual), making statistical and machine learning techniques more beneficial for studies of water resources (Diez-Sierra and del Jesus 2020).

Using data-driven predictive analytics to make intelligent decisions is a prominent application area for machine learning (Cao 2017 and Mahdavinejad *et al*., 2018). Predictive analytics basis is the identification and utilisation of relationships between explanatory variables and predicted variables from earlier events in order to forecast the future (Han *et al.*, 2011). The quantity and quality of the data and the performance of the learning algorithms are both necessary for machine learning model to be successful (Sarker 2021).

With the development of soft computational data-driven models, studies have forecasted the dynamics of rainfall and groundwater separately using different machine learning algorithms in recent decades. Geostatistics-based comparative adequacy of the approaches for predicting GWL dynamics and the application of ML algorithms to univariate time series GWL data are, however, rare in previous research. The

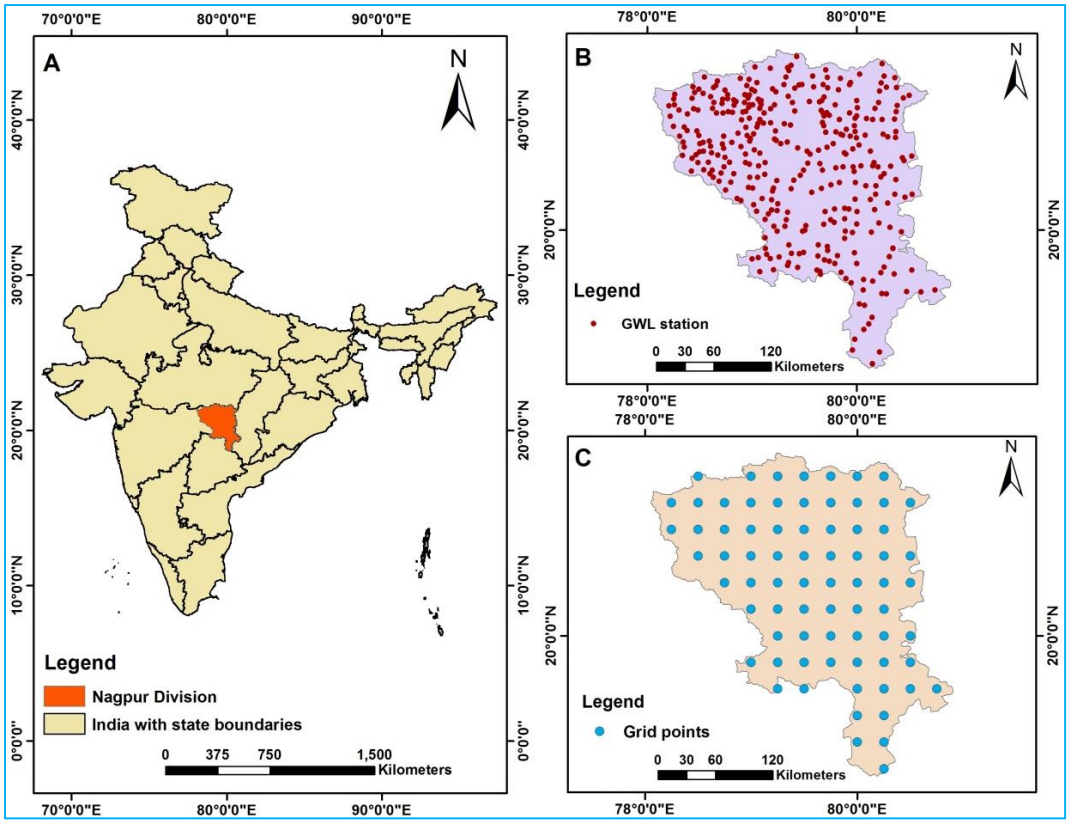

Fig. 1. Index Map of Central India (Nagpur Division), Maharashtra, India (A); Map showing the location of observation wells (GWL station) (B); and Map showing the location of grid points for rainfall data (C)

events known as the ENSO and the Indian Ocean Dipole (IOD) are closely linked to atmospheric-oceanic phenomena originate in Pacific and Indian Ocean respectively. As these phenomena affect global and regional rainfall patterns, they are extensively investigated over the few decades. Numerous research (Lau and Weng, 2001; Barsugli and Sardeshmukh, 2002; Yufu *et al.*, 2002; Hartmann *et al.*, 2008; Chattopadhyay *et al*., 2010; Shukla *et al*., 2011) have attempted to establish the link between these climatic phenomena for daily, monthly and seasonal rainfall occurrence across the world. The link between ENSO and ISMR on an intraseasonal time scale has not been extensively studied by researchers, despite the fact that their association on a seasonal scale is widely established (Joseph *et al*., 2011). Recently, it has also been discovered that IOD can have an impact on ENSO (Izumo *et al.*, 2010, 2014, 2016; Zhou *et al*., 2015; Jourdain *et al*., 2016). Both phenomena have an effect on ISMR), according to several studies carried out over the years (Ashok *et al*., 2001, Behera and Yamagata 2003). Through an atmospheric bridge, ENSO is linked to the Indian Ocean, where it has the ability to both initiate and control the evolution of IOD events (Hong *et al*., 2008; Du *et al*., 2013; Fan *et al*., 2017). Conducting regional studies has been important aspects due to varied spatial variation of rainfall occurrence in India. But, dynamics of rainfall and groundwater together has not been attempted. Hence, the dynamics of rainfall, groundwater levels, their future projections along with assessing the teleconnections of rainfall with ENSO and IOD phenomena have been explored in the present study with the following objectives;

(*i*) To assess the dynamics of rainfall and groundwater,

(*ii*) To find the relationship of rainfall and groundwater levels in the study area,

(*iii*) To attempt predictions using machine learning algorithms,

(*iv*) To find out the best and worst-fitted model for rainfall and groundwater levels.

(*v*) To investigate the relationship of ENSO and IOD on ISMR.

No similar studies have been conducted in the central India which is an agrarian region with a lot of potential for such studies.

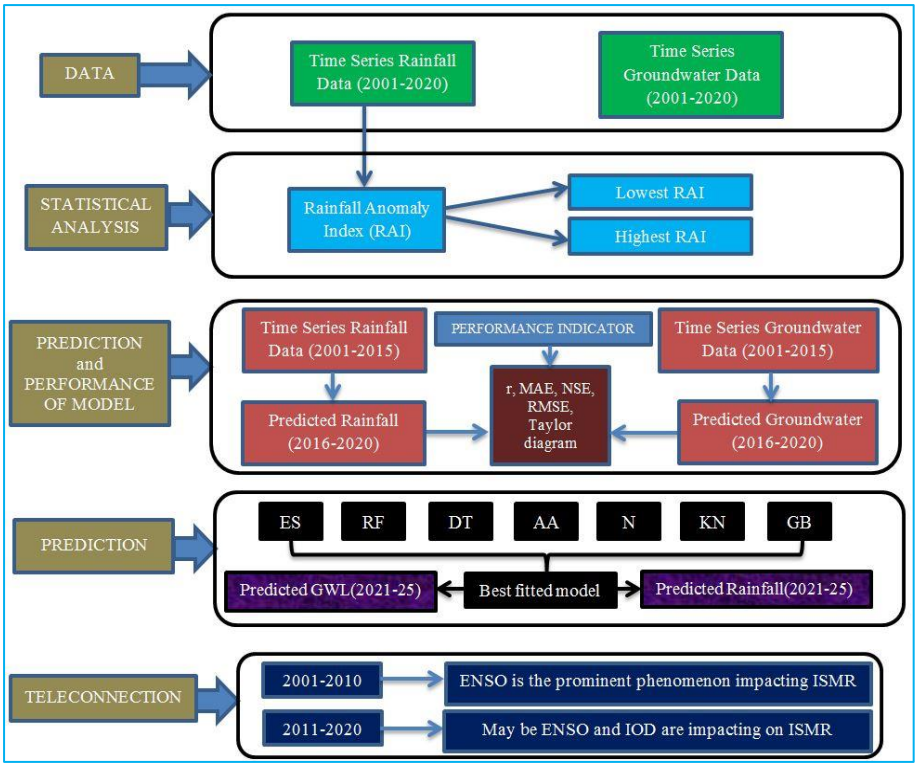

Fig. 2. Flowchart depicting the overall methodology of the present study

2. Study area

The study has been performed in central India and for this Nagpur division has been selected. The coordinates of Nagpur division in Maharashtra are 21°09′36″ N and 79°04′48″ E respectively. Six districts constitute the Nagpur division: Bhandara, Gondia, Chandrapur, Gadchiroli, Nagpur, and Wardha. It is a division of the state with abundant natural resources. The Vidarbha region is consisting of the divisions of Amravati and Nagpur. The study area along with the location of observation wells and rainfall grid points is depicted in Fig. 1.

Climatology : The southwest monsoon, which usually arrives in June and lasts until September, brings much-needed rainfall to the area. Study area experiences a tropical wet and dry climate. The highest average annual rainfall in the study area is observed in Gadchiroli district which records rainfall of 1486 mm, while the lowest rainfall of the study area of 977 mm is observed in Wardha district. The average annual rainfall of Nagpur, Bhandara, Gondia and Chandrapur district is 1094 mm, 1216 mm, 1308 mm and 1243 mm respectively. The average annual rainfall in Nagpur division which includes all the six districts, *viz*., Wardha, Nagpur, Bhandara, Gondia, Gadchiroli and Chandrapur is around 1220 mm. Almost all districts experience extreme temperature variations, with very hot summers and very cold winters. Maximum temperature observed in the region is approx. 46 °C and the minimum temperature observed is around 5 °C.

Hydrogeology : The hydrogeology within the study area varies across different districts of Nagpur division. It contains multiple principal aquifers that store and transmit groundwater. Wardha district is occupied with basaltic aquifer systems (rock formation); Nagpur district is occupied with basalt, gneiss, schist, and limestone; Bhandara district is occupied with gneiss, schist, and alluvium; Gondia district is occupied with gneiss and granite; Gadchiroli district is occupied with gneiss, granite, and sandstone; and Chandrapur district is occupied with sandstone, gneiss, limestone, shale, and basalt.

3. Data used

Monthly Rainfall and Seasonal GWL data have been downloaded from the geoportal of India Water Resources Information System (India-WRIS) for the study period of 2001-2020 over six districts that come under the Nagpur

division (Central India). The number of observation wells (GWL stations) considered for the study area is 345. The total number of IMD (India Meteorological Department) grids covered in the present study area is 73. The grid size is 0.25 degree × 0.25 degree. GWL of pre-monsoon (April-June), monsoon (July to September), post-monsoon *rabi* (October-December) and post-monsoon *kharif* (January to March) of available districts have been studied. For the above-mentioned season, seasonal average rainfall has been obtained over the study area from the monthly rainfall of each district. Niño 3.4 SST (Oceanic Niño Index-ONI) and Dipole Mode Index (DMI) data are collected for the period 2001-2020 from physical sciences laboratory (PSL), National Oceanic and Atmospheric Administration (NOAA).

4. Methodology

A detailed flowchart of the research processes adopted during the present study is shown in Fig. 2. The rainfall and groundwater level data for 20 years (2001-2020) has been selected for the analysis. Rainfall Anomaly Index, first developed and used by Van Rooy (1965) consists of the following equations for indicating positive anomalies in eqn. (1) and negative anomalies in eqn. (2):

$$\mathrm{RAI} = 3\times\left(\frac{N-\overline{N}}{\overline{M}-\overline{N}}\right)\ldots\text{for positive anomalies} \qquad (1)$$

$$\mathrm{RAI} = -3\times\left(\frac{N-\overline{N}}{\overline{X}-\overline{N}}\right)\ldots\text{for negative anomalies} \qquad (2)$$

where: N = current time series rainfall, in other words (time series-month/season/year) when RAI will be generated (mm); $\overline{N}$ = time series average rainfall of the historical series (mm); $\overline{M}$ = average of the ten highest time series precipitations of the historical series (mm); $\overline{X}$ = average of the ten lowest time series precipitations of the historical series (mm); and positive anomalies have their values above average and negative anomalies have their values below average.

To find out the best and worst forecaster model, 15 years (2001-2015) of rainfall and groundwater level data were used. The first 10 years (66.67%) data was used as training (calibrating) values and the next 5 years data (33.33%) was used as testing (validating) values to predict values from 2016 to 2020 using different machine learning algorithms. Seven machine learning (ML) algorithms, *i.e.*, Naïve (N), Decision Tree (DT), Exponential Smoothing (ES), Random Forest (RF), K-Neighbor (KN), Gradient Booster (GB) and Auto ARIMA (AA) were used to forecast rainfall and groundwater. The algorithms were implemented using the sktime and scikit-learn library in the Jupyter notebook of Anaconda navigator (version 2.3.2). Accuracy assessment parameters or Performance indicators such as correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), Nash-Sutcliffe error (NSE), and Taylor diagram were used to evaluate the output of the seven algorithms for rainfall as well as groundwater (Pham *et al*., 2022). A Taylor diagram (Taylor 2001) is a graphic design that incorporates the parameters SD, r and RMSE. A statistical measure of variation or dispersion between values in a set of data is the standard deviation. The square root of the variance is used to calculate the standard deviation. The tendency is for data points to be nearer the mean (or expected value), the lower the standard deviation. A higher standard deviation, on the other hand, denotes a wider range of values. Taylor diagram fitted with different models is assessed regarding the observed value concerning SD, r and RMSE. The model with the highest r, lowest SD and lowest RMSE is considered to be the best fit. This locates the best-fit model near the observed line. Now, data from 2001-2020 is used for the future prediction (2021-2025) of rainfall and groundwater levels using above mentioned different machine learning algorithms. The original time series rainfall and groundwater data for 20 years are split into two sets: the training set keeping 14 years (70%) of the data for fitting the models (training) and the remaining 6 years data (30%) for evaluating their prediction skill (testing). The different machine-learning algorithms used are discussed below.

Naïve Forecaster : Naïve Forecaster is a forecaster that makes forecasts using simple strategies. Forecasting is done based on naive assumptions about continuing past trends. The Naïve Forecaster can also be used for multivariate data and it then applies internally to the Column Ensemble Forecaster, so each column is forecasted with the same strategy (Hyndman and Athanasopoulos 2021).

Exponential Smoothing : It is also known as Holt-Winters exponential smoothing forecaster. It is a method for univariate time series forecasting that can be expanded to include data with a recurring pattern or seasonal component (Hyndman and Athanasopoulos 2014).

Auto ARIMA : Automatically discover the optimal order for an ARIMA model. A statistical analysis model known as an ARIMA (autoregressive integrated moving average) employs time-series data to help researchers better comprehend a data collection or forecast future trends. If a statistical model forecasts future values using data from the past, it is said to be autoregressive.

Nearest Neighbors Regression : In cases where the data labels are continuous rather than discrete variables, neighbors-based regression can be used. A query point's label is determined by computing the mean of the labels of its closest neighbors. K-Neighbors Regressor implements learning based on each query point's k nearest neighbours, where k is an integer value that the user specifies. K-Neighbour regression is used to forecast a data point based on details about the neighboring observed point (Bilali *et al.*, 2021).

Random forest : Random forests (RF), an ensemble learning technique, are a group of decision trees created using bootstrap aggregation, also known as bagging (Breiman, 2001). It is a type of ensemble machine learning used to solve problems involving classification and regression (Malakar *et al*., 2021). RF enhances the performance of the decision tree model and effectively prevents overfitting by combining all the decision trees into a single model.

Decision trees : A decision tree is a structure featuring nodes and branches which is like a flowchart (Berk, 2008; He *et al*., 2013). An efficient machine learning algorithm for supervised learning. It employs the "divide and conquer" strategy for both classification and regression (Myles *et al*., 2004). It is related to a simple tree-like structure of decision nodes and branches. It has the issue of overfitting; therefore RF is used as an assembly of several decision trees that solves this issue (Rodriguez-Galiano *et al.*, 2015). Decision trees are susceptible to overfitting and are sensitive to the training data.

Gradient boosting : It is a popular boosting algorithm. Each predictor in gradient boosting corrects the error of its predecessor (Friedman 2001). The main difference between RF and GBR is the ensemble technique while the latter is based on the boosting technique. Gradient Boosting Regressor is the class name for the gradient boosting regression in scikit-learn. The boosting method builds base models consecutively, in contrast to bagging.

In the present study, the performance of the model is evaluated using four statistical techniques: correlation coefficient, mean absolute error, root mean square error, and Nash-Sutcliffe efficiency (Ding *et al*., 2020; Yan *et al*., 2021; Fu *et al*., 2023). An important statistical method for determining the kind of relationship and the strength of it between two variables is the correlation coefficient (r) (Taylor 1990). The following eqn. (3) was utilized in the current study to calculate a correlation coefficient to assess the link between actual and predicted groundwater levels.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (3)$$

Where, r is the correlation coefficient, x_i is a sample of values for the x variable, $\bar{x}$ is the mean of those values, y_i is a sample of values for the y variable, $\bar{y}$ is the mean of those values.

The average model prediction error is expressed as mean absolute error (MAE) and root mean square error (RMSE) in units of the relevant variable.

Mean absolute error can be expressed in eqn. (4) as:

$$\mathrm{MAE} = \frac{1}{n} \times \sum_{i=1}^{n} \left| O_i - P_i \right| \quad (4)$$

where, $\left| O_i - P_i \right|$ = the absolute errors

The NSE metric is defined by Nash and Sutcliffe in 1970 and is given in eqn. (5).

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(\mathrm{OBS}_i - \mathrm{SIM}_i \right)^2}{\sum_{i=1}^{n} \left(\mathrm{OBS}_i - \overline{OBS} \right)^2} \quad (5)$$

where, OBS_i is the observation value (actual) and SIM_i is the forecast value (predicted) and $\overline{OBS}$ is average of observation values. It accepts values in the range of $-\infty$ to 1. While an NSE≤0 denotes that the assessed model itself is not a better predictor than the mean of the observed values. The better the forecast (perfect fit), the closer it is to 1, while NSE values above 0.65 signify acceptable forecasts (Moriasi *et al*., 2007; Ritter and Munoz *et al*., 2013).

Due to its adaptability to be used to many kinds of mathematical models, the NSE is a commonly used indicator in hydrology (Gupta and Kling, 2011; McCuen *et al*., 2006). Root mean square error represents the model's absolute fit to the data, or how closely the values of the observed data points fit those predicted by the model. A commonly used statistical metric for evaluating the performance of models is root mean square error (RMSE) (Chai and Draxler 2014). The following eqn. (6) was used in the current work to conduct RMSE in order to assess the model's accuracy based on the observed and predicted values.

$$\mathrm{RMSE} = \sqrt{1/n \times \sum_{i=1}^{n} \left(y_i^{\mathrm{obs}} - y_i^{\mathrm{pred}} \right)^2} \quad (6)$$

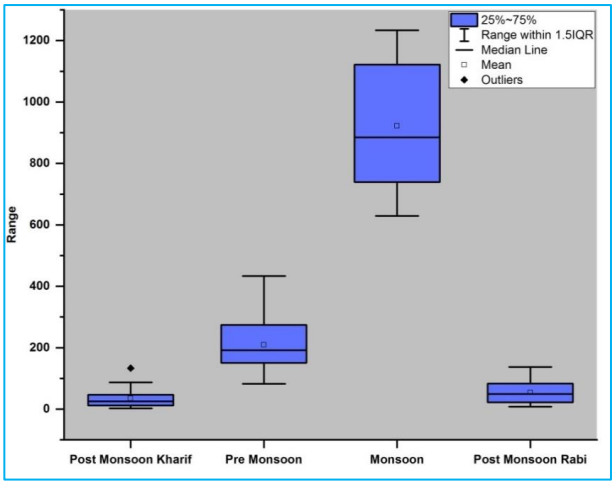

Fig. 3. Distribution of rainfall (mm) in different season from 2001 to 2020

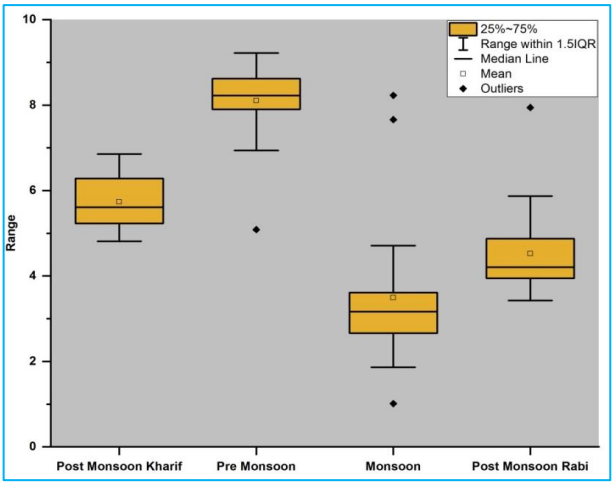

Fig. 4. Distribution of groundwater levels (m bgl) in different season from 2001 to 2020

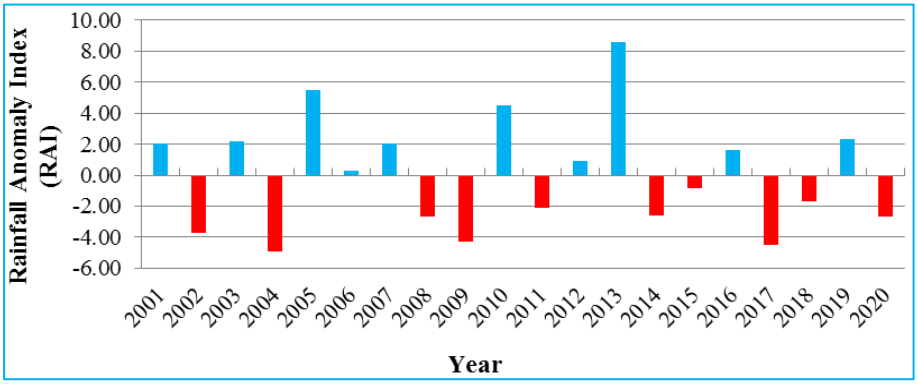

Fig. 5. Positive anomaly (blue color) and negative anomaly (red color) of the present study years

where, y_i^{obs} is the value that was observed, y_i^{pred} is the value that was predicted, and *n* is the number of samples. It accepts values in the range of 0 and +∞. The forecast is more accurate the closer it is to zero. Better fit is indicated by lower RMSE values.

5. Results and discussion

5.1. *Visual Interpretation or Dynamics of rainfall and groundwater*

The box plot of rainfall and groundwater represents the variations over the study area for the 20 years data shown in Figs. 3 and 4 respectively. One outlier can be observed in post monsoon *kharif* season of the rainfall box plot. In the groundwater box plot, there are three outliers in the monsoon season, and one outlier each in the pre-monsoon and post-monsoon *rabi* seasons. With different intensities, the positive values seen in Fig. 5 correspond to rainy or wet years, whereas the negative values correspond to dry years. It may be observed in rainfall anomaly index chart that 10 years show positive RAI values and 10 years show negative RAI values. Van Rooy (1965) classified the rainfall anomaly index into different types of humid and dry events which could be yearly, monthly, or seasonal. Classification of RAI years is given (Table 1) for the present study area. The highest positive values of RAI were 5.46, 4.51 and 8.6 for the years 2005, 2010 and 2013 respectively, classified as extremely wet (Table 1). The lowest negative values of RAI were -3.72, -4.89, -4.29 and -4.51 for the years 2002, 2004, 2009 and 2017 respectively, classified as extremely dry (Table 1). The drought years and the rainy years can be visualized using RAI, during the period from 2001 to 2020 (Table 1), where duration and intensity of the periods may also be identified. The maximum rainfall over the study area was received in 2005, 2010 and 2013 while the minimum rainfall was recorded in the years2002, 2004, 2009 and 2017. The least and most rainfall has occurred over the study area in 2004 and 2013 respectively (Fig. 5).

TABLE 1

The classification of the RAI used by Van Rooy (1965) is as follows along with present study years

	RAI Range	Classification	Study years
RAI (Rainfall Anomaly Index)	≥ 3.00	Extremely wet	2005, 2010 and 2013
	2.00 to 2.99	Very wet	2001, 2003, 2007 and 2019
	1.00 to 1.99	Moderately wet	2016
	0.50 to 0.99	Slightly wet	2012
	0.49 to –0.49	Near normal	2006
	-0.50 to –0.99	Slightly dry	2015
	-1.00 to –1.99	Moderately dry	2018
	-2.00 to –2.99	Very dry	2008, 2011, 2014 and 2020
	≤-3.00	Extremely dry	2002, 2004, 2009 and 2017

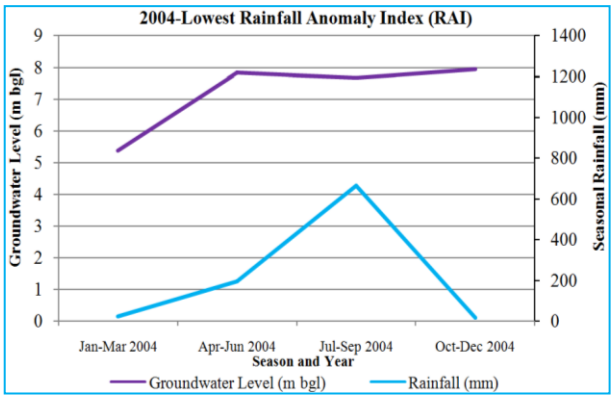

Fig. 6. Effect of rainfall on groundwater level in lowest rainfall anomaly (2004)

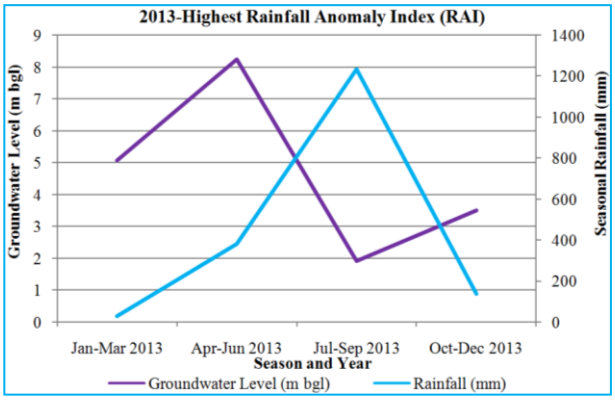

Fig. 7. Effect of rainfall on groundwater level in highest rainfall anomaly (2013)

TABLE 2

The relationship of rainfall and groundwater using Pearson correlation coefficient (r)

Pearson Correlation	Groundwater level (2001-2020)
Rainfall (2001-2020)	-0.443**

**. Correlation coefficient is significant at the 0.01 level (2-tailed)

The relationship between rainfall and groundwater has been analysed for the year 2004 (Fig. 6) when the minimum rainfall was recorded. Groundwater levels are steady during the monsoon and post-monsoon season for the years when there is less rainfall during monsoon time. It is observed that the groundwater levels are decreasing during monsoon and increasing during the post-monsoon season in the year 2013 which is the highest rainfall year (Fig. 7). Pearson correlation coefficient has been obtained to compare the rainfall and groundwater relationship within the study area which is given as Table 2.

5.2. *Findings of model and Performance of Model*

The predicted values for different machine learning algorithms with the actual values from 2016-2020 for rainfall and groundwater are shown in Table 3 and Table 4 respectively. Here, Table 5 and Table 6 indicate different accuracy assessment parameters for rainfall and

TABLE 3

The predicted rainfall values from 2016-2020 by different forecaster and actual rainfall values for the same period

Rainfall (mm) Season & Year	Gradient booster	Random Forest	Naïve	Exponential smoothing	Auto ARIMA	Decision Tree	K-Neighbor	Actual Rainfall
	Predicted Rainfall							
Jan-Mar 2016	20.3	53.6	28.1	32.5	38.0	28.1	39.1	35.8
Apr-Jun 2016	167.2	146.9	379.4	318.6	257.6	203.6	233.2	186.8
Jul-Sep 2016	948.3	1108.5	1233.7	1174.7	1224.4	750.9	1001.5	1022.2
Oct-Dec 2016	52.3	46.8	136.8	107.8	113.9	11.0	76.3	81.0
Jan-Mar 2017	93.7	14.5	62.5	49.6	30.1	133.2	23.7	11.2
Apr-Jun 2017	195.2	268.9	91.9	162.5	137.7	85.0	166.4	153.5
Jul-Sep 2017	1051.6	892.6	871.6	914.2	886.5	1233.7	934.5	708.6
Oct-Dec 2017	44.2	18.7	25.4	28.3	16.0	136.8	35.4	54.7
Jan-Mar 2018	35.5	65.3	87.2	56.4	46.7	62.5	39.8	20.2
Apr-Jun 2018	219.6	171.3	342.2	228.5	215.8	155.7	199.7	223.4
Jul-Sep 2018	1006.3	936.0	727.7	915.7	941.7	1233.7	1038.3	852.6
Oct-Dec 2018	35.0	28.4	11.0	63.0	25.5	84.8	53.9	14.5
Jan-Mar 2019	38.4	39.2	28.1	33.1	32.6	87.2	24.5	26.7
Apr-Jun 2019	203.7	209.2	379.4	324.0	323.9	203.6	204.0	82.1
Jul-Sep 2019	899.2	915.4	1233.7	1194.8	1229.5	980.0	937.9	1204.7
Oct-Dec 2019	27.2	23.9	136.8	109.6	126.4	44.3	36.7	57.6
Jan-Mar 2020	36.3	49.1	62.5	50.5	47.8	87.2	30.6	78.8
Apr-Jun 2020	210.1	186.7	91.9	165.3	112.7	85.0	203.3	200.4
Jul-Sep 2020	843.3	904.5	871.6	929.8	878.4	797.4	973.1	713.9
Oct-Dec 2020	26.6	25.6	25.4	28.8	21.1	104.0	44.9	53.8

TABLE 4

The predicted groundwater level values from 2016-2020 by different forecaster and the actual groundwater level values for the same period

Groundwater (m bgl) Year	Gradient booster	Random Forest	Naïve	Exponential smoothing	Auto ARIMA	Decision Tree	Kneighbor	Actual GWL
	Predicted GWL							
Jan-Mar 2016	4.9	5.7	5.1	5.3	5.8	5.3	5.6	6.6
Apr-Jun 2016	6.7	8.4	8.2	7.8	8.1	8.2	8.3	8.7
Jul-Sep 2016	3.1	3.2	1.9	3.4	3.5	1.9	4.6	3.4
Oct-Dec 2016	7.2	4.3	3.5	4.3	4.4	7.9	5.1	3.8
Jan-Mar 2017	6.6	5.5	4.9	5.2	5.3	6.4	5.9	5.7
Apr-Jun 2017	8.4	8.1	6.9	7.5	7.4	8.8	8.5	8.5
Jul-Sep 2017	4.6	3.9	3.8	3.2	3.7	4.7	4.4	4.4
Oct-Dec 2017	4.4	4.9	4.6	4.2	4.5	4.0	5.1	5.0
Jan-Mar 2018	5.6	6.4	5.5	5.5	5.7	4.8	5.9	6.7
Apr-Jun 2018	7.6	8.1	8.0	7.8	7.9	7.6	8.4	8.2
Jul-Sep 2018	3.0	4.3	3.1	4.1	4.7	3.5	4.6	1.0
Oct-Dec 2018	3.9	5.0	4.6	4.1	4.5	3.5	5.1	5.9
Jan-Mar 2019	5.5	6.4	5.1	5.2	5.7	5.2	5.9	6.3
Apr-Jun 2019	7.6	8.4	8.2	7.7	8.0	6.9	8.5	9.1
Jul-Sep 2019	3.8	3.9	1.9	3.4	3.5	6.9	4.4	2.4
Oct-Dec 2019	4.5	4.9	3.5	4.3	4.4	7.9	5.1	4.9
Jan-Mar 2020	5.8	6.4	4.9	5.1	5.3	6.9	5.6	5.1
Apr-Jun 2020	8.3	8.2	6.9	7.3	7.4	9.2	8.4	5.1
Jul-Sep 2020	3.2	4.3	3.8	3.1	3.7	2.8	4.6	1.9
Oct-Dec 2020	4.0	5.0	4.6	4.1	4.6	3.8	5.1	4.2

TABLE 5

Accuracy assessment values for different forecaster/predictor (Rainfall)

Algorithm	R	MAE	RMSE/RMSD	NSE
Gradient booster	0.95	74.335	119.6	0.896
Random Forest	0.964	71.81	102.527	0.923
Naïve	0.965	90.24	119.642	0.896
Exponential smoothing	0.981	68.06	101.04	0.926
AutoARIMA	0.984	67.54	97.517	0.931
Decision Tree	0.904	120.12	177.4	0.771
KN	0.961	67.255	111.221	0.91

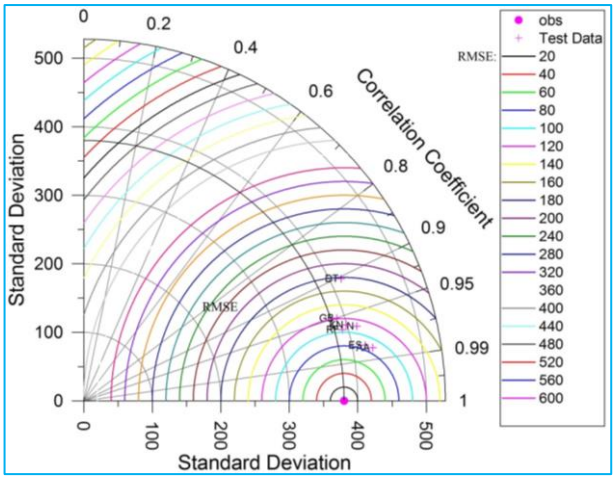

Fig. 8. Taylors diagram showing the r, SD and RMSE for rainfall (2016-2020)

TABLE 6

Accuracy assessment values for different forecaster/predictor (Groundwater)

Algorithm	R	MAE	RMSE/RMSD	NSE
Gradient booster	0.721	1.220	1.53	0.515
Random Forest	0.831	0.880	1.302	0.649
Naïve	0.864	1.015	1.176	0.713
Exponential smoothing	0.835	1.015	1.252	0.675
AutoARIMA	0.829	0.960	1.265	0.668
Decision Tree	0.507	1.755	2.179	0.015
KN	0.809	1.010	1.456	0.560

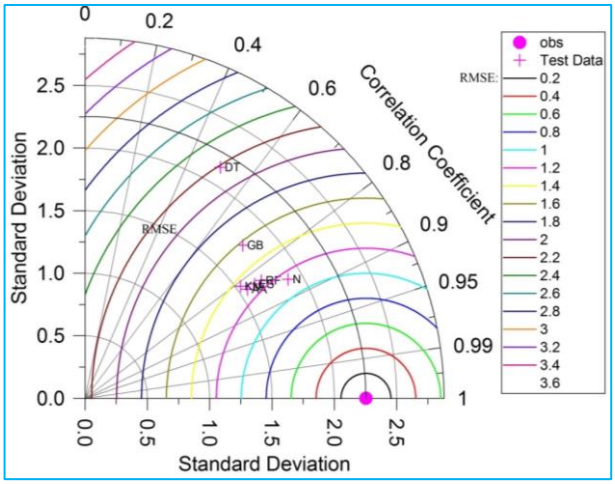

Fig. 9. Taylors diagram showing the r, SD and RMSE for groundwater (2016-2020)

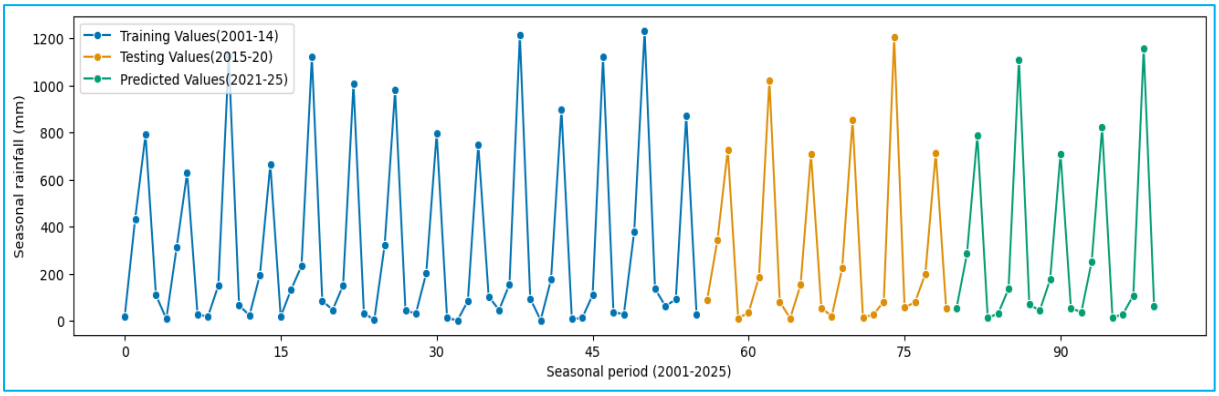

Fig. 10. The training values (2001-2014 : blue color), testing values (2015-2020 : orange color) and predicted values (2021-2025 : green color) of seasonal rainfall obtained by Auto ARIMA Forecaster

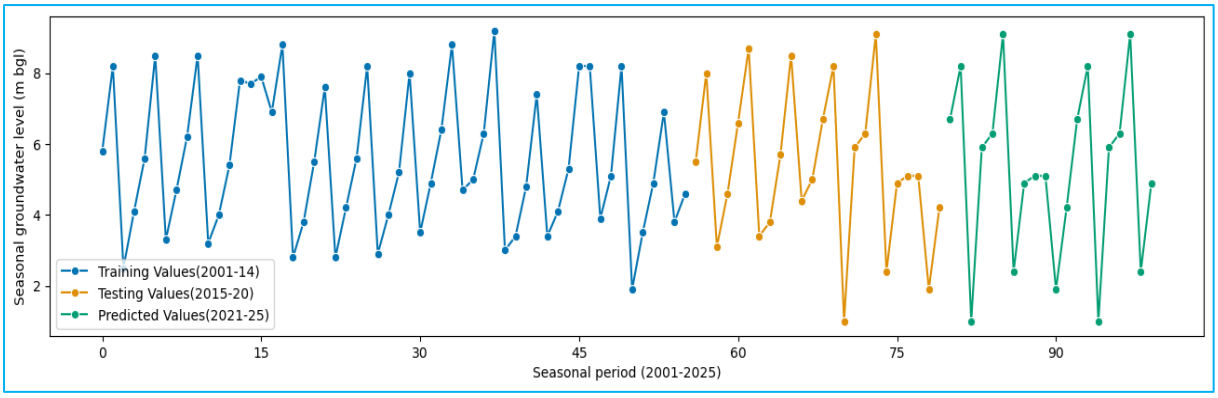

Fig. 11. The training values (2001-2014 : blue color), testing values (2015-2020 : orange color) and predicted values (2021-2025 : green color) of seasonal groundwater obtained by naive Forecaster

TABLE 7

The predicted values using best fitted Auto ARIMA and Naïve forecaster of rainfall and groundwater respectively

Season	Rainfall (mm)	Groundwater (m bgl)
Jan-Mar 2021	54.9	6.7
Apr-Jun 2021	284.9	8.2
Jul-Sep 2021	787.9	1.0
Oct-Dec 2021	12.7	5.9
Jan-Mar 2022	31.4	6.3
Apr-Jun 2022	136.3	9.1
Jul-Sep 2022	1110.2	2.4
Oct-Dec 2022	69.7	4.9
Jan-Mar 2023	43.8	5.1
Apr-Jun 2023	176.1	5.1
Jul-Sep 2023	711.1	1.9
Oct-Dec 2023	54.2	4.2
Jan-Mar 2024	36.9	6.7
Apr-Jun 2024	253.0	8.2
Jul-Sep 2024	821.4	1.0
Oct-Dec 2024	13.6	5.9
Jan-Mar 2025	29.0	6.3
Apr-Jun 2025	108.2	9.1
Jul-Sep 2025	1159.1	2.4
Oct-Dec 2025	63.4	4.9

groundwater data respectively for the period 2016-2020. Higher values of r and NSE indicate the best model performer and *vice-versa*. Lower values of MAE and RMSE indicate the best model performer and *vice-versa*. The Taylor diagram of rainfall parameter which indicates the best and worst model is shown in Fig. 8 while Fig. 9 shows the same for the groundwater parameter. Auto ARIMA and Exponential smoothing are showing close results but Auto ARIMA is best suited model for rainfall prediction if follow values of accuracy assessment parameters in Table 5 and visual interpretation in Fig. 8. Auto ARIMA, Naïve, Exponential smoothing and Random Forest are showing close results but Naive is best suited model for groundwater prediction if follow values of accuracy assessment parameters in Table 6 and visual interpretation in Fig. 9.

5.3. *Predictions of Rainfall and groundwater*

Groundwater levels and rainfall values were predicted for the period 2021-25. Future predictions of rainfall and groundwater are shown in Fig. 10 and Fig. 11 respectively. Rainfall predictions were plotted using Auto ARIMA forecaster and Groundwater predictions were plotted using Naïve forecaster as they performed as best fitted model for rainfall data and groundwater data in present study. Each point is indicating the season, *i.e.*, post-monsoon *kharif* (January to March), pre-monsoon (April-June), monsoon (July to September) and post-monsoon *rabi* (October-December) starting from 2001 to 2025. The predicted values for the period of 2021-2025 are shown in Table 7.

5.4. *Assessing impact of ENSO and IOD on ISMR*

Variations in the Eurasian snow cover (Blanford 1884; Hahn and Shukla 1976; Bamzai and Shukla 1999; Kripalani and Kulkarni 1999), IOD (Saji *et al.*, 1999;

TABLE 8

Pearson correlation coefficient (r) of ONI-rainfall and DMI-rainfall for the different study period

Year	ONI–Rainfall (r)	DMI–Rainfall (r)	ENSO Event	IOD Event
2001	0.71	0.01	Neutral	Neutral
2002	0.26	-0.10	El Niño	Neutral
2003	-0.25	0.55	Neutral	Neutral
2004	0.18	-0.31	El Niño	Neutral
2005	-0.10	-0.30	Neutral	Neutral
2006	0.16	0.16	El Niño	Positive
2007	-0.16	0.21	La Niña	Neutral
2008	0.60	0.65	La Niña	Neutral
2009	0.24	-0.81	El Niño	Neutral
2010	-0.44	-0.15	La Niña	Negative
2011	0.59	0.22	La Niña	Neutral
2012	0.67	0.85	La Niña	Positive
2013	-0.27	-0.31	Neutral	Neutral
2014	-0.07	-0.78	Neutral	Negative
2015	0.10	0.45	El Niño	Positive
2016	-0.44	-0.78	El Niño	Negative
2017	0.29	0.37	Neutral	Positive
2018	0.13	-0.09	Neutral	Positive
2019	-0.83	0.28	El Niño	Positive
2020	-0.13	-0.18	La Niña	Neutral

Kripalani and Kumar 2004), anomalies in the sea surface temperature (SST), emitted longwave radiation (OLR), wind (Sreekala 2011), topography and convective available potential energy (Goswami *et al.*, 2010) are all factors that affect the interannual variability of ISMR. Wang *et al.*, 2019 infer that ENSO has a greater impact on the eastern pole's IOD intensity than the western pole's, and its influence is stronger during a negative IOD phase than during a positive phase. Roy *et al.* (2017) found a clear connection between the Walker circulation and Indian summer monsoon rainfall around central India in models. Weak (strong) ISMR is frequently associated with El Niño (La Niña) conditions over the Pacific Ocean. According to Pothapakula *et al.*, 2020, the findings of the observations and reanalysis data, IOD and ENSO both have an impact on the interannual variability of ISMR in the majority of the Indian subcontinent. Both exhibit positive net synergy over central India, the monsoon core region, and net redundant information over the southern part of India.

Here, we have tried correlation analysis (*i.e*., Pearson correlation coefficient-5% significance level) to find out the relationship between the ONI-rainfall and the DMI-rainfall. Pearson correlation coefficient (r) is a powerful tool that allows for measuring the strength of linear relationships between two variables. A different combination of periods has been used to link the relationship. The decadal correlation coefficients of DMI-rainfall (2001-2010 and 2011-2020) are -0.01 and 0.01 respectively. The decadal correlation coefficients of ONI-rainfall (2001-2010 and 2011-2020) are -0.07 and -0.01 respectively. The correlation coefficients of ONI-rainfall and DMI-rainfall for the study period (2001-2020) are -0.04 and 0.00 respectively. Hence, it may be observed that ONI and DMI values do not always show good correlation with the ISMR. The correlation values obtained from yearly data does not give any clear indication of the impact of ENSO and IOD on ISMR. The results are shown in Table 8. From the results, it can be concluded that mere correlation analysis will not be sufficient to find

TABLE 9

ENSO and IOD events with actual and predicted rainfall for the study period

Year	ENSO event	IOD event	Actual	AutoARIMA	
2001	Neutral	Neutral	1354.4	Training (Calibrating)	
2002	El Niño	Neutral	979.5		
2003	Neutral	Neutral	1360.5		
2004	El Niño	Neutral	903.3		
2005	Neutral	Neutral	1574.6		
2006	El Niño	Positive IOD	1239.8		
2007	La Niña	Neutral	1352.0		
2008	La Niña	Neutral	1045.5		
2009	El Niño	Neutral	942.2		
2010	La Niña	Negative IOD	1513.2		
2011	La Niña	Neutral	1086.1		
2012	La Niña	Positive IOD	1281.2		
2013	Neutral	Neutral	1778.0		
2014	Neutral	Negative IOD	1051.3		
2015	El Niño	Positive IOD	1168.1		
2016	El Niño	Negative IOD	1325.8	1633.9	Testing (Validation)
2017	Neutral	Positive IOD	928	1070.3	
2018	Neutral	Positive IOD	1110.7	1229.7	
2019	El Niño	Positive IOD	1371.1	1712.4	
2020	La Niña	Neutral	1046.9	1060	
2021	La Niña	Positive IOD	1632.9	1140.4	
2022	La Niña	Positive IOD	-	1347.6	
2023	El Niño	-	-	985.2	
2024	-	-	-	1124.9	
2025	-	-	-	1359.7	

the impact of ENSO and IOD on ISMR. To understand this, many other factors (*viz*., wind, pressure, *etc*.) need to be considered while establishing the impact of ENSO and IOD on ISMR.

Average annual rainfall (normal rainfall condition) over the study area is observed to be 1213.6 mm for the study period. During 2001-2010, El Niño showed significant results giving less rainfall compared to La Niña events (Table 9) 2002, 2004 and 2009, which are El Niño years in terms of ENSO phenomenon and neutral years in terms of IOD phenomenon, observed deficit rainfall over the study area. Except 2006, which is El Niño and Positive IOD year observed normal rainfall conditions. El Niño occurrences that coincided with a Positive IOD often had normal summer monsoon rainfall conditions (Ummenhofer *et al.*, 2011). 2001, 2003 and 2005 were the neutral years with above-normal rainfall. The year 2010 was observed to have above-normal rainfall, which is a La Niña and negative IOD event (Wang *et al.*,2019). During this decade, the ENSO phenomenon was more prominent as compared to the IOD phenomenon.

Although 2011 was a La Niña year, there was less rainfall than average. Year 2013 is neutral in both phenomena, yet it observed the highest rainfall during the study period 2015 and 2019 are El Niño and positive IOD years which recorded near normal rainfall conditions. As

per the trend or pattern observed on rainfall due to the influence of ENSO and IOD, it is unfair to conclude that one phenomenon is impacting more than the other during 2011-2020 decade. The teleconnections of IOD, ENSO, and ISMR had uncertainties. In several earlier studies, both the atmospheric bridge and the oceanic channel have been presented as potential mechanisms for the influence of IOD on ENSO (Wijffels and Meyers 2004; Annamalai *et al*., 2005; Izumo *et al.*, 2010; Wieners *et al.*,2016).

The rainfall data of five years (2021-2025) was predicted using several machine learning algorithms and the years 2021, 2023 and 2024 were estimated to have below-average rainfall. Above-average rainfall is anticipated for 2022 and 2025, 2021 and 2022 were La Niña years while 2023 is declared an El Niño year by many forecasting agencies [based on input from the WMO (World Meteorological Organisation) Global Producing Centers of Long-Range Forecasts and expert assessment and by the US Climate Prediction Centre (CPC)]. As per the predicted rainfall data, 2023 is anticipated to have the lowest rainfall among the five years.

The predicted rainfall over the Central region (the study area) might be related to the associated Walker and Hadley circulations as well as mean increase in SST across the Indian and Pacific Oceans. In fact, greater amounts of greenhouse gases are known to cause climate process behaviours and associations that are possibly non-stationary, and as a result, it is probable that ENSO/IOD characteristics and forcing are going to change over time.

6. Conclusions

Prediction of rainfall and groundwater levels which are the main climatic parameters, is a prerequisite for water resource planning and management both in terms of surface water availability and groundwater availability. Based on the relationship between rainfall and groundwater levels of highest and lowest rainfall anomaly index (RAI) years, it is concluded that rainfall is affecting the groundwater level in the study area. Nowadays, Machine learning algorithms are more prominently used for predictions in many fields for better management of existing available resources. In this process, it is very important to find the best-fitted model for the prediction of rainfall as well as groundwater. In the case of rainfall predictions, the Auto ARIMA forecaster is found to be the best-fitted model and in the case of groundwater predictions, the naïve forecaster is found to be the best-fitted model. The decision tree forecaster is found to be a worst-fitted model in both the cases. As per predicted results, the year 2023 shows less rainfall while the year 2025 shows more rainfall. The lowest groundwater level (higher bgl) was observed during the pre-monsoon season of 2022 and 2025 years. The highest groundwater level (lower bgl) was observed during the pre-monsoon season of 2023. It is observed that interannual to interdecadal variations of climate drivers (ENSO and IOD) during 2001 to 2020 have impacted the ISMR over the study area. Study indicates that the ENSO phenomenon was more prominent during 2001-2010 and both may be the driving factors impacting ISMR during 2011-2020.The neutral years (2001, 2003, 2005 and 2013) in the study area during the period (2001-2020) are seen to record higher than average rainfall. Study attempts to relate the ENSO and IOD years with predicted rainfall using machine learning however a detailed analysis on regional scale is required to comment on impacts of ENSO and IOD on ISMR.

Future research may include the use of various algorithms and a hybrid model with various training and testing ratios for the prediction of rainfall and groundwater in the study area. A detailed study of the impact of ENSO and IOD on ISMR will be considered in future studies.

Acknowledgments

The authors thank the geoportal of India Water Resources Information System (India-WRIS), India Meteorological Department (IMD) and National Oceanic and Atmospheric Administration (NOAA) for data availability and the anonymous reviewers for their comments and suggestions, and their colleagues for their support.

Disclaimer : The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Ahada, C. P. and Suthar, S., 2018, "Groundwater nitrate contamination and associated human health risk assessment in southern districts of Punjab, India", *Environmental Science and Pollution Research*, **25**, 25336-25347. doi : https://doi.org/10.1007/s11356-018-2581-2.

Annamalai, H., Xie, S. P., McCreary, J. P. and Murtugudde, R., 2005, "Impact of Indian Ocean sea surface temperature on developing El Niño", *J. Clim.*, **18**, 302-319. doi : https://doi.org/10.1175/JCLI-3268.1.

Ashok, K., Guan, Z. and Yamagata, T., 2001, "Impact of the Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO", *Geophysical research letters*, **28**, 23, 4499-4502. doi : https://doi.org/10.1029/2001G L0132 94.

Bamzai, A. S. and Shukla, J., 1999, "Relation between Eurasian snow cover, snow depth, and the Indian summer monsoon: an observational study", *J. Clim*., **12**, 3117-3132. doi:

https://doi.org/10.1175/15200442(1999)012<3117:RBESCS>2.0.CO;2.

Bari, S., Rahman, M., Hussain, M. and Ray, S., 2015, "Forecasting monthly precipitation in Sylhet city using ARIMA model", *Civil and Environmental Research*, **7**, 1, 69-77.

Barsugli, J. J. and Sardeshmukh, P. D., 2002, "Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin", *Journal of Climate*, **15**, 23, 3427-3442. doi : https://doi.org/10.1175/15200442(2002)015%3C3427:GASTTS%3E2.0.CO;2.

Behera, S. K. and Yamagata, T., 2003, "Influence of the Indian Ocean dipole on the Southern Oscillation", *Journal of the Meteorological Society of Japan*, Ser. II, **81**, 1, 169-177. doi : https://doi.org/10.2151/jmsj.81.169.

Berk, R. A., 2008, "Statistical learning from a regression perspective", **14**, New York: *Springer*.

Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S. and Sridhar, V., 2018, "Spatio-temporal trends of rainfall across Indian river basins", *Theoretical and applied climatology*, **132**, 1-2, 419-436. doi : https://doi.org/10.1007/s00704-017-2095-8.

Blanford, H. H., 1884, "On the connection of Himalayan snowfall and seasons of drought in India", *Proc. R. Soc. Lond*., **37**, 3-22.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., 2015, "Time series analysis: Forecasting and control", John Wiley & Sons.

Breiman, L., 2001, "Random forests. Machine Learning", **45**, 1, 5-32. doi : https://doi.org/10.1023/a:1010933404324.

Cao, L., 2017, "Data science: a comprehensive overview", ACM Computing Surveys (CSUR), **50**, 3, 1-42. doi : https://doi.org/10.1145/3076253.

Carlson, R. F., MacCormick, A. and Watts, D. G., 1970, "Application of linear random models to four annual streamflow series", *Water Resources Research*, **6**, 4, 1070-1078. doi : https://doi.org/10.1029/WR006i004p01070.

Chai, T. and Draxler, R. R., 2014, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature", *Geoscientific model development*, **7**, 3, 1247-1250. doi : https://doi.org/10.5194/gmd-7-1247-2014.

Chattopadhyay, G., Chattopadhyay, S. and Jain, R., 2010, "Multivariate forecast of winter monsoon rainfall in India using SST anomaly as a predictor: neurocomputing and statistical approaches", *Comptes. Rendus. Geoscience*, **342**, 10, 755-765. doi : https://doi.org/10.1016/j.crte.2010.06.004.

Chattopadhyay, S. and Chattopadhyay, G., 2010, "Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN", *Comptes. Rendus. Geoscience*, **342**, 2, 100-107. doi : https://doi.org/10.1016/j.crte.2009.10.016.

Chaudhari, H. S., Shinde, M. A. and Oh, J. H., 2010, "Understanding of anomalous Indian summer monsoon rainfall of 2002 and 1994", *Quaternary International*, **213**, 1-2, 20-32. doi : https://doi.org/10.1016/j.quaint.2008.05.009.

Diez-Sierra, J. and Del Jesus, M., 2020, "Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods", *Journal of Hydrology*, **586**, 124789. doi : https://doi.org/10.1016/j.jhydrol.2020.124789.

Ding, Y. K., Zhu, Y. L., Feng, J., Zhang, P. C. and Cheng, Z. R., 2020, "Interpretable spatio-temporal attention LSTM model for flood forecasting", *Neurocomputing*, **403** (Aug): 348-359. doi : https://doi.org/10.1016/j.neucom.2020.04.110.

Du, Y., Cai, W. and Wu, Y., 2013, "A new type of the Indian Ocean Dipole since the Mid-1970s", *J. Clim*., **26**, 959-972. doi : https://doi.org/10.1175/jcli-d-12-00047.1.

El, Bilali, A., Taleb, A. and Brouziyne, Y., 2021, "Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region", *Journal of African EarthSciences*, **181**, 104244. doi : https://doi.org/10.1016/j.jafrearsci.2021.104244.

Fan, L., Liu, Q., Wang, C. and Guo, F., 2017, "Indian Ocean Dipole modes associated with different types of ENSO development", *J. Clim*., **30**, 2233-2249. doi : https://doi.org/10.1175/jcli-d-16-0426.1.

Friedman, J. H., 2001, "Greedy function approximation: a gradient boosting machine", *Annals of statistics*, 1189-1232. doi : https://www.jstor.org/stable/2699986.

Fu, Y., Zhou, X., Li, B. and Zhang, Y., 2023, "Daily Water Level Time Series Prediction Using ECRBM-Based Ensemble Optimized Neural Network Model", *Journal of Hydrologic Engineering*, **28**, 1, 04022036. doi : https://doi.org/10.1061/(ASCE)HE.1943-5584.0002219.

Gadgil, S. and Gadgil, S., 2006, "The Indian monsoon, GDP and agriculture", *Econ. Political Wkly*., **41**, 4887-4895. doi : https://www.jstor.org/stable/4418949.

Geetha, A. and Nasira, G., 2016, "Time-series modelling and forecasting: Modelling of rainfall prediction using ARIMA model", *International Journal of Society Systems Science*, **8**, 4, 361-372. doi : https://doi.org/10.1504/IJSSS.2016.081411.

Goswami, B. B., Mukhopadhyay, P., Mahanta, R.and Goswami, B. N., 2010, "Multiscale interaction with topography and extreme rainfall events in the northeast Indian region", *J. Geophys.Res*., **115**, D12114. doi : https://doi.org/10.1029/2009JD012275.

Goswami, B. N., Venugopal, V., Sengupta, D. and Madhusoodanan, M. S. and Xavier, P. K., 2006, "Increasing Trend of Extreme Rain Events Over India in a Warming Environment", *Science*, **314**, 5804, 1442-1445. doi : https://www.jstor.org/stable/20032936.

Guhathakurta, P. and Rajeevan, M., 2008, "Trends in the rainfall pattern over India", *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **28**, 1453-1469. doi : https://doi.org/10.1002/joc.1640.

Gupta, H. V. and Kling, H., 2011, "On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics", *Water Resour. Res*., **47**, W10601. doi : http://dx.doi.org/10.1029/2011WR010962.

Hahn, D. G. and Shukla, J., 1976, "An apparent relationship between Eurasian snow cover and Indian monsoon rainfall", *J. Atmos. Sci*., **33**, 2461-2462. doi : https://doi.org/10.1175/1520-0469(1976)033%3C2461:AARBES%3E2.0.CO;2.

Han, J., M. Kamber and J. Pei, 2011, "Data mining: Concepts and techniques: Concepts and techniques", 3. izd. Amsterdam: Elsevier; 2011.

Hartmann, H., Becker, S. and King, L., 2008, "Predicting summer rainfall in the Yangtze River basin with neural networks", *International Journal of Climatology*, **28**, 7, 925-936. doi : https://doi.org/10.1002/joc.1588.

He, X., Zhao, T. and Yang, D., 2013, "Prediction of monthly inflow to the Danjiangkou reservoir by distributed hydrological model

and hydroclimatic teleconnections", *Journal of Hydroelectric Engineering*, **32**, 3, 4-9.

Hikouei, I. S., Eshleman, K. N., Saharjo, B. H., Graham, L. L., Applegate, G. and Cochrane, M. A., 2023, "Using machine learning algorithms to predict groundwater levels in Indonesian tropical peatlands", *Sci. Total Environ*., **857**, 3. doi : https://doi.org/10.1016/j.scitotenv.2022.159701.

Hipel, K. W. and McLeod, A. I., 1994, "Time series modelling of water resources and environmental systems", Elsevier.

Hong, C. C., Lu, M. M. and Kanamitsu, M., 2008, "Temporal and spatial characteristics of positive and negative Indian Ocean Dipole withand without ENSO", *J. Geophys. Res. Atmos*. doi : https://doi.org/10.1029/2007J D009151.

https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

https://psl.noaa.gov/gcos_wgsp/Timeseries/Data/dmi.had.long.data

Hyndman, R. J. and Athanasopoulos, G., 2014, "Forecasting: Principles and practice", otexts. Online athttp://otexts. org/fpp.

Hyndman, R. J. and Athanasopoulos, G., 2021, "Forecasting: principles and practice", 3rdedition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 22 September 2022.

Islam, T., Rico-Ramirez, M. A., Han, D., Srivastava, P. K. and Ishak, A. M., 2012, "Performance evaluation of the TRMM precipitation estimation using ground-based radars from the GPM validation network", *Journal of Atmospheric and Solar-Terrestrial Physics*, **77**, 194-208. doi : https://doi.org/10.1016/j.jastp.2012.01.001.

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S.K., Luo, J. J., Cravatte, S., Masson, S. and Yamagata, T., 2010,"Influence of the state of the Indian Ocean Dipole on the following year's El Niño", *Nature Geoscience*, **3**, 3, 168-172. doi : https://doi.org/10.1038/ngeo760.

Izumo, T., Lengaigne, M., Vialard, J., Luo, J. J., Yamagata, T. and Madec, G., 2014, "Influence of Indian Ocean Dipole and Pacific recharge on following year's El Niño: interdecadal robustness", *Climate Dynamics*, **42**, 291-310. doi : https://doi.org/10.1007/s00382-012-1628-1.

Izumo, T., Vialard, J., Dayan, H., Lengaigne, M. and Suresh, I., 2016, "A simple estimation of equatorial Pacific response from windstress to untangle Indian Ocean Dipole and Basin influences on El Niño", *Climate dynamics*, **46**, 2247-2268. doi : https://doi.org/10.1007/s00382-015-2700-4.

Joseph, S., Sahai, A. K., Chattopadhyay, R. and Goswami, B. N., 2011, "Can El Niño–Southern Oscillation (ENSO) events modulate intraseasonal oscillations of Indian summer monsoon?", *Journal of Geophysical Research: Atmospheres*, **116**, D20. doi : https://doi.org/10.1029/2010JD015510.

Jourdain, N. C., Lengaigne, M., Vialard, J., Izumo, T. and Gupta, A. S., 2016, "Further insights on the influence of the Indian Ocean dipole on the following year's ENSO from observations and CMIP5 models", *Journal of Climate*, **29**, 2, 637-658. doi : https://doi.org/10.1175/JCLI-D-15-0481.1.

Kenda, K., ˇCerin, M., Bogataj, M., Senoˇzetnik, M., Klemen, K., Pergar, P., Laspidou, C. and Mladeni´c, D., 2018, "Groundwater Modeling with Machine Learning Techniques: Ljubljana polje Aquifer", Proceedings. **2**, 11. doi : https://doi.org/10.3390/proceedings2110697.

Koch, J., Berger, H., Henriksen, H. J. and Sonnenborg, T. O., 2019, "Modelling of the shallow water table at high spatial resolution using random forests", *Hydrol. Earth Syst. Sci*., **23**, 11, 4603-4619. doi : https://doi.org/10.5194/hess-23-4603-2019.

Lau, K. and Weng, H., 2001, "Coherent modes of global SST and summer rainfall over China: an assessment of the regional impacts of the 1997-98 El Nino", *Journal of Climate*, **14**, 6, 1294-1308. doi : https://doi.org/10.1175/1520-0442(2001)014%3C1294:CMOGSA%3E2.0.CO;2.

Kripalani, R. H. and Kulkarni, A. A., 1999, "Climatology and variability of historical Soviet snow depth data: some new perspectives in snow-Indian monsoon teleconnections", *Clim. Dyn.*, **15**, 475-489. doi : https://doi.org/10.1007/s003820050294.

Kripalani, R. H. and Kumar, P., 2004, "Northeast monsoon rainfall variability over south peninsular India *vis-à-vis* the Indian Ocean Dipole Mode", *Int. J. Climatol*., **24**, 1267-1282. doi : https://doi.org/10.1002/joc.1071.

Kripalani, R. H., Kulkarni, A., Sabade, S.S. and Khandekar, M. L., 2003, "Indian monsoon variability in a global warming scenario", *Natural hazards*, **29**, 189-206. doi : https://doi.org/10.1023/A:1023695326825.

Mahammad, S. and Islam, A. 2021, "Evaluating the groundwater quality of Damodar Fan Delta (India) using fuzzy-AHP MCDM technique", *Applied Water Science*, **11**. doi : https://doi.org/10.1007/s13201-021-01408-2.

Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P. and Sheth, A.P., 2018, "Machine learning for Internet of Things data analysis: A survey", *Digital Communications and Networks*, **4**, 3, 161-175. doi : https://doi.org/10.1016/j.dcan.2017.10.002.

Malakar, P., Sarkar, S., Mukherjee, A., Bhanja, S. and Sun, A.Y., 2021, "Use of machine learning and deep learning methods in groundwater", *In Global groundwater*, (545-557). Elsevier. doi : https://doi.org/10.1016/B978-0-12-818172-0.00040-2.

Margat, J. and Van der Gun, J., 2013, "Groundwater around the world: a geographic synopsis", Crc. Press.

Maulud, D. and Abdulazeez, A. M., 2020, "A Review on Linear Regression Comprehensive in Machine Learning", *Journal of Applied Science and Technology Trends*, **1**, 140-147. doi : https://doi.org/10.38094/jastt1457.

McCuen, R. H., Knight, Z. and Cutter, A.G., 2006, "Evaluation of the Nash–Sutcliffe Efficiency Index", *J. Hydrol. Eng*., **11**, 597-602. doi : https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597).

Mertz, O., Mbow, C., Reenberg, A. and Diouf, A., 2009, "Farmers' perceptions of climate change and agricultural adaptation strategies in rural Sahel", *Environmental management*, **43**, 804-816. doi : https://doi.org/10.1007/s00267-008-9197-0.

Mooley, D. and Parthasarathy, A, B., 1984, "Fluctuations of All-India summer monsoon rainfall during 1871-1978", *Clim. Change*, **6**, 287-301. doi : https://doi.org/10.1007/BF00142477.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R.D. and Veith, T. L.,2007, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations", *Trans. ASABE*, **50**, 885-900. doi : http://dx.doi.org/10.13031/2013.23153.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A. and Brown, S. D., 2004, "An introduction to decision tree modeling", *Journal of Chemometrics: A Journal of the Chemometrics Society*, **18**, 6, 275-285. doi : https://doi.org/10.1002/cem.873.

Nikumbh, A. C., Chakraborty, A. and Bhat, G. S., 2019, "Recent spatial aggregation tendency of rainfall extremes over India", *Scientific reports*, **9**, 1, 10321. doi : https://doi.org/10.1038/s41598-019-46719-2.

Pham, Q. B., Kumar, M., Nunno, F. D., Elbeltagi, A., Granata, F., Islam, A. R., Anh, D. T., 2022, "Groundwater level prediction using machine learning algorithms in a drought-prone area", *Neural Computing and Applications*, **34**, 10751-10773. doi : https://doi.org/10.1007/s00521-022-07009-7.

Pothapakula, P. K., Primo, C., Sørland, S. and Ahrens, B., 2020, "The synergistic impact of ENSO and IOD on Indian summer monsoon rainfall in observations and climate simulations-an information theory perspective", *Earth System Dynamics*, **11**, 4, 903-923. doi : https://doi.org/10.5194/esd-11-903-2020.

Prasanna, V., 2014, "Impact of monsoon rainfall on the total foodgrains yield over India", *J. Earth Sys. Sci*., **123**, 5, 1129-1145. doi : https://doi.org/10.1007/s12040-014-0444-x.

Preethi, B. and Revadekar, J. V., 2012, "Kharif food grain yield and daily summer monsoon precipitation over India", *Int. J .Climatol*. doi : https://doi.org/10.1002/joc.3565.

Rahman, M. A., Yunsheng, L. and Sultana, N., 2017, "Analysis and prediction of rainfall trends over Bangladesh using Mann-Kendall, Spearman's rho tests and ARIMA model", *Meteorology and Atmospheric Physics*, **129**, 4, 409-424. doi ; https://doi.org/10.1007/s00703-016-0479-4.

Ritter, A. and Munoz-Carpena, R., 2013, "Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments", *Journal of Hydrology*, **480**, 33-45. doi : https://doi.org/10.1016/j.jhydrol. 2012.12.004.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M.J.O.G.R., 2015, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines", *Ore Geology Reviews*, **71**, 804-818. doi ; https://doi.org/10.1016/j.oregeorev.2015.01.001.

Roy, I., Tedeschi, R. G. and Collins, M., 2017, "ENSO teleconnections to the Indian summer monsoon in observations and models", *International Journal of Climatology*, **37**, 4, 1794-1813. doi : https://doi.org/10.1002/joc.4811.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N. and Yamagata, T., 1999, "A dipole mode in the tropical Indian Ocean", *Nature*, **401**, 360-363. doi : https://doi.org/10.1038/43854.

Sanikhani, H., Kisi, O., Mirabbasi, R. and Meshram, S. G., 2018, "Trend analysis of rainfall pattern over the Central India during 1901-2010", *Arabian Journal of Geosciences*, **11**, 15, 437. doi : https://doi.org/10.1007/s12517-018-3800-3.

Sarker, I. H. and Kayes, A.S.M., 2020, "ABC-RuleMiner: User behavioral rule-based machine learning method for context-aware intelligent services", *Journal of Network and Computer Applications*, 168, 102762. doi : https://doi.org/10.1016/j.jnca. 2020.102762.

Sarker, I. H., 2021, "Machine learning: Algorithms, real-world applications and research directions", *SN computer science*, **2**, 3, 160. doi : https://doi.org/10.1007/s42979-021-00592-x.

Sarker, I. H., Kayes, A.S.M. and Watters, P., 2019, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage", *Journal of Big Data*, **6**, 1, 1-28. doi : https://doi.org/10.1186/s40537-019-0219-y.

Shukla, R. P., Tripathi, K. C., Pandey, A. C., Das, I.M.L., 2011, "Prediction of Indian summer monsoon rainfall using Niño indices: a neural network approach", *Atmospheric Research*,**102**, 1-2, 99-109. doi : https://doi.org/10.1016/ j.atmosres.2011.06.013.

Sreekala, P. P., Rao, S. V. B. and Rajeevan, M., 2011, "Northeast monsoon rainfall variability over south peninsular India and its teleconnections", *Theor. Appl.Climatol*., **108**, 73-83. doi : https://doi.org/10.1007/s00704-011-0513-x,

Srivastava, P. K., Han, D., Rico-Ramirez, M. A. and Islam, T., 2014, "Sensitivity and uncertainty analysis of mesoscale model downscaled hydro-meteorological variables for discharge prediction", *Hydrological Processes*, **28**, 4419-4432. doi : https://doi.org/10.1002/hyp.9946.

Srivastava, P. K., Mehta, A., Gupta, M., Singh, S. K. and Islam, T., 2015, "Assessing impact of climate change on Mundra mangrove forest ecosystem, Gulf of Kutch, western coast of India: a synergistic evaluation using remote sensing", *Theoretical and Applied Climatology*, **120**, 3-4, 685-700. doi : https://doi.org/10.1007/s00704-014-1206-z.

Taylor, K. E., 2001, "Summarizing multiple aspects of model performance in a single diagram", *Journal of geophysical research: atmospheres*, **106**, D7, 7183-7192. doi : https://doi.org/10.1029/2000JD900719.

Taylor, R., 1990, "Interpretation of the correlation coefficient: a basic review", *Journal of diagnostic medical sonography*, **6**, 1, 35-39. doi : https://doi.org/10.1177/875647939000600106.

Ummenhofer, C. C., Gupta, A. S., Li, Y., Taschetto, A. S. and England, M. H., 2011, "Multi-decadal modulation of the El Niño-Indian monsoon relationship by Indian Ocean variability", *Environmental Research Letters*, **6**, 3, p034006. doi : 10.1088/1748-9326/aba137.

Valipour, M., 2015, "Long-term runoff study using SARIMA and ARIMA models in the United States", *Meteorological Applications*, **22**, 3, 592-598.https://doi.org/10.1002/met.1491.

Valipour, M., Banihabib, M. E. and Behbahani, S. M. R., 2013, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir", *Journal of Hydrology*, **476**, 433-441. doi : https://doi.org/10.1016/j.jhydrol.2012.11.017.

Van Rooy, M. P., 1965, "A rainfall anomally index independent of time and space", Notos.

Varikoden, H., Revadekar, J. V., Kuttippurath, J. and Babu, C. A., 2019, "Contrasting trends in southwest monsoon rainfall over the Western Ghats region of India", *Climate Dynamics*, **52**, 7-8, 4557-4566. doi : https://doi.org/10.1007/s00382-018-4397-7.

Wanders, N., Bachas, A., He, X. G., Huang, H., Koppa, A., Mekonnen, Z. T., Pagán, B. R., Peng, L. Q., Vergopolan, N., Wang, K. J., Xiao, M.,Zhan, S., Lettenmaier, D. P. and Wood, E. F., 2017,"Forecasting the Hydroclimatic signature of the 2015/16 El Niño event on the Western United States", *Journal of Hydrometeorology*, **18**, 1, 177-186. doi : https://doi.org/ 10.1175/JHM-D-16-0230.1.

Wang, H., Kumar, A., Murtugudde, R., Narapusetty, B. and Seip, K. L., 2019, "Covariations between the Indian Ocean dipole and ENSO: a modeling study", *Climate Dynamics*, **53**, 5743-5761. doi : https://doi.org/10.1007/s00382-019-04895-x.

Webster, P. J., Magana, V. O., Palmer, T. N., Shukla, J., Tomas, R. A., Yanai, M. U. and Yasunari, T., 1998, "Monsoons: Processes, predictability, and the prospects for prediction", *Journal of Geophysical Research: Oceans*, **103**, C7, 14451-14510. doi : https://doi.org/10.1029/97JC02719.

Wieners, C. E., de Ruijter, W. P. M., Ridderinkhof, W., von der, Heydt A. S. and Dijkstra, H. A., 2016, "Coherent tropical Indo-Pacific interannual climate variability", *J. Clim*., **29**, 4269-4291. doi : https://doi.org/10.1175/JCLI-D-15-0262.1.

Wijffels, S. E. and Meyers, G. M., 2004, "An intersection of oceanic wave guides: variability in the Indonesian through flow region", *J. Phys. Oceanogr*., **34**, 1232-1253. doi : https://doi.org/10.1175/15200485(2004)034%3C1232:AIOOWV%3E2.0.CO;2.

Yan, L., Chen, C. W., Hang, T. T. and Hu, Y. C., 2021, "A stream prediction model based on attention-LSTM", *Earth Sci. Inform*.,**14**, 2, 723-733. doi : https://doi.org/10.1007/s12145-021-00571-z.

Yufu, G., Yan, Z. and Jia, W., 2002, "Numerical simulation of the relationships between the 1998 Yangtze River valley floods and SST anomalies", *Advances in Atmospheric Sciences*, **19**, 3, 391-404. doi : https://doi.org/10.1007/s00376-002-0074-0.

Zhang, Z. and Moore, J. C., 2014, "Mathematical and physical fundamentals of climate change", Elsevier.

Zhou, Q., Duan, W., Mu, M. and Feng, R., 2015, "Influence of positive and negative Indian Ocean Dipoles on ENSO via the Indonesian Throughflow: Results from sensitivity experiments", *Advances in Atmospheric Sciences*, **32**, 783-793. doi : https://doi.org/10.1007/s00376-014-4141-0.