scientific reports

OPEN

Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas

Owais Ali Wani[1], Syed Sheraz Mahdi[1,9]✉, Md. Yeasin[2], Shamal Shasang Kumar[3], Alexandre S. Gagnon[4], Faizan Danish[5], Nadhir Al-Ansari[6]✉, Salah El-Hendawy[7] & Mohamed A. Mattar[8]✉

Predicting rainfall is a challenging and critical task due to its significant impact on society. Timely and accurate predictions are essential for minimizing human and financial losses. The dependence of approximately 60% of agricultural land in India on monsoon rainfall implies the crucial nature of accurate rainfall prediction. Precise rainfall forecasts can facilitate early preparedness for disasters associated with heavy rains, enabling the public and government to take necessary precautions. In the North-Western Himalayas, where meteorological data are limited, the need for improved accuracy in traditional modeling methods for rainfall forecasting is pressing. To address this, our study proposes the application of advanced machine learning (ML) algorithms, including random forest (RF), support vector regression (SVR), artificial neural network (ANN), and k-nearest neighbour (KNN) along with various deep learning (DL) algorithms such as long short-term memory (LSTM), bi-directional LSTM, deep LSTM, gated recurrent unit (GRU), and simple recurrent neural network (RNN). These advanced techniques hold the potential to significantly improve the accuracy of rainfall prediction, offering hope for more reliable forecasts. Additionally, time series techniques, including autoregressive integrated moving average (ARIMA) and trigonometric, Box-Cox transform, arma errors, trend, and seasonal components (TBATS), are proposed for predicting rainfall across the altitudinal gradients of India's North-Western Himalayas. This approach can potentially revolutionise how we approach rainfall forecasting, ushering in a new era of accuracy and reliability. The effectiveness and accuracy of the proposed algorithms were assessed using meteorological data obtained from six weather stations at different elevations spanning from 1980 to 2021. The results indicate that DL methods exhibit the highest accuracy in predicting rainfall, as measured by the root mean squared error (RMSE) and mean absolute error (MAE), followed by ML algorithms and time series techniques. Among the DL algorithms, the accuracy order was bi-directional LSTM, LSTM, RNN, deep LSTM, and GRU. For the ML algorithms, the accuracy order was ANN, KNN, SVR, and RF. These findings suggest that altitude significantly affects the accuracy of the models, highlighting the need for additional weather stations in this mountainous region to enhance the precision of rainfall prediction.

[1]Division of Agronomy, Faculty of Agriculture Wadoora, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), Jammu and Kashmir 193201, India. [2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India. [3]Department of Agronomy (Rootcrops), Ministry of Agriculture & Waterways (MOA & W), Suva City 679, Fiji. [4]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK. [5]Department of Mathematics, School of Advanced Sciences, VIT-AP University, Inavolu, Andhra Pradesh 522237, India. [6]Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187 Lulea, Sweden. [7]Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia. [8]Department of Agricultural Engineering, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia. [9]Advanced Centre for Rainfed Agriculture (ACRA), Dhiansar, Bari-Brahmana-181133, SKUAST-Jammu, UT-J&K, India. ✉email: syedapbau@gmail.com; nadhir.alansari@ltu.se; mmattar@ksu.edu.sa

Climate change implications on food security are significantly connected to its profound impact on agriculture. Anticipating the conditions for the upcoming planting season becomes challenging due to the prevailing uncertainty caused by the unpredictable nature of climate variability, often proving detrimental to agricultural activities[1]. As a result, farmers and farming decision-makers heavily rely on their understanding of regional climatic patterns when making crucial decisions about ploughing, seeding, and managing their crops. However, traditional approaches have become less reliable in the changing climate[2]. Enhanced climate predictions offer hope for improving decision-making in the agricultural sector. These advanced forecasts have the potential to mitigate the adverse effects of factors such as a poor or delayed monsoon season and provide an opportunity to leverage projected favourable weather conditions[3]. Embracing these predictions can help farmers and other agricultural professionals navigate uncertainty more effectively, safeguarding their crops and yields while laying the groundwork for sustainable farming practices that adapt to climate change[4]. Additionally, heavy precipitation can lead to flooding, impacting infrastructure, transport networks, and human livelihood[5]. Therefore, it would be advantageous to the decision-making process if the potential magnitude of rainfall over a region could be quantified in advance. Predicting rainfall is crucial for improving agricultural output and ensuring a country's residents' access to food and clean water[6].

The link between climate predictions and agriculture highlights the crucial role of accurate forecasts in shaping the future of food security. In recent decades, improving rainfall forecasting has been a focal point in the scientific community[7]. To improve the accuracy of rainfall forecasts, it is essential to integrate various data sources and utilise advanced modeling techniques. Numerical weather prediction models utilise mathematical equations to simulate the atmosphere's behaviour and interactions with various climatic factors. These models incorporate historical data, real-time observations from weather stations, satellite imagery, and sophisticated computational algorithms[8]. Rainfall is intricately linked to a network of climatic factors, each influencing and being influenced by others. These factors include maximum and minimum temperatures, atmospheric pressure, relative humidity, and wind speed[9]. The interconnectedness of these variables creates a complex system that plays a crucial role in maintaining overall climate balance. Numerical weather prediction models aim to provide more precise forecasts by examining the intricate relationships between climatic elements and rainfall. The intricate and ever-changing nature of the atmosphere and the complex interplay of various climatic factors have challenged the improvement of rainfall forecasting. Although there has been notable progress in enhancing rainfall prediction techniques, several factors continue to contribute to the complexity of the task. One significant obstacle is the inherent unpredictability of atmospheric processes[10]. The atmosphere is dynamic and chaotic, with sudden shifts and unexpected interactions. This unpredictability suggests that even minor variations in initial conditions can lead to vastly different outcomes over time[11]. This phenomenon presents a substantial challenge for meteorologists and climate scientists working to create accurate models for predicting rainfall patterns. Another crucial aspect complicating rainfall forecasting is the need for long-term historical data to construct reliable prediction models. Historical data aids scientists in identifying rainfall patterns, trends, and cycles. However, obtaining comprehensive and high-quality historical data is not always straightforward, particularly in areas with limited monitoring infrastructure[12].

Rainfall prediction involves complex stochastic and nonlinear behaviours, which can be addressed using advanced techniques such as data mining, artificial intelligence (AI), ML, and DL. ML methods can reveal hidden patterns in historical rainfall data and have been proposed as an alternative modeling approach for nonlinear and dynamic systems[5]. For this reason, in recent years, ML approaches have emerged as powerful successors to traditional data mining techniques in the domain of rainfall prediction, reflecting the growing recognition of ML methods' capabilities in tackling the intricate challenges of predicting precipitation patterns[13–15,16]. demonstrated that ML methods are superior to traditional deterministic methods for rainfall prediction. Examples of ML models include RF, SVR, and SVM. RF employs an ensemble of decision trees to make predictions, enhancing accuracy and handling complex data relationships[17]. SVR, a form of SVM adapted for regression tasks, effectively captures nonlinear patterns in data. SVM is a versatile model for classification and regression, creating optimal decision boundaries through support vectors[18].

DL, as a subset of ML, has also demonstrated significant potential in enhancing predictive capabilities by utilising sophisticated neural networks inspired by the interconnected neurons of the human brain[19]. DL techniques encompass various algorithms of ANNs. These networks consist of interconnected layers of nodes, or "neurons," each processing and transforming input data before passing it to the next layer. The depth and intricacy of these networks enable them to capture complex patterns and relationships within extensive datasets. This capability is particularly valuable in domains characterised by high-dimensional and nonlinear data, such as climate science and meteorology[20]. DL methods are closely related to traditional ML methods, albeit differing in architectural complexity and the hierarchy of feature extraction[21]. While both DL and traditional ML strive to identify patterns in data, DL models excel in autonomously learning data representations at multiple levels of abstraction. This implies that DL models can automatically discover intricate features within raw data without explicit feature engineering, which often demands domain expertise and can be time-consuming[22]. DL application in rainfall prediction involves training neural networks on historical climate data. These networks are designed to identify hidden correlations, nonlinear relationships, and temporal dependencies among crucial variables for accurate rainfall forecasts[23]. As the neural networks process and learn from these datasets, they refine their internal representations, progressively enabling them to make increasingly accurate predictions. DL methods effectively harness the interconnectedness of climatic factors in rainfall forecasting, considering diverse variables such as temperature, humidity, wind speed, and atmospheric pressure and recognising their combined impact on rainfall patterns[24]. This comprehensive approach provides an advantage over traditional methods that may struggle to capture the intricate interactions among these variables. Furthermore, DL's ability to process vast amounts of data aligns well with the requirements of meteorological forecasting, where historical climate records span many decades and comprise an array of variables. The capacity of DL models to identify subtle

trends, nonlinear dependencies, and intricate temporal patterns within these datasets can lead to more accurate and reliable rainfall predictions[25].

Examples of DL approaches include ANN, RNN, KNN, and GRU. ANN simulates interconnected neurons to capture complex relationships in data, enhancing learning capabilities. RNN specialises in sequence modeling, preserving the memory of previous inputs for tasks like language processing[26]. KNN is a learning algorithm that makes predictions based on the proximity of data points in the feature space. GRU, a variant of RNN, addresses vanishing gradient issues in RNN and improves long-range dependency capture[27]. Additionally, there are time series models such as ARIMA, LSTM, trigonometric, Box-Cox transform, ARMA, and TBATS models. These models allow data analysts to model and forecast time series data across various applications. For instance, ARIMA combines autoregressive and moving average components with differencing for handling non-stationary data. LSTM is an RNN that captures intricate temporal relationships in sequences[28,29]. Trigonometric models leverage sinusoidal functions to capture cyclic patterns, which are ideal for data with periodic fluctuations. The Box-Cox transform stabilises variance and enhances normality in data[30]. ARMA models blend autoregressive and moving average components, while TBATS considers complex trends and seasonal patterns[17]. Using linear regression as an ML approach aims to predict rainfall by establishing relationships with other atmospheric variables[31,3233]. conducted a thorough comparative study to assess the effectiveness of statistical modeling and regression techniques in predicting rainfall using environmental features. The study highlighted the superior performance of regression techniques over statistical modeling when forecasting rainfall patterns. Additionally, their investigation demonstrated that the RF model exhibited enhanced predictive accuracy among ML algorithms compared to SVM and Decision Tree methods. This study contributes valuable insights into rainfall prediction, shedding light on the advantages of specific modeling approaches. The DL LSTM model has proven to be effective for rainfall prediction. In a study on forecasting rain in the Hyderabad region of India[20,34], put forth an enhanced LSTM model, which they compared with other models such as Holt-Winters, ARIMA, extreme learning machine (ELM), and RNN. The results were verified using ANN to predict the monthly average rainfall[35]. The findings indicate that out of the three different types of networks (layer recurrent, cascaded feed forward back propagation, and feed forward back propagation), the feed forward back propagation network type yielded the best results.

Previous studies have utilised a linear regression model to pinpoint crucial characteristics for predicting rainfall, including solar radiation, detectable water vapour, and daily patterns[36]. found that temperature, wind, and cyclones can be utilised to forecast rain in growth of agriculture sector and the farmers can take their decisions accordingly. Some researchers have also utilised atmospheric features such as temperature, relative humidity, pressure, and wind speed to accurately predict rainfall using ML techniques such as ANN, RF, and multiple linear regression models[37,38]. Therefore, the combination of ML, DL, and time series modeling represents a promising frontier for advancing our understanding of climatic patterns and improving our ability to predict rainfall.

This study focuses on exploring rainfall data in the North-Western Himalayas with several key objectives. Firstly, we aim to gauge the efficacy of advanced ML algorithms in predicting rainfall patterns. Secondly, we seek to evaluate the accuracy of DL models specifically tailored for this geographical region. Thirdly, we intend to compare the performance of ML and DL methods against traditional time series techniques commonly used in meteorological predictions. Additionally, our study aims to analyze how variations in altitude affect the precision of these predictive models. Lastly, based on our findings, we aim to propose a holistic approach to enhance the overall accuracy of rainfall prediction in the region. Through these objectives, we aim to contribute valuable insights that could advance both scientific understanding and practical applications in meteorological forecasting in mountainous terrains. This study takes a multidimensional approach to unravel the complexities of precipitation trends in a region known for its intricate altitude-dependent climate variations. By integrating these innovative methodologies, we aim to decode the intricate relationships between environmental factors and rainfall, ultimately developing more accurate and localised predictive models.

Materials and methods

Case study region, meteorological data and pre-processing. The study area covers the North-Western Himalayan region of India, centered at approximately 33° 5′ 24′′ N latitude and 74° 47′ 24′′ E longitude. Figure 1 illustrates the location of the study area covered. The study map was generated by QGIS 3.30.0. The six locations of meteorological stations were located on map across the altitudinal gradient. The primary data was gathered from the head office of the Indian Meteorological Department (IMD). The dataset encompassed daily data for maximum and minimum temperature and rainfall spanning 40 years (1980–2020). The initial data pre-processing stage involved several tasks: data conversion, handling missing values, encoding categorical variables, and partitioning the dataset into training and testing sets. Since the raw data exhibited both missing values and extreme values or outliers, corrective actions were taken. Missing values were replaced by not available (N.A). Subsequently, the dataset underwent encoding procedures before being ready for experimentation. Relevant attributes crucial for rainfall prediction were identified and isolated. The dataset was then divided, with 80% designated for training and 20% for testing, serving as input for the models.

Models

This study centres on rainfall prediction through a combined approach involving ML, DL, and time series methodologies. The analysis encompassed three distinct ML algorithms: ANN, SVR, and RF; three DL algorithms: RNN, LSTM, and GRU; and two time series algorithms: ARIMA and TBATS. These various algorithms utilised input variables moderately to strongly associated with rainfall, drawn from multiple environmental factors. The model parameters and information about models is given in supplementary Table 1S. The study determined and
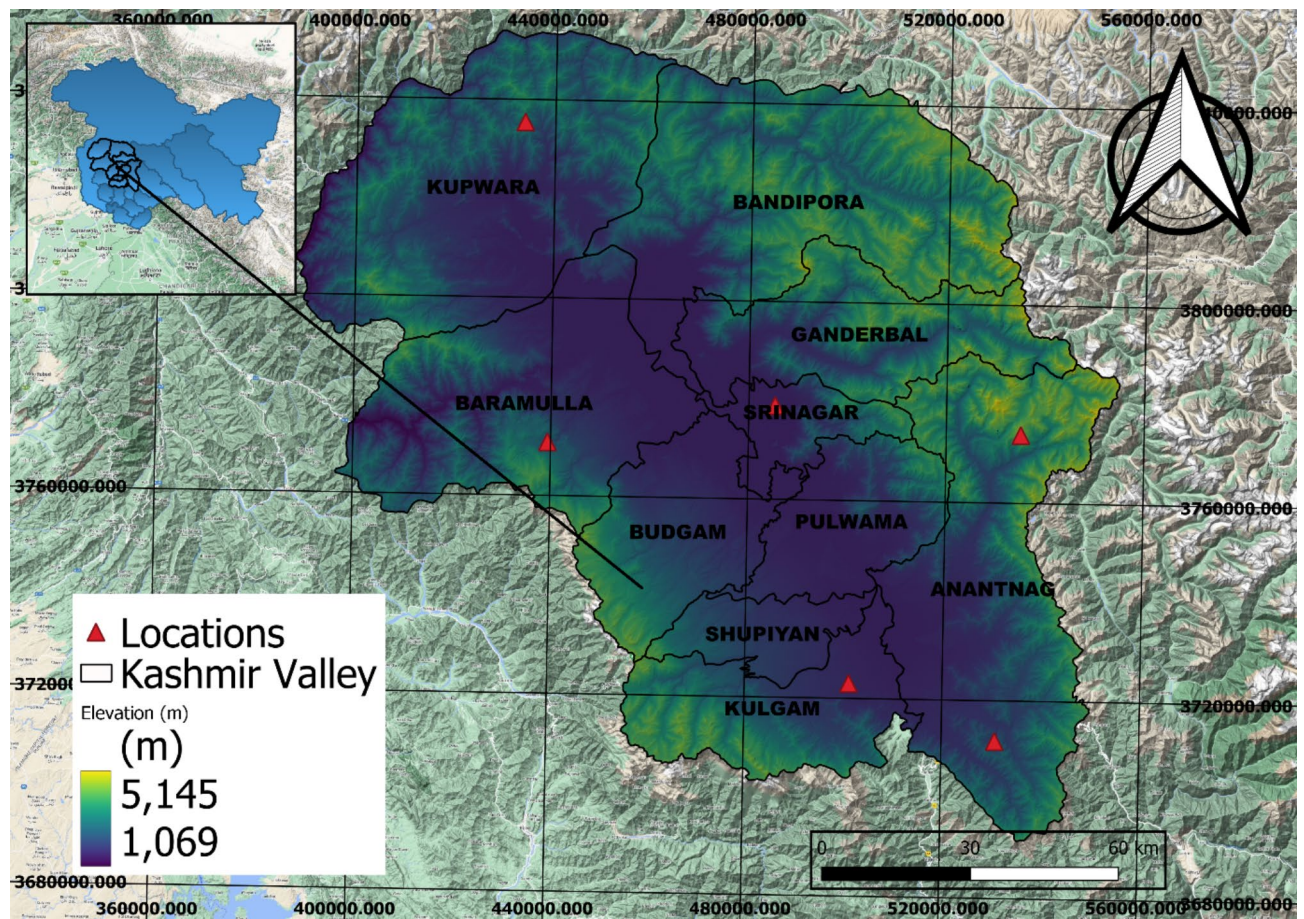
Fig. 1. Study area map (map was created using QGIS 3.30.0 https://www.qgis.org/).

reported the most effective models and algorithms by assessing performance using the RMSE and MAE metrics, other accuracy assement like Bias and R2 of various models across the altitudes is given in Table 1S.

Machine learning (ML) algorithms

Three ML algorithms were utilized to predict the rainfall based on the behaviour of the data set. The algorithms are discussed below:

Artificial neural network (ANN)

ANN is an ML model inspired by the brain's structure and function. The architecture of an ANN typically consists of an input layer, one or more hidden layers, and an output layer. The input layer receives and passes the input data to the hidden layers. The hidden layers use weights and biases to transform the input data into a new representation that is more suitable for the task. The output layer produces the final predictions based on the information processed by the hidden layers. The weights and biases are learned during training, allowing the model to make increasingly accurate predictions. The number of hidden layers, the size of the hidden layers, and the activation functions used determine the capacity of the model and its ability to learn complex relationships between inputs and outputs. In Fig. 2, y_t is considered as a function of its lag values $y_{t-1}, y_{t-2}, \ldots, y_{t-k}$ and respective weights.

$$y_t = a_0 + \sum_{j=1}^{q} a_j g(\beta_{0j} + \sum_{i=1}^{k} \beta_{ij} y_{t-i}) + \epsilon_t \tag{1}$$

where $a_j \, (j = 0,1,2,\ldots,q)$ and $\beta_{ij} \, (j = 0,1,2,\ldots,q; i = 1,2,.,k)$ are the model's connection weights; k is the number of nodes in the input layer, and q is the number of nodes in the hidden layer.

Support vector regression (SVR)

SVR, developed by Vapnik et al.[39], is a type of ML model used for regression and time series problems. The SVR model architecture consists of the following components. First, the input features are transformed into a high-dimensional space using a kernel function, such as a radial basis function (RBF) or a polynomial function. The transformed features are used to train a model. The model finds the optimal boundary by maximising the
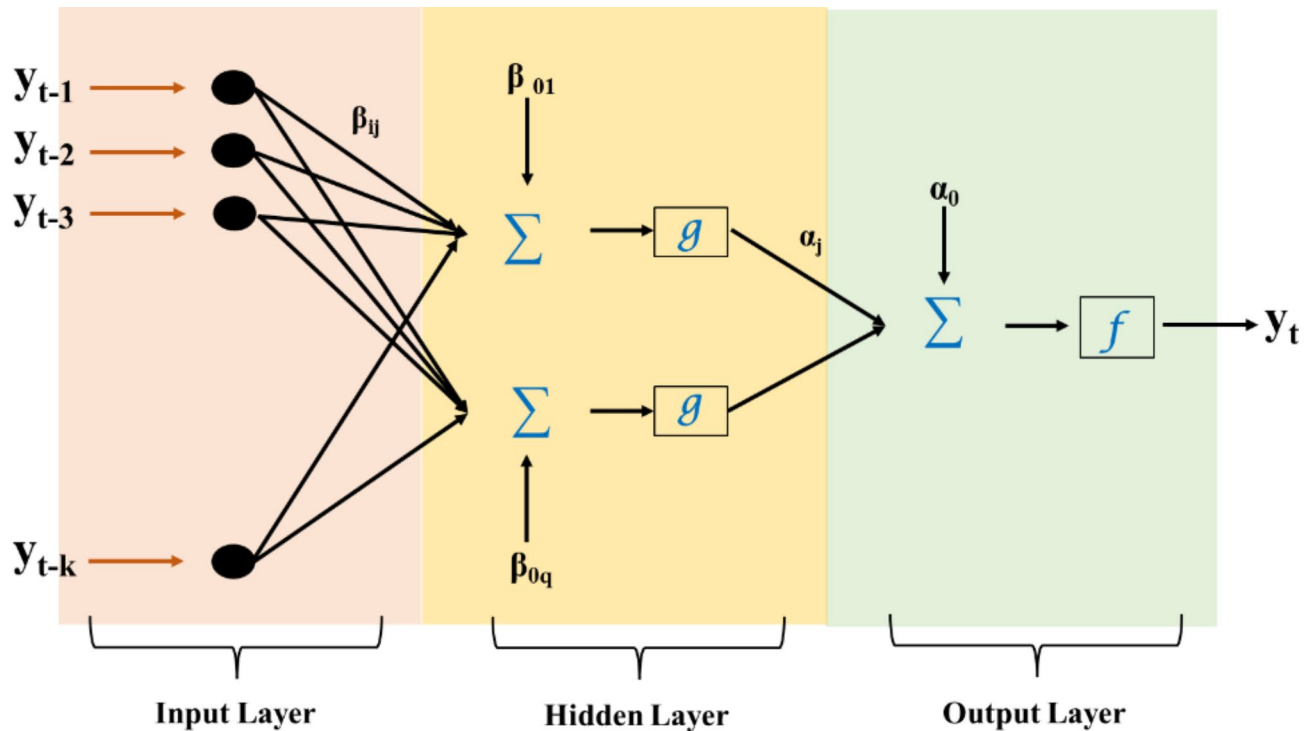
Fig. 2. Model architecture of ANN.

margin between the data points and the boundary. The model parameters are estimated using a cost function penalising deviations from the target values. The model makes predictions for new input data by transforming the features into high-dimensional space and applying the linear regression model. The kernel function and the regularisation parameter (C) are hyperparameters that can be tuned to improve the model's performance. The main advantage of the SVR model is that it can handle complex non-linear relationships between inputs and outputs using a suitable kernel function.

Random forest (RF)

RF is an ML algorithm based on decision trees. It builds multiple decision trees and combines their predictions to improve the overall accuracy and stability of the model. Many decision trees are trained on randomly selected subsets of the training data. Each decision tree is trained to make predictions for the target variable by recursively splitting the data into subsets based on the values of the input features. The final prediction for a new input sample is the average of the predictions made by all the decision trees. The random selection of features at each split helps to decorrelate the trees and reduce overfitting. The bootstrapped training data samples used to train each tree help reduce the prediction variance. The architecture of the RF model is given in Fig. 3.

Deep learning (DL) algorithms

To estimate daily rainfall through current environmental data, three specific algorithms viz., RNN, LSTM, and GRU were selected for the empirical investigation. Consequently, these three DL algorithms were tested and juxtaposed to determine the superior methods for anticipating daily rainfall quantities.

Recurrent neural network (RNN)

An RNN is a DL model that processes sequential data. It has a feedback loop that allows information to be passed from one sequence step to the next. The core architecture of an RNN consists of a repeating module with a hidden state, which stores information about the previous time step, and an input-to-hidden layer, which takes the current input and last hidden state as inputs and outputs the current hidden state. The hidden state is then used as input to a hidden-to-output layer, which generates the output for the present time step. The weights of the input-to-hidden and hidden-to-output layers are shared across all time steps, allowing the model to learn patterns in the sequence data. The internal structure of an RNN model can be described using the following equations:

Input Layer: The input layer receives the input sequence denoted as x(t), with t representing the time step.

Hidden Layer: The hidden layer contains the recurrent neuron that computes the hidden state h(t) using the following equation:

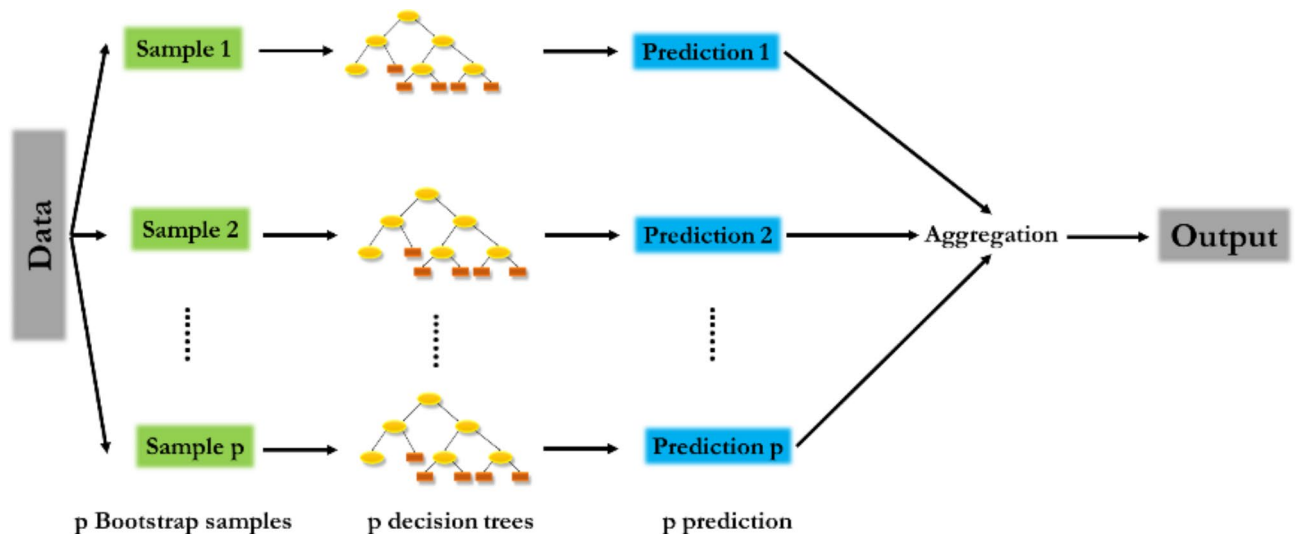$$h(t) = f(W_{hx}*x(t) + W_{hh} \cdot h(t-1) + b_h) \tag{2}$$

Fig. 3. Architecture of RF.

where, f is the activation function (e.g., tanh), W_{hx} is the weight matrix for the input-to-hidden layer connection, W_{hh} is the weight matrix for the hidden-to-hidden layer connection, and b_h is the bias vector.

Output Layer: The output layer computes the output y(t) using the following equation:

$$y(t) = W_{hy} * h(t) + b_y \tag{3}$$

where, W_{hy} is the weight matrix for the hidden-to-output layer connection, and b_y is the bias vector.

These equations describe the basic structure of a simple RNN model. More complex RNN models may have additional layers or use different types of activation functions or recurrent neurons. This architecture can be extended with multiple hidden layers to form a deep RNN or multiple hidden states to create an LSTM or GRU.

Long short-term memory (LSTM) and its variant

LSTM is an RNN type widely used for processing sequential data such as time series, speech signals, and text. It was introduced by[40] as a solution to the vanishing gradient problem that occurs in traditional RNNs. The LSTM model is an RNN type that can remember previous inputs for a long time. The architecture of an LSTM network consists of memory cells, input gates, forget gates and output gates. The memory cells are responsible for storing information for a prolonged period, whereas the gates control the flow of information in and out of the memory cells. The input gate controls the flow of information into the memory cell, allowing the network to decide what information to retain. The forget gate decides what information should be discarded based on the current input and previous memory state. The output gate determines what information should be outputted based on the current input and previous memory state. The structure of an LSTM unit can be described in the following Fig. 4 and equations:

Input gate: Controls the amount of new information allowed into the cell state. It is calculated using the following equation:

$$i_t = \sigma \{W_i * (h_{t-1} * x_t)\} + b_i \tag{4}$$

where, i_t is the input gate at time step t, W_i is the weight matrix, h_{t-1} is the concatenation of the previously hidden state $t-1$ and the current input x_t, and b_i is the bias term. The sigmoid activation function σ ensures that the input gate is between 0 and 1.

Forget gate: Controls the amount of information that will be forgotten from the cell state. It is calculated using the following equation:

$$f_t = \sigma \{W_f * (h_{t-1} * x_t)\} + b_f \tag{5}$$

where, f_t is the forget gate at time step t, W_f is the weight matrix, and b_f is the bias term.

Cell state: A continuous memory of the model that stores information from the past. It is updated using the following equation:
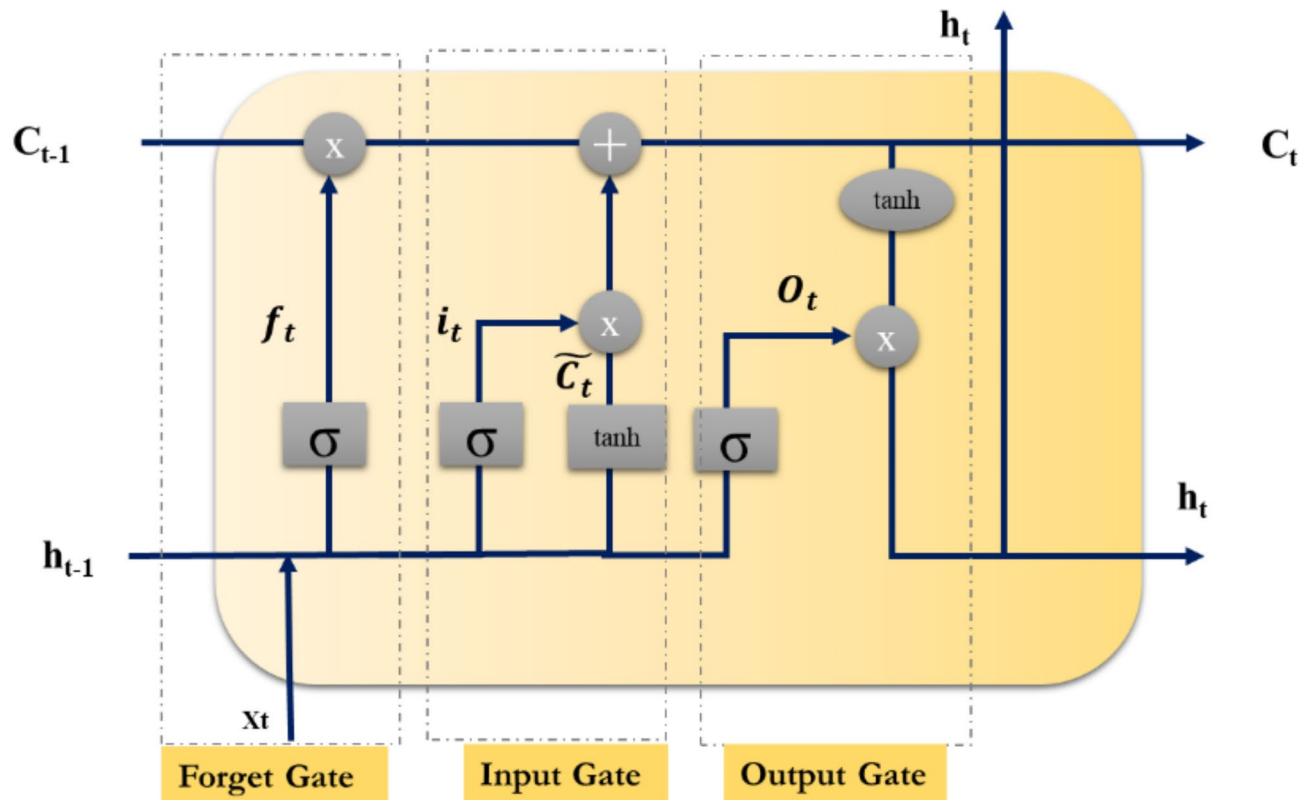
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

Fig. 4. Structure of an LSTM unit.

where, C_t is the cell state at time step t, f_t is the forget gate, i_t is the input gate, C_{t-1} is the previous cell state, and $\widetilde{C_t}$ is a candidate cell state calculated as follows:

$$\widetilde{C_t} = tanh\,\{W_c * (h_{t-1} * x_t)\} + b_c \tag{7}$$

W_c and b_c are the weight and bias of the candidate's input.

Output gate: Controls the amount of information output from the cell state to the hidden state. It is calculated using the following equation:

$$o_t = \sigma\,\{W_o * (h_{t-1} * x_t)\} + b_o \tag{8}$$

where, o_t is the output gate at time step t, W_o is the weight matrix, and

b_o is the bias term.

Hidden state: A representation of the current state of the model. It is calculated using the following equation:

$$h_t = o_t * tanh\,(C_t) \tag{9}$$

where, h_t is the hidden state at time step t, o_t is the output gate, and C_t is the cell state.

The above equations form the core of an LSTM unit, and multiple units can be stacked together to create a multi-layer LSTM network. A deep LSTM model is a variant of the LSTM model with multiple layers of LSTM units. A bidirectional LSTM processes the input sequences in two ways: in the forward direction (from start to end) and backward order (from end to start).

Gated recurrent unit (GRU)

GRU is a modification of the RNN model with two gates: the reset gate and the update gate. These gates help regulate the flow of information and prevent the vanishing gradient problem that can occur in traditional RNNs. The architecture of a GRU is like an LSTM network but with fewer parameters.

Time series algorithms

ARIMA and TBATS were chosen to forecast daily rainfall intensity using real-time environmental data. These algorithms were tested and compared to determine the best approach for accurate daily rainfall predictions.

Autoregressive integrated moving average (ARIMA)

ARIMA is a popular time series forecasting algorithm to model and predict univariate data. For the ARIMA model, the time series data must be stationary, meaning its mean, variance, and covariance are constant over time. If the data is not stationary, a differencing step is applied to make it stationary. The ARIMA model is selected based on the values of p, d, and q, where p is the number of AR terms, d is the number of differencing steps applied, and q is the number of MA terms. The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to identify the patterns in the time series data and to determine the number of autoregressive (AR) and moving average (MA) terms to include in the ARIMA model. The model is fitted to the time series data using maximum likelihood estimation or a variation thereof. The model is validated using residual analysis and by comparing the predicted values with the actual values for a portion of the time series data set aside for validation. The ARIMA model predicts the future values of the time series data.

Exponential smoothing state space model with Box–Cox transformation, ARMA errors, trend and seasonality components (TBATS)

TBATS is a time series forecasting algorithm for univariate time series data. It is a hybrid algorithm that combines exponential smoothing and ARIMA models, making it suitable for modeling complex time series patterns like seasonality, trend, and irregularity. The algorithm models the time series as a combination of multiple components, including a trend, a seasonal component, and an irregular component. The algorithm estimates the model's parameters using maximum likelihood estimation and predicts future values by combining the different elements. TBATS has been found to perform well on various time series datasets, including those with multiple seasonal patterns, non-stationary trends, and irregular fluctuations.

Statistical evaluation

Generally, it is necessary to assess the performance of a developed prediction model and compare it with other models using specific statistical measures. However, it is crucial to employ multiple statistical indices because different models may yield similar or nearly identical values for a particular index. This similarity makes it challenging to definitively determine which model performs better than the others. Each statistical index evaluates the model's performance from a single perspective of how well its outputs match the desired values. Therefore, evaluating models across multiple statistical indices is advisable to comprehensively assess each model's performance and conduct a robust comparative analysis, ultimately identifying the most suitable modeling approach. Two performance metrics were commonly used to evaluate the effectiveness of predictive models: the RMSE, and MAE. RMSE measures the average magnitude of the errors between predicted and actual values. Lower RMSE values indicate better model accuracy, as they signify smaller discrepancies between predicted and observed data points. Similarly, MAE quantifies the average absolute difference between predicted and actual values, with lower values indicating more accurate predictions. The RMSE and MAE can be defined as

$$\text{RMSE} = \frac{1}{n}\sqrt{\sum\nolimits_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2}$$

$$\text{MAE} = \frac{1}{n}\sum\nolimits_{t=1}^{n}\left|Y_t - \widehat{Y}_t\right|$$

where Y_t is the actual value, $\widehat{Y}_t$ is the fitted value and n is the number of observations.

Result and discussion

The descriptive statistics of temperature across altitudinal gradients in the North-Western Himalayas reveal significant variability, which is crucial for understanding local climate dynamics and its implications for rainfall patterns. Mean maximum temperatures ranged from 11.65 to 20.21 ℃ across locations L1 to L6, reflecting diverse thermal regimes influenced by altitude (Table 1). These findings are consistent with previous studies emphasizing the role of elevation in temperature variation[41]. The wide range of maximum temperature values, from −5.70 to 37.60 ℃, reveals the broad climatic spectrum within the study area, indicative of both alpine and lower altitude climates. Similarly, mean minimum temperatures ranged from 2.50 to 7.64 ℃ across the same locations, highlighting cooler conditions at higher elevations compared to lower lying areas (Table 2). This gradient is critical for understanding temperature-dependent processes such as evapotranspiration and precipitation formation[42]. Standard error and skewness values for both maximum and minimum temperatures further elucidate the distribution characteristics, indicating slight asymmetry in temperature data distributions but with relatively low variability, suggesting robustness in measurements[34,43,44]. These temperature statistics serve as a foundational framework for predicting rainfall patterns in the region. Studies have shown that temperature gradients directly influence atmospheric stability, moisture content, and the onset of precipitation events[45–48]. By integrating altitude-specific temperature data into rainfall models, researchers can enhance the accuracy of predictive models tailored to mountainous regions[49].

The analysis of rainfall patterns across different altitudinal locations (L1 to L6) in the North-Western Himalayas provided valuable insights into the spatial variability and predictive challenges associated with precipitation modeling in mountainous regions. Mean rainfall values ranging from 1.94 to 4.09 mm illustrate the significant diversity in precipitation levels across relatively small geographical distances (Table 3). This variability is further underscored by the wide range of minimum and maximum rainfall values observed at each location, reflecting the complex interplay of topography, atmospheric dynamics, and local climatic conditions. The standard error and skewness values offer additional perspectives on the distributional characteristics of rainfall data. Higher

Descriptive statistics	Maximum temperature					
	L1	L2	L3	L4	L5	L6
Mean	18.62	16.60	20.02	19.31	11.65	20.21
Standard error	0.07	0.07	0.07	0.07	0.07	0.08
Median	19.96	17.90	21.10	20.70	12.22	21.60
Mode	28.00	25.00	29.00	27.50	21.60	30.50
Standard deviation	8.76	8.35	8.92	8.44	7.96	9.30
Kurtosis	− 1.11	− 1.14	− 1.07	− 1.00	− 1.19	− 1.13
Skewness	− 0.29	− 0.29	− 0.28	− 0.37	− 0.15	− 0.28
Range	41.00	40.20	43.60	40.80	38.70	43.30
Minimum	− 5.70	− 8.00	− 6.60	− 5.10	− 9.60	− 5.70
Maximum	35.30	32.20	37.00	35.70	29.10	37.60

Table 1. Descriptive statistics of maximum temperature at different altitudinal gradients in the North-Western Himalayas.

Descriptive statistics	Minimum temperature					
	L1	L2	L3	L4	L5	L6
Mean	6.48	3.04	7.64	6.44	2.50	6.39
Standard error	0.06	0.06	0.06	0.06	0.06	0.06
Median	6.40	3.00	7.40	6.10	2.80	6.20
Mode	0.40	5.00	0.00	0.00	12.40	0.40
Standard deviation	6.95	7.09	7.51	7.12	7.20	7.19
Kurtosis	− 0.98	− 0.57	− 1.11	− 0.94	− 1.08	− 1.05
Skewness	0.01	− 0.07	0.09	0.08	− 0.11	0.12
Range	38.60	64.60	37.00	36.40	37.90	39.10
Minimum	− 15.70	− 18.60	− 11.80	− 13.60	− 19.80	− 15.70
Maximum	22.90	46.00	25.20	22.80	18.10	23.40

Table 2. Descriptive statistics of minimum temperature at different altitudinal gradients in the North-Western Himalayas.

Descriptive statistics	Rainfall					
	L1	L2	L3	L4	L5	L6
Mean	2.98	3.51	1.94	3.32	4.09	2.97
Standard Error	0.07	0.07	0.05	0.09	0.08	0.06
Median	0.00	0.00	0.00	0.00	0.00	0.00
Mode	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation	8.26	8.70	6.30	10.37	10.13	7.65
Kurtosis	44.80	33.66	56.41	53.54	49.59	21.75
Skewness	5.46	4.77	6.18	6.02	5.61	4.14
Range	149.50	138.90	130.30	206.00	189.20	85.00
Minimum	0.00	-5.50	0.00	0.00	-0.20	0.00
Maximum	149.50	133.40	130.30	206.00	189.00	85.00

Table 3. Descriptive statistics of rainfall at different altitudinal gradients in the North-Western Himalayas.

skewness values, such as those observed for locations L2, L3, L4, and L5, indicate asymmetric distributions with a tendency towards higher or lower rainfall extremes, potentially influenced by localized weather phenomena like orographic lifting or convective processes[44,50]. Conversely, lower skewness values suggest a more symmetrical distribution of rainfall, indicative of more moderate and consistent precipitation patterns[51–54]. The application of ML models (RF, SVR, ANN, and KNN) to predict rainfall from meteorological variables reveals varying model performances across different altitudinal gradients. RF and SVR models generally exhibited higher RMSE values during both training and testing phases, suggesting challenges in capturing the complex nonlinear relationships between predictors and rainfall outcomes in mountainous terrain. These findings align with previous studies emphasizing the sensitivity of statistical models to spatial heterogeneity and the need for robust

validation strategies in mountainous regions[55–57]. In contrast, ANN and KNN models demonstrated relatively lower RMSE values, indicating their potential for better capturing spatial variability and nonlinearity in rainfall patterns across diverse altitudes. The superior performance of these models may stem from their ability to learn complex patterns and relationships inherent in meteorological data, including altitude-dependent factors such as orographic effects and microclimatic variations[58,59]. These findings highlight the importance of integrating altitude-specific meteorological data and employing advanced modeling techniques to enhance the accuracy of rainfall forecasts in mountainous regions. Future research directions could include refining model inputs by incorporating additional environmental variables (e.g., terrain characteristics, vegetation cover) and exploring ensemble modeling approaches to mitigate uncertainties associated with individual model performances[60–64]. Furthermore, incorporating high-resolution satellite data and ground-based observations could further improve the spatial representation of rainfall patterns and support more robust model validations in complex terrain settings.

The application of ML, DL, and time series modeling techniques to predict rainfall from meteorological variables across altitudinal gradients in the North-western Himalayas offers valuable insights into model performance and predictive accuracy (Table 4). ML models such as RF, SVR, ANN, and KNN exhibited varying levels of effectiveness as indicated by their train and test RMSE values across altitudinal levels (L1 to L6) (Fig. 5). These differences underscored the sensitivity of these models to spatial variations in climatic conditions and the complex interactions between meteorological predictors and rainfall patterns. DL models, including LSTM, bi-directional LSTM, deep LSTM, GRU, and RNN, also demonstrated varying RMSE values across altitudinal gradients. These models leverage sequential dependencies in data and have shown promise in capturing temporal patterns and non-linear relationships in climatic variables, which are crucial for accurate rainfall prediction in dynamic mountainous environments[40,65]. Moreover, time series modeling approaches such as ARIMA and TBATS provided insights into the temporal variability of rainfall across altitudes. These models exhibited varying train and test RMSE values, reflecting their ability to capture both short-term fluctuations and long-term trends in rainfall data[56,66–68]. The observed differences in RMSE values across these modeling techniques highlight the importance of selecting appropriate methodologies that account for the complex spatial and temporal dynamics inherent in mountainous regions. The higher RMSE values in some models indicate challenges in accurately capturing local-scale variations and extreme weather events, which are critical for effective water resource management and disaster preparedness in mountain ecosystems[69–71]. Integrating diverse modeling approaches and leveraging advanced statistical techniques enhance our understanding of rainfall variability in mountainous areas.

The DL models, including LSTM, RNN, Bidirectional LSTM, deep LSTM, and GRU, exhibited varying degrees of effectiveness across the study locations. LSTM consistently demonstrated robust performance, achieving lower RMSE and MAE compared to other DL models in several instances[30]. For example, at Location 1, LSTM achieved a training RMSE of 26.4051 and a testing RMSE of 38.2848, indicating its proficiency in capturing the complex relationships between temperature variables and rainfall patterns. Bidirectional LSTM also showed competitive performance, particularly noteworthy for its lower MAE scores across multiple locations. However, it displayed marginally higher RMSE values compared to LSTM, suggesting potential variability in predictive accuracy across different evaluation metrics. On the other hand, Deep LSTM and GRU, while generally performing adequately, exhibited higher RMSE and MAE values compared to LSTM and Bidirectional LSTM (Fig. 6). This observation suggests that these models might have struggled more with capturing the intricate nuances of temperature-rainfall relationships specific to the north-western Himalayan region. In contrast, traditional ML models such as ANN, KNN, SVR, RF, and the time series approach ARIMA consistently demonstrated higher RMSE and MAE values across all locations[72]. Specifically, KNN, SVR, and RF exhibited notably higher error metrics, highlighting their limitations in accurately capturing the non-linear dependencies inherent in rainfall prediction tasks compared to DL models.

The DL models consistently outperformed both ML and time series models across all six locations examined (Fig. 7). Specifically, LSTM and Bidirectional LSTM emerged as the top-performing models, achieving the lowest RMSE and MAE scores for locations L1 through L6, respectively. Meanwhile, RNN, LSTM, Bidirectional LSTM, and Deep LSTM were identified as superior models based on their performance in terms of test MAE across the same locations. Overall, the DL algorithms demonstrated superior accuracy, with Bi-directional LSTM showing the highest effectiveness, followed by LSTM, RNN, Deep LSTM, and GRU in descending order. In contrast, the ML models performed relatively better than traditional time series methods, with ANN leading, followed by KNN, SVR, and RF. Time series models, represented by TBATS and ARIMA, ranked lowest in accuracy (Fig. 8). The preference for DL models in predicting rainfall patterns in the North-Western Himalayas can be attributed to several factors. Firstly, the complex and non-linear nature of the data, influenced by altitude variations, topography, and atmospheric stability, poses challenges for ML and time series models in accurately capturing these relationships[73]. DL models, such as LSTM and Bidirectional LSTM, are specifically designed to handle such complexities by automatically learning intricate data patterns[27,72]. This capability reduces the need for manual feature engineering and enhances prediction accuracy by effectively managing noise inherent in rainfall and temperature data. DL models benefit from the availability of large historical datasets in the region, which are essential for training these models effectively. The extensive data enable DL models to generalize well and remain robust against temporal variations in the data, thereby improving overall prediction accuracy. This study makes a significant contribution to overcoming challenges in agricultural planning and climate adaptation by applying advanced Machine Learning (ML) and Deep Learning (DL) techniques to rainfall prediction. As traditional forecasting methods struggle with the increasing unpredictability of climate, this research enhances forecast accuracy through sophisticated algorithms, allowing for more informed decisions by farmers and agricultural planners. By integrating complex climatic data and assessing model effectiveness in the North-Western Himalayas, the study provides crucial insights into improving predictive precision and managing varied

Models	L1				L2				L3			
	Train RMSE	Test RMSE	Train MAE	Test MAE	Train RMSE	Test RMSE	Train MAE	Test MAE	Train RMSE	Test RMSE	Train MAE	Test MAE
LSTM	26.4051	38.2848	18.3032	22.272	28.6268	32.6531	18.6174	20.2858	20.9055	22.4477	14.0341	14.8548
RNN	26.4314	38.4852	17.9959	22.048	34.4668	36.1414	28.5762	28.7417	23.2679	24.2392	18.0615	18.1207
Bidirectional LSTM	28.5082	39.0882	20.9806	24.3876	28.4072	32.4625	18.7237	20.3433	20.738	22.2265	13.966	14.6357
Deep LSTM	29.2431	39.3847	22.2205	25.5191	30.2622	33.0527	22.1146	23.2161	22.7085	23.8022	17.0902	17.3248
GRU	29.5688	39.4416	22.7908	25.8663	31.0385	33.4672	23.8292	24.8423	22.0524	23.0137	14.7668	15.453
TBATS	30.4407	39.9763	23.7599	26.9549	33.7797	35.5372	26.5045	26.1142	23.6468	24.5349	17.885	17.2726
ANN	29.5231	40.6845	21.3692	23.5784	31.4014	37.3677	23.3202	27.4649	22.0548	24.0753	15.5666	15.8668
KNN	39.137	44.8777	26.0212	26.8164	41.1062	42.5591	29.3555	30.6486	29.0738	29.6286	19.4788	20.0675
SVR	37.0156	46.6131	25.3716	30.5223	40.2145	42.0613	29.3366	30.4928	28.7573	29.8899	19.6855	20.3037
RF	36.5499	48.5559	25.5494	31.5407	42.4912	43.1474	29.9277	29.323	30.6046	33.4445	19.9559	21.6242
ARIMA-X	33.3473	48.9682	24.8915	40.2202	33.7015	35.8058	27.2914	28.2475	23.2583	24.5984	17.7394	18.571

Models	L4				L5				L6			
	Train RMSE	Test RMSE	Train MAE	Test MAE	Train RMSE	Test RMSE	Train MAE	Test MAE	Train RMSE	Test RMSE	Train MAE	Test MAE
LSTM	37.6824	41.882	25.7388	23.2889	36.9767	32.2304	22.3422	21.5748	27.0726	27.1765	19.3609	20.128
RNN	34.9175	42.6456	21.6094	22.2904	36.6707	32.2461	22.3529	21.7823	27.7091	28.0215	19.5725	20.5426
Bidirectional LSTM	34.9437	42.6471	21.2045	21.8476	36.815	34.6105	25.4816	25.6676	26.5043	26.9585	18.4087	19.5685
Deep LSTM	36.629	44.0575	25.7735	26.5994	38.4883	35.9311	26.3701	26.5954	28.9267	29.2136	22.0771	22.8641
GRU	36.6263	44.1257	25.4547	26.1814	41.3881	35.9506	30.7579	27.3958	28.2329	28.6365	21.3745	22.0169
TBATS	39.4747	45.1084	30.271	29.8734	39.0508	37.0425	27.7653	28.1527	30.4776	30.9879	23.0755	24.0622
ANN	37.902	45.2663	28.3823	29.2586	39.0998	37.0643	27.8667	28.2929	26.633	30.0485	19.4238	20.9501
KNN	45.3611	51.0009	32.2122	38.5397	45.1007	40.61	31.9854	27.5973	37.01	37.7642	25.2801	27.1359
SVR	48.4088	52.6685	32.5023	33.006	47.412	44.517	32.5436	30.7088	36.3563	37.1531	25.3094	26.9675
RF	47.8864	53.3362	32.6218	32.8244	47.9481	45.5619	32.3674	31.2958	38.7174	37.8615	25.8642	25.9618
ARIMA-X	51.0779	58.2526	33.3209	35.5394	51.7974	53.9499	33.4704	34.7775	31.444	30.9085	24.5291	24.6846

Table 4. Performance metrics of predictive models at different altitudinal gradients in the North-Western Himalayas.
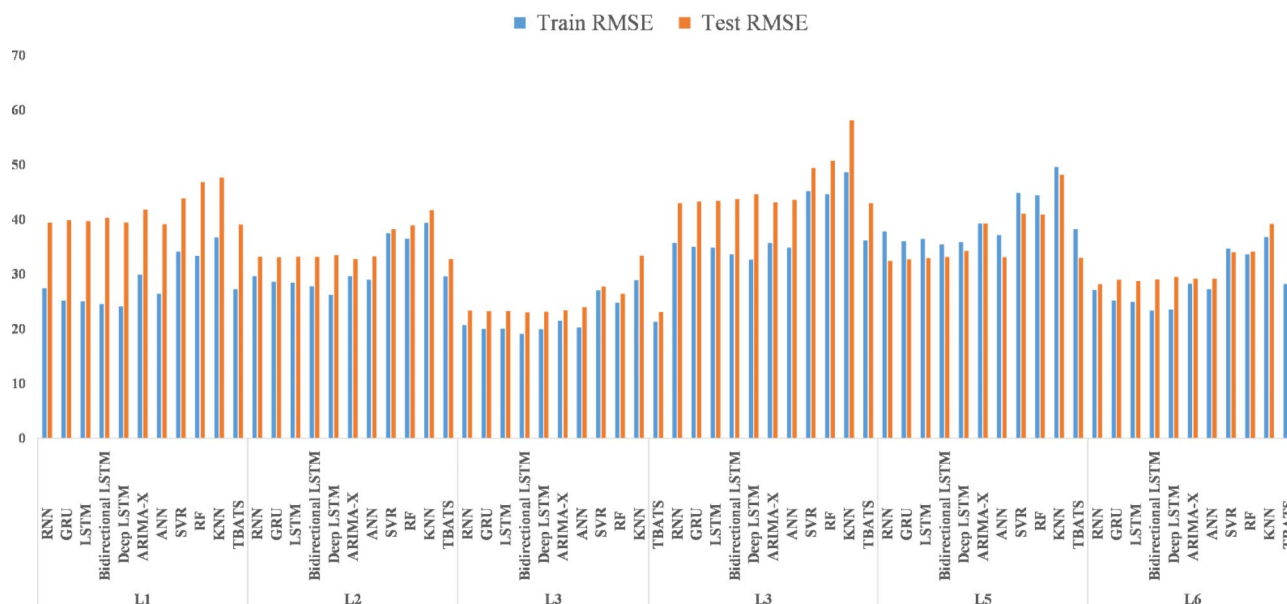
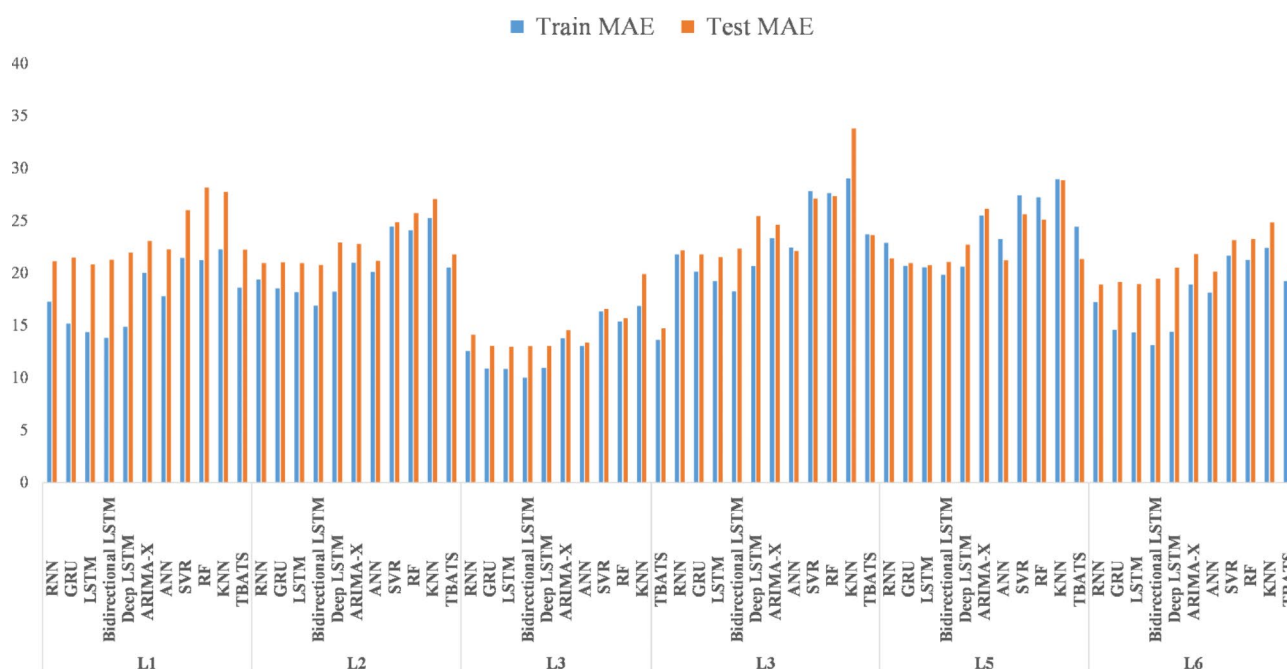**Fig. 5**. Bar plot of RMSE across altitudinal gradient.

Fig. 6. Bar plot of MAE across altitudinal gradient.

environmental conditions. This advancement supports better agricultural practices, infrastructure development, and overall food security in the face of climate variability.

The use of advanced Machine Learning (ML) and Deep Learning (DL) techniques for rainfall prediction, as outlined in this study, represents a significant advancement in meteorological forecasting with important implications for engineers and stakeholders, especially in semi-arid regions. Techniques such as Random Forest (RF), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) offer notable improvements in rainfall prediction accuracy. This enhanced precision is crucial for effective planning and management in areas dependent on seasonal rains. However, applying these techniques in practice brings several considerations. A key issue is balancing the accuracy of the models with their complexity. While these advanced methods can deliver highly accurate forecasts, they also introduce increased computational demands and complexity. Engineers and stakeholders must assess whether the benefits of improved accuracy justify the challenges associated with implementing and maintaining these sophisticated systems. The models

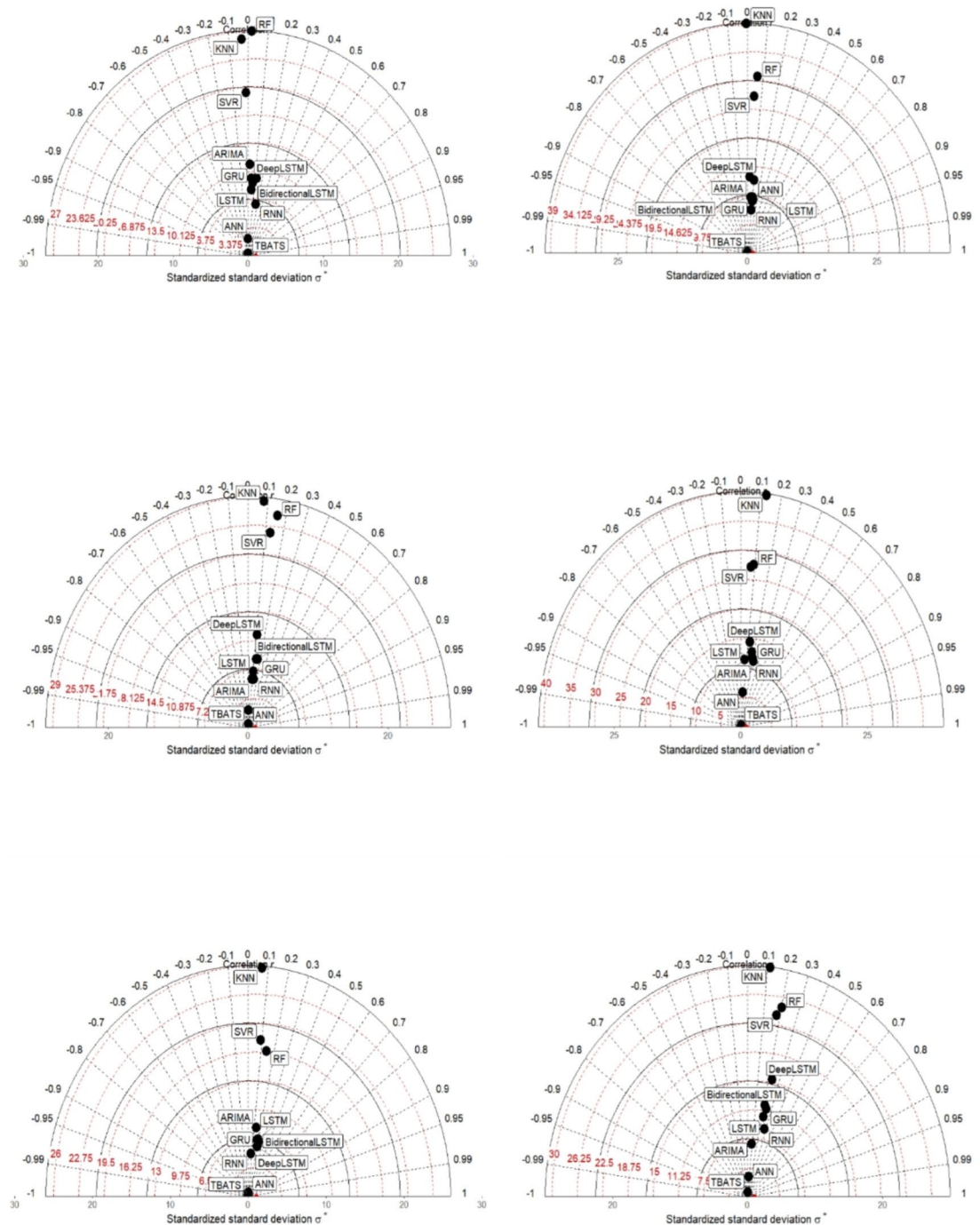**Fig. 7**. Distribution of weather variables across altitudinal gradient.

Fig. 8. Taylor diagram of altitude wise comparing accuracy of models.

require significant computational power and specialized expertise, which might be a hurdle for regions with limited technological resources. Proper integration of these advanced models into existing forecasting systems is necessary to ensure that the accuracy improvements justify the operational demands. Another important consideration is the dependency on extensive and high-quality datasets. These advanced models perform best with large amounts of data, which may not always be available, especially in regions with sparse meteorological information. This points to a need for better data collection infrastructure. Engineers should focus on enhancing

data acquisition systems and developing strategies to handle limited data to fully utilize these forecasting techniques.

The adaptability of these methodologies to other meteorological variables like humidity, wind speed, and solar radiation is also significant. While the principles behind these methods are versatile, each climatic variable has unique challenges that may require specialized approaches. Although the methods used for rainfall prediction can be extended to other variables, engineers need to tailor the models to each specific context to ensure accuracy and effectiveness. Operational integration is another practical challenge. Incorporating these advanced models into existing systems for agricultural planning, disaster management, and infrastructure development requires careful consideration. Stakeholders must address how to integrate these models into practical decision-making processes, which includes not only technical adaptation but also training and capacity building for users. Ensuring that personnel are properly trained to use these advanced tools is essential for maximizing their benefits. The benefits of improved rainfall forecasting are considerable. Enhanced prediction accuracy enables better agricultural planning, such as more precise irrigation scheduling and effective water resource management. It also supports infrastructure design and disaster preparedness by helping engineers create infrastructure that can withstand extreme weather events. For policymakers, advanced forecasting techniques provide a foundation for developing data-driven policies that promote sustainable resource management and enhance climate resilience. In summary, while advanced techniques for rainfall prediction offer considerable advantages, they also require careful management of model complexity, data needs, and operational integration. Addressing these challenges effectively can lead to significant improvements in decision-making, infrastructure planning, and policy development, ultimately supporting more resilient and sustainable practices in meteorology and related fields. The model parameters were presented in Table 2S.

Conclusions

In conclusion, the comprehensive analysis of temperature and rainfall patterns across altitudinal gradients in the North-Western Himalayas portrays the significant variability and complexity inherent in mountainous climatic systems. The study revealed diverse thermal regimes influenced by altitude, with mean maximum temperatures ranging from 11.65 to 20.21 ℃ and mean minimum temperatures from 2.50 to 7.64 ℃ across different locations. Altitude is a critical factor shaping temperature variations in the region. The wide range of temperature values reflects both alpine and lower altitude climates, crucial for understanding local climate dynamics, evapotranspiration processes, and precipitation formation. Statistical analyses, including standard error and skewness, further elucidate the distribution characteristics of temperature data, emphasizing robustness in measurements despite slight asymmetry. This foundational framework of temperature statistics not only enhances our understanding of regional climate but also serves as a crucial basis for predicting rainfall patterns. Integrating altitude-specific temperature data into predictive models improves the accuracy of rainfall forecasts by accounting for temperature gradients that influence atmospheric stability, moisture content, and precipitation onset. Advanced modeling techniques such as ML, DL, and time series analysis provided deeper insights into rainfall variability across diverse altitudinal gradients. DL models, particularly LSTM and Bidirectional LSTM, demonstrated superior performance in capturing complex climatic relationships compared to traditional ML and time series methods. Their ability to handle non-linear data dynamics and leverage extensive historical datasets underscores their effectiveness in predicting rainfall patterns in dynamic mountainous environments. Moving forward, continued research efforts should focus on refining model inputs by incorporating additional environmental variables and exploring ensemble modeling approaches to further enhance prediction accuracy. High-resolution satellite data and ground-based observations will play pivotal roles in improving spatial representation and validating models in complex terrain settings. By advancing our understanding and predictive capabilities, we can better manage water resources and mitigate risks associated with climate variability in mountain ecosystems. In real-world applications for rainfall monitoring and warning systems, standalone methods such as statistical, physical, and data-driven models offer distinct advantages but also face notable limitations. Statistical models, like ARIMA, are straightforward and computationally efficient, making them practical for immediate use. However, their tendency to assume linear relationships and their limited flexibility can lead to suboptimal performance when faced with sudden climatic changes or complex weather patterns. Physical models, which simulate atmospheric and hydrological processes, can provide in-depth forecasts by considering intricate variable interactions. Despite their accuracy in well-defined conditions, these models are often hindered by high computational demands and complexity, making them less feasible for real-time applications in regions with limited resources. Additionally, physical models may also encounter local optima if not precisely calibrated for specific regional conditions. Data-driven models, such as machine learning and deep learning techniques, are adept at identifying complex, non-linear patterns and can be adapted to various datasets. While they have the potential for high precision, their effectiveness depends on access to extensive and high-quality data and significant computational power. These models are also prone to overfitting, where they excel with historical data but may underperform with new or different data, and their intricate nature often results in lower interpretability. Overall, while advanced data-driven methods offer substantial improvements in rainfall forecasting, their practical implementation must carefully address data quality, computational requirements, and ongoing maintenance to ensure reliable and actionable predictions.

Data availability

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

References

1. Skendžić, S., Zovko, M., Živković, I. P., Lešić, V. & Lemić, D. The impact of climate change on agricultural insect pests. *Insects*. **12**, 440 (2021).
2. Zuma-Netshiukhwi, G., Stigter, K. & Walker, S. Use of traditional weather/climate knowledge by farmers in the South-Western Free State of South Africa: agrometeorological learning by scientists. *Atmosphere*. **4**, 383–410 (2013).
3. Jones, J. W., Hansen, J. W., Royce, F. S. & Messina, C. D. Potential benefits of climate forecasting to agriculture. *Agric. Ecosyst. Environ*. **82**, 169–184 (2000).
4. Chen, C. et al. Forecast of rainfall distribution based on fixed sliding window long short-term memory. *Eng. Appl. Comput. Fluid Mech*. **16**, 248–261 (2022).
5. Abobakr Yahya, A. S. et al. Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water*. **11**, 1231 (2019).
6. Grote, U. Can we improve global food security? A socio-economic and political perspective. *Food Secur*. **6**, 187–200 (2014).
7. Trenberth, K. E. & Asrar, G. R. Challenges and opportunities in water cycle research: WCRP contributions. *Earth's Hydrol. Cycle* **46**, 515–532 (2014).
8. Franzke, C. L., O'Kane, T. J., Berner, J., Williams, P. D. & Lucarini, V. Stochastic climate theory and modeling. *Wiley Interdiscip. Rev. Clim. Change*. **6**, 63–78 (2015).
9. Wang, H. et al. Association of meteorological factors with infectious diarrhea incidence in Guangzhou, southern China: a time-series study (2006–2017). *Sci. Total Environ*. **672**, 7–15 (2019).
10. Garbrecht, J. D., Nearing, M. A., Zhang, J. X. & Steiner, J. L. Uncertainty of climate change impacts on soil erosion from cropland in central Oklahoma. *Appl. Eng. Agric*. **32**, 823–836 (2016).
11. Resnicow, K. & Vaughan, R. A chaotic view of behavior change: a quantum leap for health promotion. *Int. J. Behav. Nutr. Phys. Activity*. **3**, 1–7 (2006).
12. Praveen, B. et al. Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. *Sci. Rep.* **10**, 10342 (2020).
13. Khambra, G. & Shukla, P. Novel machine learning applications on fly ash based concrete: an overview. *Mater. Today: Proc*. **80**, 3411–3417 (2023).
14. Alkesaiberi, A., Harrou, F. & Sun, Y. Efficient wind power prediction using machine learning methods: a comparative study. *Energies*. **15**, 2327 (2022).
15. Patil, S. S. & Vidyavathi, B. A machine learning approach to weather prediction in wireless sensor networks. *Int. J. Adv. Comput. Sci. Appl*. **13** (2022).
16. Namitha, K., Jayapriya, A. & Kumar, G. S. In *Proceedings of the Third International Symposium on Women in Computing and Informatics*. 492–495.
17. Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S. & Ahrentzen, S. Random forest based hourly building energy prediction. *Energy Build*. **171**, 11–25 (2018).
18. Levis, A. & Papageorgiou, L. Customer demand forecasting via support vector regression analysis. *Chem. Eng. Res. Des*. **83**, 1009–1018 (2005).
19. Calderaro, J., Seraphin, T. P., Luedde, T. & Simon, T. G. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J. Hepatol*. **76**, 1348–1361 (2022).
20. Indrakumari, R., Poongodi, T. & Singh, K. Introduction to deep learning. *Adv. Deep Learn. Eng. Sci. Pract. Approach*. **1**, 1–22 (2021).
21. Vieira, S., Pinaya, W. H. & Mechelli, A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci. Biobehav. Rev.*. **74**, 58–75 (2017).
22. Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. *Electron. Markets*. **31**, 685–695 (2021).
23. Ozcanli, A. K., Yaprakdal, F. & Baysal, M. Deep learning methods and applications for electrical power systems: a comprehensive review. *Int. J. Energy Res*. **44**, 7136–7157 (2020).
24. Kang, J. et al. Prediction of precipitation based on recurrent neural networks in Jingdezhen, Jiangxi Province, China. *Atmosphere*. **11**, 246 (2020).
25. Mayer, M. J. Benefits of physical and machine learning hybridization for photovoltaic power forecasting. *Renew. Sustain. Energy Rev*. **168**, 112772 (2022).
26. Almási, A. D., Woźniak, S., Cristea, V., Leblebici, Y. & Engbersen, T. Review of advances in neural networks: neural design technology stack. *Neurocomputing*. **174**, 31–41 (2016).
27. Saha, S., Baral, S. & Haque, A. DEK-Forecaster: A novel deep learning model integrated with EMD-KNN for traffic prediction. arXiv preprint arXiv:2306.03412 (2023).
28. Ban, W., Shen, L., Chen, J. & Yang, B. Short-term prediction of wave height based on a deep learning autoregressive integrated moving average model. *Earth Sci. Inf*. **16**, 2251–2259 (2023).
29. Wani, O. A. et al. Climate plays a dominant role over land management in governing soil carbon dynamics in North Western Himalayas. *J. Environ. Manage*. **338**, 117740 (2023).
30. Godahewa, R., Bergmeir, C., Webb, G. I. & Montero-Manso, P. An accurate and fully-automated ensemble model for weekly time series forecasting. *Int. J. Forecast*. **39**, 641–658 (2023).
31. Thirumalai, C., Harsha, K. S., Deepak, M. L. & Krishna, K. C. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)*. 1114–1117 (IEEE).
32. Prabakaran, S., Kumar, P. N. & Tarun, P. S. M. Rainfall prediction using modified linear regression. *ARPN J. Eng. Appl. Sci*. **12**, 3715–3718 (2017).
33. Tharun, V., Prakash, R. & Devi, S. R. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. 1507–1512 (IEEE).
34. Kumar, S. S., Wani, O. A., Krishna, J. R. & Hussain, N. Impact of climate change on soil health. *Int. J. Environ. Sci*. **7**, 70–90 (2022).
35. Abhishek, K., Kumar, A., Ranjan, R. & Kumar, S. In *2012 IEEE Control and System Graduate Research Colloquium*. 82–87 (IEEE).
36. Chaudhari, M. & Choudhari, D. Study of various rainfall estimation & prediction techniques using data mining. *Am. J. Eng. Res*. **6**, 137–139 (2017).
37. Vijayan, R., Mareeswari, V., Mohankumar, P., Gunasekaran, G. & Srikar, K. Estimating rainfall prediction using machine learning techniques on a dataset. *Int. J. Sci. Technol. Res*. **9**, 440–445 (2020).
38. Gnanasankaran, N. & Ramaraj, E. A multiple linear regression model to predict rainfall using Indian meteorological data. *Int. J. Adv. Sci. Technol*. **29**, 746–758 (2020).
39. Vapnik, V. & Chervonenkis, A. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognit. Image Anal*. **1**, 283–305 (1991).
40. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput*. **9**, 1735–1780 (1997).
41. Palazzi, E., Filippi, L. & von Hardenberg, J. Insights into elevation-dependent warming in the Tibetan Plateau-Himalayas from CMIP5 model simulations. *Clim. Dyn*. **48**, 3991–4008 (2017).

42. Patel, J. B. Analysis of the microbial diversity associated with the Lesotho highlands through culture-independent approaches (2020).
43. Dunn, R. J., Willett, K. M. & Parker, D. E. Changes in statistical distributions of sub-daily surface temperatures and wind speed. *Earth Sys. Dyn.* **10**, 765–788 (2019).
44. Nazir, S. F., Singh, L., Shah, B. A. & Ali, O. R. I. Rice-wheat cropping system under changing climate scenario: a review. *Int. J. Chem. Stud.* **8**, 1907–1914 (2020).
45. Gimeno, L. et al. Oceanic and terrestrial sources of continental precipitation. *Rev. Geophys.* **50** (2012).
46. Moharana, L., Sahoo, A. & Ghose, D. In *IOP Conference Series: Earth and Environmental Science.* 012054 (IOP Publishing).
47. Sahoo, A., Behera, S. & Sharma, N. in *AIP Conference Proceedings.* (AIP Publishing).
48. Sahoo, A. & Ghose, D. K. in *Smart Intelligent Computing and Applications, Volume 1: Proceedings of Fifth International Conference on Smart Computing and Informatics (SCI* 307–317 (Springer, 2021).
49. Körner, C. et al. Creative use of mountain biodiversity databases: the Kazbegi research agenda of GMBA-DIVERSITAS. *Mt. Res. Dev.* **27**, 276–281 (2007).
50. Hornberger, G. M., Wiberg, P. L., Raffensperger, J. P. & D'Odorico, P. *Elements of Physical Hydrology* (JHU, 2014).
51. Rashid, M. M., Beecham, S. & Chowdhury, R. K. Statistical characteristics of rainfall in the Onkaparinga catchment in South Australia. *J. Water Clim. Change.* **6**, 352–373 (2015).
52. Sahoo, A. & Ghose, D. K. Imputation of missing precipitation data using KNN, SOM, RF, and FNN. *Soft. Comput.* **26**, 5919–5936 (2022).
53. Sahoo, B. B., Jha, R., Singh, A. & Kumar, D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys.* **67**, 1471–1481 (2019).
54. Sahoo, B. B., Panigrahi, B., Nanda, T., Tiwari, M. K. & Sankalp, S. Multi-step ahead urban water demand forecasting using deep learning models. *SN Comput. Sci.* **4**, 752 (2023).
55. Dal Molin, M., Schirmer, M., Zappa, M. & Fenicia, F. Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: the case study of the Thur catchment. *Hydrol. Earth Syst. Sci.* **24**, 1319–1345 (2020).
56. Singh, G., Batra, N., Salaria, A., Wani, O. A. & Singh, J. Groundwater quality assessment in Kapurthala district of central plain zone of Punjab using hydrochemical characteristics. *J. Soil Water Conserv.* **20**, 43–51 (2021).
57. Babu, S. et al. Biochar implications in cleaner agricultural production and environmental sustainability. *Environ. Science: Adv.* **2**, 1042–1059 (2023).
58. Dura, V., Evin, G., Favre, A. C. & Penot, D. Spatial variability in the seasonal precipitation lapse rates in complex topographical regions–application in France. *Hydrol. Earth Syst. Sci.* **28**, 2579–2601 (2024).
59. Altaf, S. et al. Management of green mold disease in white button mushroom (Agaricus Bisporus) and its yield improvement. *J. Fungi.* **8**, 554 (2022).
60. Fisher, R. A. & Koven, C. D. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *J. Adv. Model. Earth Syst.* **12**, eMS001453 (2018).
61. Maxwell, A. E. & Shobe, C. M. Land-surface parameters for spatial predictive mapping and modeling. *Earth Sci. Rev.* **226**, 103944 (2022).
62. Sahoo, B. B., Sankalp, S. & Kisi, O. A novel smoothing-based deep learning time-series approach for daily suspended sediment load prediction. *Water Resour. Manage.* **37**, 4271–4292 (2023).
63. Satapathy, D. P., Swain, H., Sahoo, A., Samantaray, S. & Satapathy, S. C. In *Intelligent System Design: Proceedings of INDIA 2022* 355–364 (Springer, 2022).
64. Swagatika, S., Paul, J. C., Sahoo, B. B., Gupta, S. K. & Singh, P. Improving the forecasting accuracy of monthly runoff time series of the Brahmani River in India using a hybrid deep learning model. *J. Water Clim. Change.* **15**, 139–156 (2024).
65. Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R. & Schmidhuber, J. LSTM: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 2222–2232 (2016).
66. Momani, P. & Naill, P. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. *Am. J. Environ. Sci.* **5**, 599 (2009).
67. Bouznad, I. E. et al. Trend analysis and spatiotemporal prediction of precipitation, temperature, and evapotranspiration values using the ARIMA models: case of the Algerian highlands. *Arab. J. Geosci.* **13**, 1281 (2020).
68. Hasan, N. A., Dongkai, Y. & Al-Shibli, F. SPI and SPEI drought assessment and prediction using TBATS and ARIMA models, Jordan. *Water.* **15**, 3598 (2023).
69. Yucel, I. & Onen, A. Evaluating a mesoscale atmosphere model and a satellite-based algorithm in estimating extreme rainfall events in northwestern Turkey. *Nat. Hazards Earth Syst. Sci.* **14**, 611–624 (2014).
70. Kara, F., Yucel, I. & Akyurek, Z. Climate change impacts on extreme precipitation of water supply area in Istanbul: use of ensemble climate modelling and geo-statistical downscaling. *Hydrol. Sci. J.* **61**, 2481–2495 (2016).
71. Jamei, M. et al. Quantitative improvement of streamflow forecasting accuracy in the Atlantic zones of Canada based on hydro-meteorological signals: a multi-level advanced intelligent expert framework. *Ecol. Inf.* **80**, 102455 (2024).
72. Saha, S. et al. Measuring landslide vulnerability status of Chukha, Bhutan using deep learning algorithms. *Sci. Rep.* **11**, 16374 (2021).
73. Huffman, G. J. & Bolvin, D. T. TRMM and other data precipitation data set documentation. *NASA Greenbelt USA.* **28**, 1 (2013).

Acknowledgements

Author contributions

Conceptualization, supervision, methodology, formal analysis, writing—original draft preparation, writing—review and editing, O.A.W., S.S.M., M.Y., S.S.K., A.S.G.; data curation, project administration, investigation, writing—review and editing, F.D., N.A.-A., S.E.-H., M.A.M. All authors have read and agreed to the published version of the manuscript.

Funding

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77687-x.

Correspondence and requests for materials should be addressed to S.S.M., N.A.-A. or M.A.M.

Reprints and permissions information is available at www.nature.com/reprints.

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