MAUSAM, **77**, 3 (July 2026), 915-926

MAUSAM

DOI: https://doi.org/10.54302/mausam.v77i3.6934
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No.51.524:551.577.2

Comprehensive analysis and predictive modeling of temperature and rainfall patterns in Southern Telangana

GUHAN VELUSAMY[1*], DHARMA RAJU AKASAPU[2], SRAVANI ALANKA[3] and NAGARATNA KOPPARTHI[4]

[1]*Project Scientist, Meteorological Centre, Airport Colony, India Meteorological Department, Begumpet, Hyderabad, Telangana, India.*

[2]*Scientist – D, India Meteorological Department, New Delhi, India.*

[3]*Scientist – D, Meteorological Centre, Airport Colony, India Meteorological Department, Begumpet, Hyderabad, Telangana, India.*

[4]*Scientist – F, Meteorological Centre, Airport Colony, India Meteorological Department, Begumpet, Hyderabad, Telangana, India.*

(*Received 16 December 2024, Accepted 19 September 2025*)

***Corresponding author's email: guhanthiran@gmail.com**

सार – यह अध्ययन दक्षिणी तेलंगाना के 12 जिलों में तापमान और वर्षा के पैटर्न का डेटा-आधारित विश्लेषण प्रस्तुत करता है, जिसके लिए 43 वर्षों (1981-2023) के ग्रिडयुक्त मौसम संबंधी डेटा का उपयोग किया गया है। वर्णनात्मक सांख्यिकी, सहसंबंध मैट्रिक्स और छह पूर्वानुमान मॉडल - रैंडम फ़ॉरेस्ट (RF), आर्टिफिशियल न्यूरल नेटवर्क (ANN), सपोर्ट वेक्टर रिग्रेशन (SVR), लॉन्ग शॉर्ट-टर्म मेमोरी (LSTM), ARIMA और TBATS - का उपयोग करते हुए, हमने अधिकतम तापमान, न्यूनतम तापमान और वर्षा के पूर्वानुमान की सटीकता का मूल्यांकन किया। RF मॉडल ने सभी मापदंडों में सबसे कम टेस्ट RMSE (0.1178) और टेस्ट MAE (0.0601) के साथ बेहतर प्रदर्शन किया, जो पारंपरिक समय श्रृंखला मॉडलों से कहीं बेहतर है। सहसंबंध विश्लेषण से तापमान में मजबूत अंतर-स्थानिक समकालिकता ($r \approx 0.98$-1.00) का पता चला, जबकि वर्षा में उच्च स्थानिक परिवर्तनशीलता ($r = 0.15$-0.77) देखी गई, जो स्थानीय जलवायु प्रभावों को इंगित करती है। विशेषता महत्व विश्लेषण ने L332 और L333 को प्रमुख भविष्यवक्ता के रूप में पहचाना, जिनके स्कोर क्रमशः 0.1209 और 0.0557 थे। नवीन योगदानों में शामिल हैं: (1) दीर्घकालिक क्षेत्रीय जलवायु डेटा पर छह मॉडलों का तुलनात्मक मूल्यांकन, (2) व्याख्यात्मकता बढ़ाने के लिए विशेषता महत्व का एकीकरण, और (3) पूर्वानुमान अंतराल विश्लेषण जो सुसंगत ऊपरी सीमाओं (~0.187) और शून्य निचली सीमाओं के साथ मॉडल स्थिरता की पुष्टि करता है। ये निष्कर्ष अर्ध-शुष्क क्षेत्रों में जलवायु अनुकूलन, कृषि योजना और संसाधन प्रबंधन के लिए उपयोगी अंतर्दृष्टि प्रदान करते हैं।

ABSTRACT. This study presents a data-driven analysis of temperature and rainfall patterns across 12 districts in Southern Telangana using 43 years of gridded meteorological data (1981–2023). Employing descriptive statistics, correlation matrices, and six predictive models, Random Forest (RF), Artificial Neural Network (ANN), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), ARIMA, and TBATS, we evaluated forecasting accuracy for maximum temperature, minimum temperature, and rainfall.The RF model demonstrated superior performance with the lowest Test RMSE (0.1178) and Test MAE (0.0601) across all parameters, outperforming traditional time series models. Correlation analysis revealed strong inter-location temperature synchrony ($r \approx 0.98$–1.00), while rainfall exhibited high spatial variability ($r = 0.15$–0.77), indicating localized climatic influences. Feature importance analysis identified L332 and L333 as dominant predictors, with scores of 0.1209 and 0.0557, respectively. Novel contributions include: (1) a comparative evaluation of six models on long-term regional climate data, (2) integration of feature importance to enhance interpretability, and (3) prediction interval analysis confirming model stability with consistent upper bounds (~0.187) and zero lower bounds. These findings offer actionable insights for climate adaptation, agricultural planning, and resource management in semi-arid regions.

Key words – Temperature, Rainfall, Summary Data, Correlation, Machine learning, Random forest.

1. Introduction

Climate change is reshaping the natural systems that sustain life, affecting not only humans, but also the land, water, and ecosystems we depend on. Recognizing these evolving patterns is essential for effective planning and resource management (Pec *et al*., 2017). Among the most critical indicators of climate change are temperature and precipitation trends, which offer valuable insights into environmental transformations and societal impacts (Renforth *et al*., 2025). India, with its diverse climatic zones, presents a compelling landscape for climatological research. Studies have shown that geographic and climatic factors significantly influence regional variations in temperature and rainfall across the country (Yaduvanshi *et al*., 2021). Southern Telangana, located on the Deccan Plateau, exemplifies this complexity with its semi-arid climate and pronounced fluctuations in both temperature and precipitation (Anand *et al*, 2022).

This study aims to investigate the spatial and temporal variations in rainfall and temperature across Southern Telangana using a hybrid approach that combines descriptive statistics, correlation analysis, and advanced machine learning algorithms. The primary objectives are to enhance the accuracy of weather predictions and to generate actionable insights into regional climatic shifts (AlShafeey., 2024). Descriptive statistics are employed to examine central tendencies, dispersion, and distribution patterns, offering a comprehensive understanding of baseline climate conditions (He *et al*., 2023). These methods are essential for detecting anomalies and long-term trends (Zamanzadeh *et al*., 2024), while correlation analysis helps uncover relationships among climatic variables across different locations.

High correlation values suggest synchronous climatic behavior, often driven by shared geographic or meteorological influences. In Southern Telangana, significant linear correlations between maximum and minimum temperatures point to consistent temperature fluctuations across the region (Gaur *et al*., 2025). The growing adoption of predictive modeling in climate research reflects the effectiveness of machine learning techniques in processing large datasets and identifying complex patterns (Hamdan *et al*., 2024). Models such as Random Forest (RF), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) are increasingly used to forecast weather parameters (Rangelov *et al*., 2023). This study evaluates the performance of these models in predicting rainfall and temperature trends in Southern Telangana (Sreelatha & Anand Raj, 2021).

Machine learning algorithms have demonstrated superior accuracy and reliability compared to traditional statistical approaches, significantly advancing the field of climate prediction (Zhang *et al*., 2025). Among these, the Random Forest model stands out for its ability to detect hidden patterns and adapt to new data inputs (Salman *et al*, 2024). This research further explores the potential of machine learning to improve regional weather forecasting and climate analysis (Yadav *et al*., 2025).

Feature importance analysis is conducted to identify the most influential meteorological variables, enabling model refinement and prioritization of key attributes (Ries *et al*., 2024). As noted by Gbaguidi *et al*. (2025), understanding the relative significance of different indicators is crucial for improving prediction accuracy. Additionally, prediction intervals are used to quantify uncertainty, an inherent aspect of climate modeling, by providing a range of expected outcomes (Gbaguid *et al*., 2025). The stability of these intervals confirms the reliability of the models used in this study (Li., 2022).

Ultimately, the findings have important implications for planning and resource management in Southern Telangana. Accurate weather forecasting supports agricultural decision-making, water resource allocation, and mitigation of extreme weather impacts (Mulla *et al*., 2023). As emphasized by Chen *et al*. (2023), the integration of machine learning into climate studies offers a powerful tool for enhancing forecast precision and deepening our understanding of climatic trends, contributing to more resilient and informed regional development strategies.

2. Data and methodology

2.1. *Study area*

This study focuses on twelve major districts in Southern Telangana, India: Hyderabad, Jogulamba Gadwal, Mahabubnagar, Medchal-Malkajgiri, Nagarkurnool, Nalgonda, Narayanpet, Rangareddy, Suryapet, Vikarabad, Wanaparthy, and Yadadri Bhuvanagiri (Fig. 1). These districts collectively represent

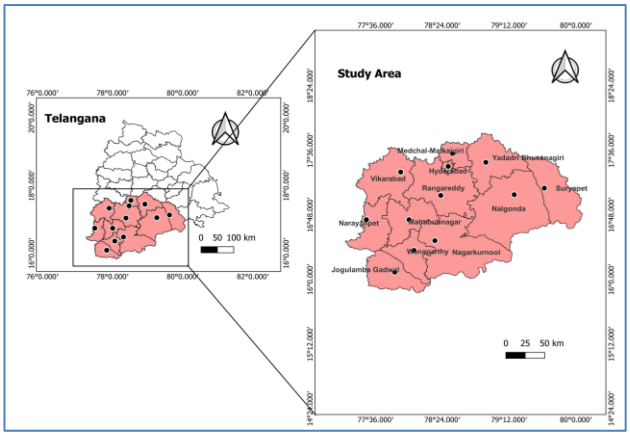


Fig. 1. Study Area Map

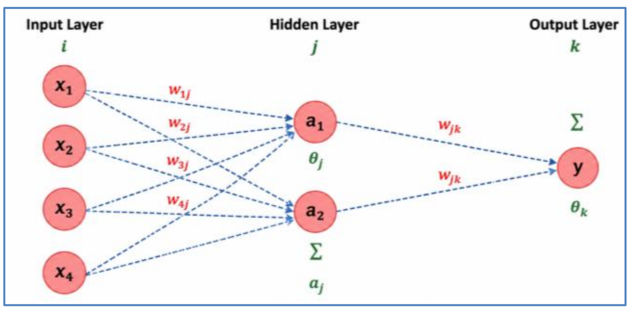


Fig. 2. ANN model architecture

the Southern Telangana region, which is characterized by a semi-arid climate and diverse topographical features. Gridded climate data, including rainfall, minimum temperature, and maximum temperature—were obtained from the India Meteorological Department (IMD) for a 43-year period spanning 1981 to 2023. To ensure data quality and model compatibility, a series of preprocessing steps were applied. Missing values were addressed using linear interpolation, and the data were normalized using Min-Max scaling. Time series data were then transformed into a supervised learning format to facilitate machine learning model training. To enhance model robustness and reduce noise, outliers were identified using z-score thresholds and subsequently removed. These preprocessing techniques ensured a clean, consistent dataset suitable for accurate climate analysis and predictive modeling across the selected districts.

2.2. *Machine learning (ML) algorithms*

Based on the data set's behavior, the ML algorithms were used to forecast the weather parameters. Below is a discussion of the algorithms:

2.3. *Artificial neural network*

The brain's structure inspired artificial neural networks (ANN) (Eq 1), a type of machine learning method (see Fig. 2). ANN typically consist of three layers: input, output, and hidden layers. In the output layer, predictions are generated, while the hidden layers manipulate incoming data using weights and biases. The number of hidden layers and activation functions in the model directly impacts its capacity and learning capabilities. Hyperparameters such as learning rate (0.01), number of hidden layers (2), and activation function (ReLU) were optimized using grid search.

$$y_t = \alpha_0 + \sum_{j=1}^{q} a_j \, g\left(\beta_{0j} + \sum_{i=1}^{k} \beta_{ij} \, y_{t-i}\right) + €_t \quad (1)$$

The model's connection weights are aj (j = 0,1, 2,...,q) and β ij (j = 0,1, 2,...,q; i = 1,2,.,k), where k and q are the numbers of nodes in the input and hidden layers, respectively.

2.4. *Support vector regression (SVR)*

A machine learning model called SVR is applied to time series and regression issues (Eq 2). By leveraging kernel functions, the system transforms input features into a high-dimensional space, identifies optimal boundaries, utilizes a cost function to estimate parameters, and predicts incoming input data. Improving performance is achievable through the fine-tuning of hyperparameters like the kernel function and regularization parameter. Support Vector Regression (SVR) can effectively manage intricate non-linear relationships.

$$f(x) = w^T \emptyset(x) + b \quad (2)$$

w: A feature space coefficient vector
$\emptyset(x)$: A feature space vector that is mapped from input x by a kernel function
b : An intercept

The radial basis function (RBF) kernel was selected after tuning kernel type, regularization parameter (C = 1.0), and epsilon (ε = 0.1) using cross-validation.

2.5. *Random forest (RF)*

The ML algorithm RF utilizes decision trees to enhance the stability and accuracy of the model, as depicted in Fig. 3. This methodology involves recursively partitioning the data based on input features, training multiple trees on randomly sampled data subsets. The average of all the trees is the final prediction, which lowers prediction variance and overfitting. Hyperparameters such as number of trees (n_estimators = 100), maximum depth (max_depth = 10), and minimum samples per leaf (min_samples_leaf = 2) were tuned using randomized search.

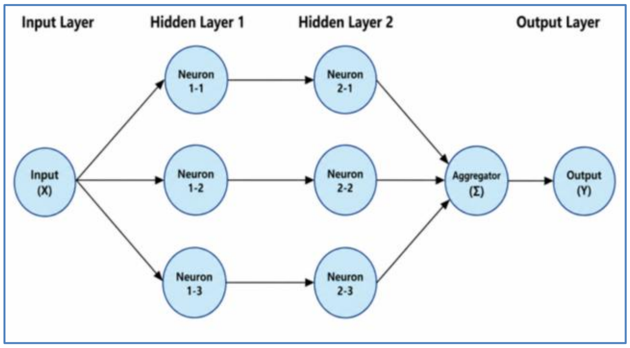


Fig. 3. RF architecture

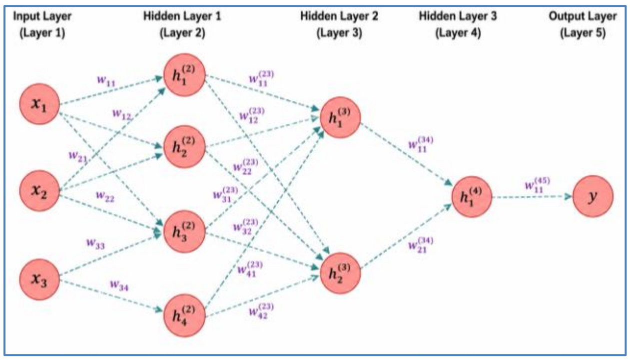


Fig. 4. The composition of LSTM unit

2.6. *Long short-term memory (LSTM)*

One kind of RNN that handles sequential data, such as text, audio signals, and time series, is called an LSTM (Fig. 4). It addresses the vanishing gradient issue in conventional RNNs and was first introduced in 1997. The input, forget, output, and memory cells that make up LSTM networks regulate the information that enters and exits the memory cells. The LSTM model was trained using a sequence length of 30 days, with a batch size of 64 and 50 epochs. Dropout (0.2) was applied to prevent overfitting.

In this context, C_t represents the cell state, f_t symbolizes the input or forget gate, C_{t-1} denotes the previous cell state, and $\sim C_t$ stands for the candidate cell state. At time step t, the candidate's input weight and bias are denoted by W_c and b_c. The hidden state h_t corresponds to the output gate represented as o_t. Moreover, the output gate and cell state are identified as o_t and C_t respectively, while Wo indicates the weight matrix employed in the process. The terminology used here distinguishes different components of the LSTM architecture.

2.7. *ARIMA (Autoregressive Integrated Moving Average)*

The ARIMA model is used for forecasting time series data in a univariate setting. It requires the data to be stationary, exhibiting a consistent mean, variance, and covariance over time. Model selection relies on autocorrelation functions and partial autocorrelation functions, determined by the parameters p, d, and q. Calibration of the model involves maximum likelihood estimation, validation through residual analysis, and comparison with actual data to assess accuracy.

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

Complex patterns including seasonality, trend, and irregularity may be predicted using the TBATS time series forecasting method, which combines exponential smoothing and ARIMA models. It uses maximum likelihood estimation to estimate parameters and forecast future values, modeling the time series as a mixture of many components.

2.9. *Statistical evaluation*

All models were trained using an 80:20 train-test split. Five-fold cross-validation was applied to ensure generalizability. Performance was evaluated using MAE and RMSE, metrics. It is essential for evaluating prediction models' performance since many models may provide numbers that are close or almost the same. Models are thoroughly assessed using a variety of statistical criteria, enabling strong comparison analysis. Better model accuracy is shown by lower values of the MAE (Eq 3) and RMSE (Eq 4), two frequently used metrics that quantify the average size of errors between projected and actual values. To evaluate the model's performance, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) measures are employed.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

3. Results and discussion

3.1. *Descriptive Analysis of Climatic Parameters in Southern Telangana*

The climatic profile of Southern Telangana, based on data from twelve distinct regions, reveals insightful patterns in maximum temperature (Table 1 & Figs. 5a-c), minimum temperature (Table 2 & Figs. 5a-c), and rainfall

TABLE 1

Descriptive statistics of Maximum Temperature (ºC) at different locations in Southern Telangana

Descriptive Statistics	Maximum temperature											
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
Mean	33.30	33.51	33.51	33.30	33.50	33.83	33.51	33.30	33.83	33.17	33.50	33.30
Standard error	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Median	32.21	32.54	32.54	32.21	32.54	32.80	32.54	32.21	32.80	32.08	32.54	32.21
Mode	30.94	31.07	31.07	30.94	30.96	32.77	31.07	30.94	32.77	31.20	30.96	30.94
Standard deviation	4.05	3.86	3.86	4.05	3.89	4.03	3.86	4.05	4.03	4.01	3.89	4.05
Kurtosis	-0.36	-0.53	-0.53	-0.36	-0.46	-0.34	-0.53	-0.36	-0.34	-0.45	-0.46	-0.36
Skewness	0.65	0.58	0.58	0.65	0.59	0.66	0.58	0.65	0.66	0.62	0.59	0.65
Range	25.48	23.12	23.12	25.48	23.67	25.13	23.12	25.48	25.13	24.51	23.67	25.48
Minimum	19.58	21.12	21.12	19.58	20.69	20.90	21.12	19.58	20.90	19.78	20.69	19.58
Maximum	45.06	44.24	44.24	45.06	44.36	46.03	44.24	45.06	46.03	44.29	44.36	45.06

TABLE 2

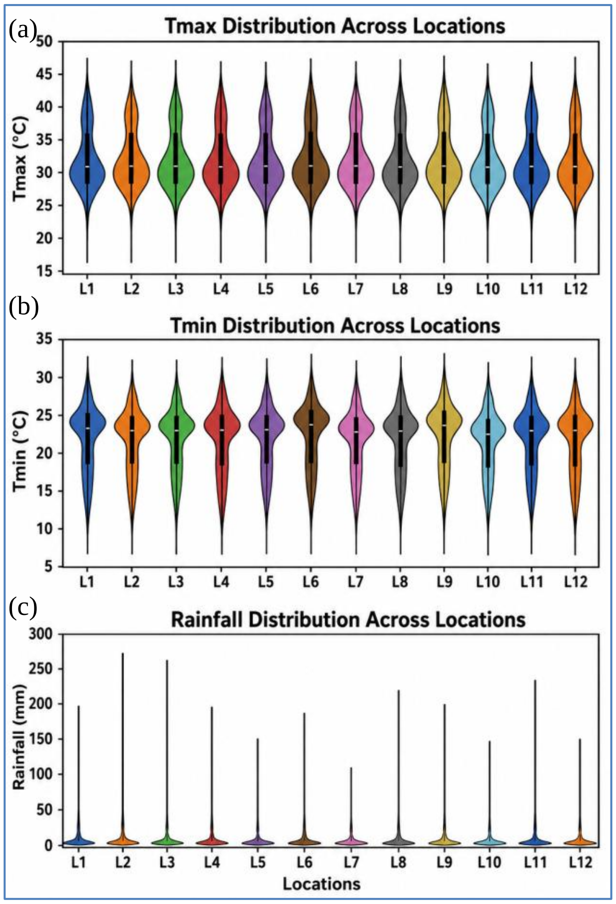
Descriptive statistics of Minimum Temperature (ºC) at different locations in Southern Telangana

Descriptive Statistics	Maximum temperature											
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
Mean	21.32	21.23	21.23	21.32	21.45	22.11	21.23	21.32	22.11	20.87	21.45	21.32
Standard error	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Median	22.29	22.00	22.00	22.29	22.25	23.19	22.00	22.29	23.19	21.67	22.25	22.29
Mode	23.01	22.06	22.06	23.01	22.55	23.92	22.06	23.01	23.92	22.05	22.55	23.01
Standard deviation	3.78	3.40	3.40	3.78	3.47	3.86	3.40	3.78	3.86	3.58	3.47	3.78
Kurtosis	-0.39	-0.29	-0.29	-0.39	-0.40	-0.45	-0.29	-0.39	-0.45	-0.31	-0.40	-0.39
Skewness	-0.47	-0.49	-0.49	-0.47	-0.45	-0.51	-0.49	-0.47	-0.51	-0.46	-0.45	-0.47
Range	20.90	20.30	20.30	20.90	19.77	21.96	20.30	20.90	21.96	20.19	19.77	20.90
Minimum	9.46	9.35	9.35	9.46	10.30	9.43	9.35	9.46	9.43	9.13	10.30	9.46
Maximum	30.36	29.65	29.65	30.36	30.07	31.39	29.65	30.36	31.39	29.32	30.07	30.36

TABLE 3

Descriptive statistics of rainfall at different locations in Southern Telangana (mm)

Descriptive Statistics	Maximum temperature											
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
Mean	2.36	1.84	3.53	2.36	1.90	2.22	1.99	2.12	2.38	2.19	2.02	2.25
Standard error	0.06	0.06	0.10	0.06	0.06	0.06	0.05	0.06	0.07	0.06	0.06	0.06
Median	0.00	0.00	0.00	0.00	0.00	0.00	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	0.00	0.00	0.00	0.00	0.00	0.00
Standard deviation	8.11	7.10	12.18	8.11	7.01	7.99	6.10	7.41	8.55	7.18	7.16	7.63
Kurtosis	75.25	390.37	85.56	75.25	81.33	85.32	39.99	114.97	84.12	47.85	132.32	61.64
Skewness	6.93	12.43	7.33	6.93	7.37	7.41	5.29	8.05	7.38	5.89	8.35	6.44
Range	190.11	342.54	271.20	190.11	142.34	181.50	100.11	214.57	193.36	130.38	230.71	142.15
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	190.11	342.54	271.20	190.11	142.34	181.50	100.11	214.57	193.36	130.38	230.71	142.15



Figs. 5(a-c). Distribution of Maximum Temperature (ºC), Minimum Temperature (ºC), and Rainfall (mm) across Different Location. (a) Maximum Temperature distribution, (b) Minimum Temperature distribution, (c) Rainfall distribution

Table 3 & Figs. 5 (a-c). The average maximum temperature ranges from 33.30 °C to 33.83 °C, showing remarkable stability across the region. A consistently low standard error of around 0.03 percent indicates minimal sample variability, reinforcing the reliability of the mean values. Median temperatures closely mirror the means, suggesting a symmetric distribution, while modal values reflect common temperature experiences with only minor regional differences (Grotjahn *et al*., 2016). Standard deviations between 3.86 and 4.05 highlight limited fluctuations in daily maximum temperatures. Negative kurtosis values from -0.53 to -0.34 suggest a slightly platykurtic distribution, with fewer extreme deviations than a normal curve. Skewness values between 0.58 and 0.66 indicate a slight right skew, implying that lower-than-average temperatures are more frequently observed. Temperature readings span from 23 °C to 25 °C, with optimal temperatures between 19 °C and 21 °C and peaks reaching 44 °C to 46 °C, showcasing the region's daily temperature variability. Similarly, minimum temperature data shows regional mean values ranging from 20.87 °C to 22.11 °C, supported by a low standard error of approximately 0.03 percent. Median values align closely with the means, confirming central tendency, while mode values vary slightly across locations, indicating the most likely minimum temperatures. Standard deviations from 3.40 to 3.86 reflect notable variability, and negative kurtosis values between -0.45 and -0.29 suggest a flatter distribution with fewer extremes. The temperature range spans 19.77 °C to 21.96 °C, with minimums between 9.13 °C and 10.30°C and maximums from 29.32 °C to 31.39 °C, highlighting significant daily variation and a symmetric distribution pattern. Rainfall analysis reveals substantial variation across the twelve regions, with total rainfall ranging from 100.11 mm to 342.54 mm. However, the daily mean rainfall values remain consistently close to 0.00 mm, reflecting the high frequency of rainless days typical of semi-arid climates. This low mean, when averaged over a long time span (1981–2023), masks the intensity of sporadic rainfall events. Standard error values between 0.05 and 0.10 affirm the precision of mean calculations, while standard deviations from 6.10 to 12.18 underscore notable daily fluctuations. Exceptionally high kurtosis values at sites L2 and L11 point to heavy-tailed distributions, indicative of rare but intense rainfall events. Skewness values ranging from 5.29 to 12.43 reveal a pronounced right skew, suggesting a predominance of low-rainfall days punctuated by occasional heavy downpours. These patterns, supported by prior studies, emphasize the erratic nature of precipitation in the region and underscore the importance of recognizing such variability for effective water resource management and agricultural planning (Muzammal *et al*., 2024).

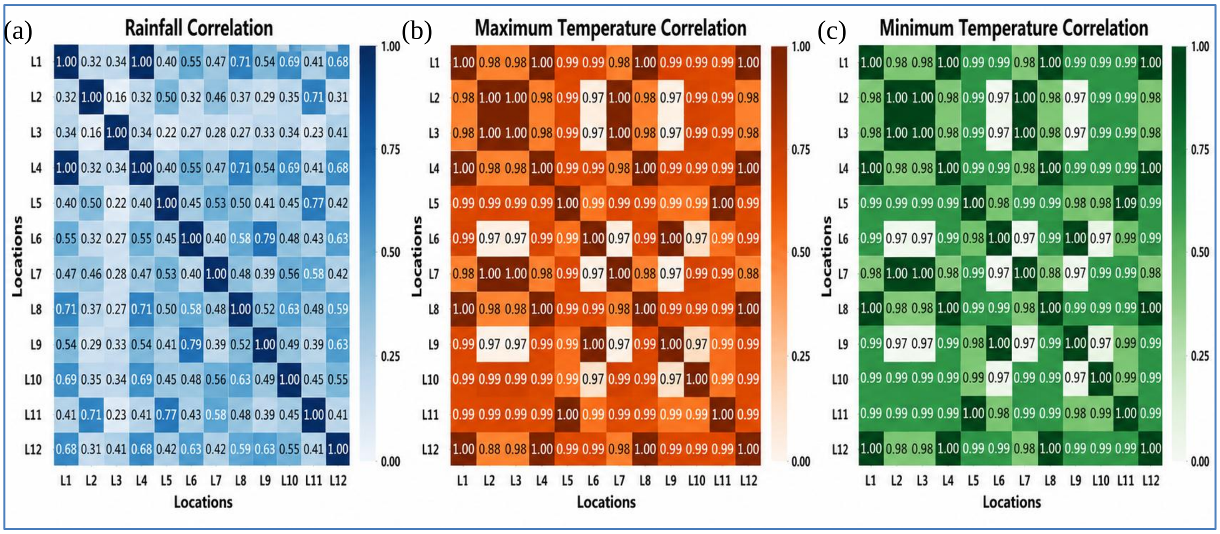
TABLE 4

Seasonal Trends and Residuals for Rainfall and Temperature in Southern Telangana

Parameter	Seasonal	Trend	Residual
Maximum Temperature	33.299	33.299	33.299
Minimum Temperature	21.318	21.318	21.318
Rainfall	-0.001	2.355	-0.002

3.2. *Correlation matrix*

The correlation analysis of maximum temperature, minimum temperature, and rainfall across various locations in Southern Telangana reveals distinct climatic patterns. The correlation matrix for maximum temperature Figs. 6(a-c) shows consistently high values, typically ranging from 0.98 to 1.00, indicating a strong linear relationship and suggesting that maximum temperatures tend to fluctuate in unison across the region. Similarly, minimum temperature correlations are also high-often between 0.98 and 1.00-demonstrating a synchronized trend across different sites, likely driven by shared



Figs. 6(a-c). Correlation Matrices of Rainfall (mm), Maximum Temperature (°C), and Minimum Temperature (°C) Across 12 Districts of Southern Telangana. (a) Rainfall correlation matrix, (b) Maximum Temperature correlation matrix, (c) Minimum Temperature correlation matrix

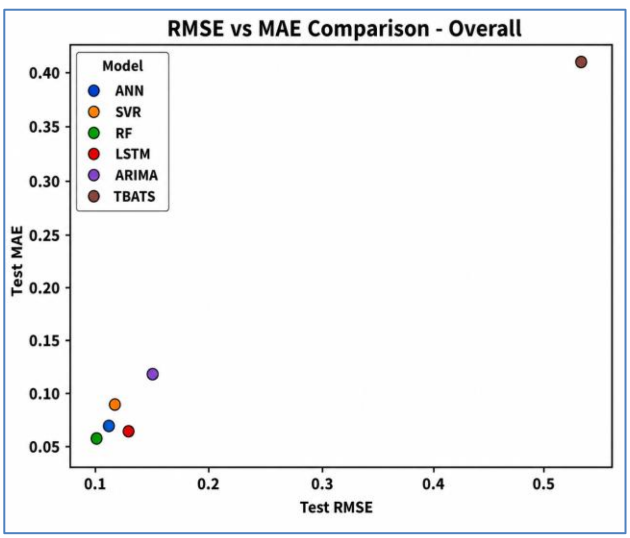


Fig. 7. Scatter plot comparing several models RMSE and MAE

climatic influences. In contrast, the rainfall correlation matrix presents a more varied pattern, with values ranging from 0.15 to 0.77 (Figs. 6a-c). This variability suggests that rainfall events are not uniformly distributed and are significantly influenced by localized factors such as microclimate conditions, geological features, and elevation (Shetty *et al.*, 2022). These results indicate that while temperature variations across Southern Telangana are relatively uniform, rainfall patterns are more spatially confined and erratic (Chandu *et al.*, 2025).The findings underscore the consistent nature of temperature changes in the region, contrasted by the unpredictable and uneven distribution of rainfall. This highlights the importance of developing region-specific weather forecasting models and climate adaptation strategies that account for the unique challenges posed by Southern Telangana's variable rainfall patterns (Alsharef & Hassan, 2024).

3.3. *Model performance*

The performance of machine learning models for parameter prediction reveals varying levels of accuracy (Fig. 7). The Random Forest (RF) model achieved the lowest Train MAE and Test MAE, demonstrating its strength in capturing complex data patterns and generalizing effectively. The Artificial Neural Network (ANN) produced strong results, with a Test RMSE of 0.125 and a Test MAE of 0.069. The Support Vector Regression (SVR) model showed room for improvement, with a Test RMSE of 0.129 and a Test MAE of 0.087. The Long Short-Term Memory (LSTM) model exhibited signs of overfitting to the training data, while the ARIMA model recorded higher error rates, indicating limited suitability for this dataset. The TBATS model also showed notable errors, suggesting its inadequacy for the current prediction task. This assessment highlights the diverse performance characteristics of different machine learning models in parameter prediction (Patel *et al.*, 2025). Fig. 8 illustrates the comparative performance of each model across all evaluation metrics. These findings emphasize the importance of careful model selection and tuning to match the specific requirements of a given task. To enhance prediction accuracy, future research may focus on optimizing the hyperparameters of the SVR and LSTM

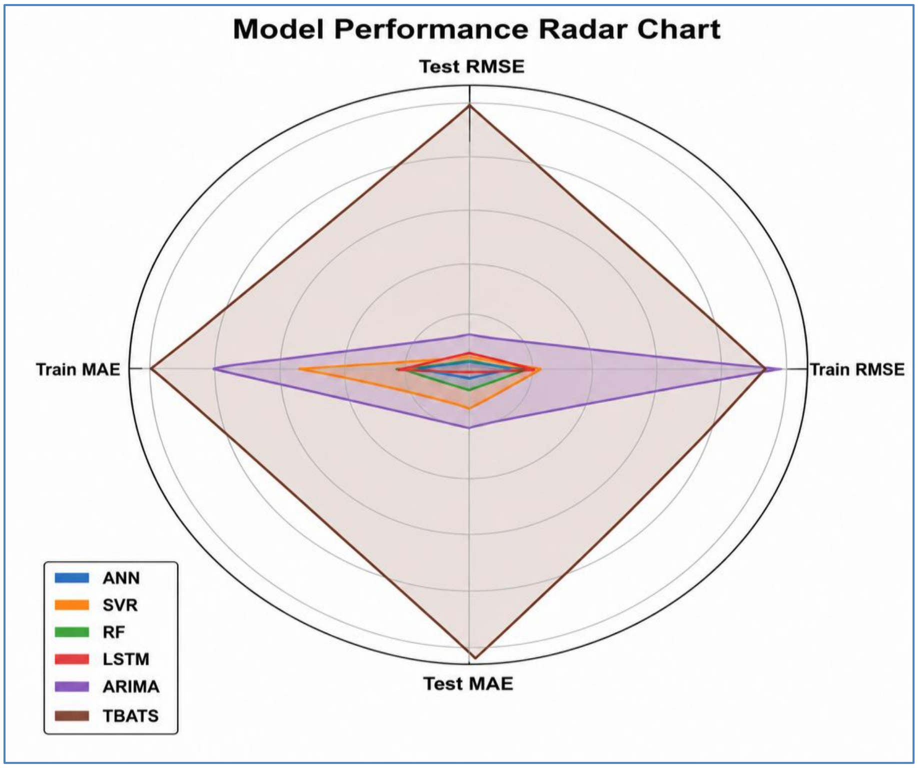


Fig. 8. A radar image that shows the performance of each model across a number of performance measures

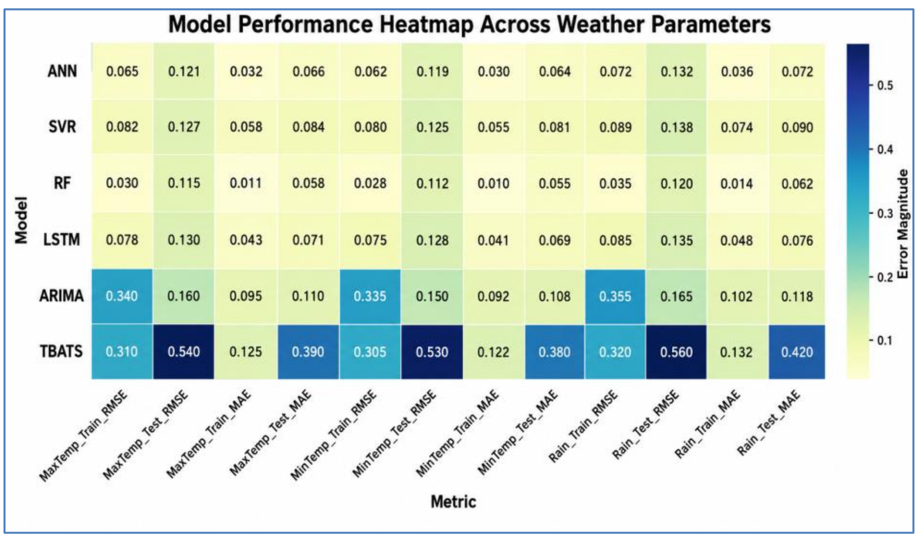


Fig. 9. Comparative Predictive Performance of Random Forest and Other Models Across Climatic Parameters in Southern Telangana

models and exploring ensemble approaches that combine the strengths of multiple algorithms (Shaikh *et al*., 2024).

3.4. *Comparative analysis of predictive models for weather parameters*

Across all three climate parameters, maximum temperature, minimum temperature, and rainfall, the Random Forest (RF) model consistently demonstrated superior predictive performance (Fig. 9). For maximum temperature, RF achieved the lowest Test RMSE of **0.115** and Test MAE of **0.058**, outperforming both traditional statistical models like ARIMA and TBATS, and advanced machine learning models such as ANN, SVR, and LSTM. Similarly, in minimum temperature prediction, RF recorded a Test RMSE of **0.112** and Test MAE of **0.055**,

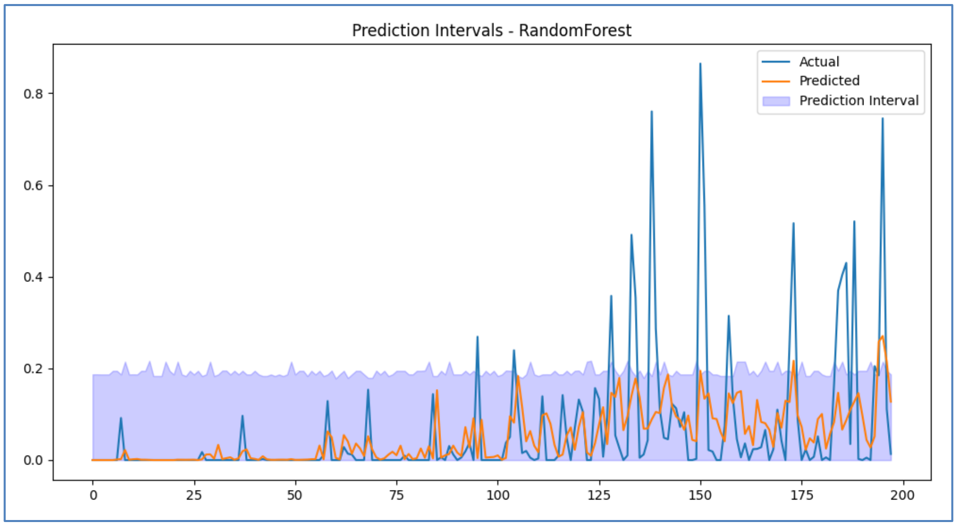


Fig. 10. Comprehensive Analysis of Prediction Intervals for Rainfall and Temperature

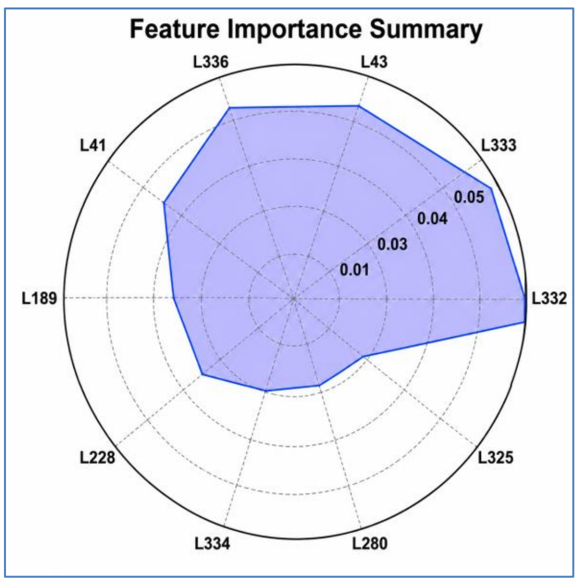


Fig. 11. Feature importance summary

indicating its strong generalization capability and minimal error margins. These results underscore RF's robustness in modeling stable temperature patterns across Southern Telangana. Rainfall prediction, which typically involves greater variability and noise, also favored the RF model. It delivered a Test RMSE of **0.120** and Test MAE of **0.062**, significantly lower than other models, especially TBATS which showed the highest error rates. While ANN and LSTM performed moderately well, their accuracy was consistently outpaced by RF across all metrics. The consistent performance of RF across diverse climatic parameters highlights its adaptability and reliability, making it a valuable tool for regional weather forecasting, agricultural planning, and climate resilience strategies (Makokha *et al*., 2025).

3.5. *Seasonal dynamics and long-term trends in rainfall and temperature*

Table 4 provides a detailed analysis of the seasonal, trend, and residual components for maximum temperature, minimum temperature, and rainfall in Southern Telangana, with all values rounded to three decimal places. The rainfall data reveals a slight negative seasonal variation of -0.001, indicating minimal fluctuations throughout the year. Its trend component, at 2.355, suggests a gradual increase in rainfall over the observation period, while the residual value of -0.002 reflects minor discrepancies between observed and predicted values, demonstrating the model's overall accuracy. In contrast, both maximum and minimum temperatures show remarkable stability, with identical values across seasonal, trend, and residual components: 33.299 for maximum temperature and 21.318 for minimum temperature. This uniformity indicates no significant seasonal or long-term variations in temperature patterns during the study period (He *et al*., 2019).

3.6. *Prediction interval summary analysis*

The prediction interval summary results for the analysis show consistent upper bounds of approximately 0.187 across multiple intervals, with occasional variations (Fig. 10). The data consistently forecasts zero or positive outcomes with no negative deviations, as seen by the lower bounds staying uniformly at 0. This points to a restricted and constant range for anticipated values, suggesting a stable and dependable prediction model with upper limits often around 0.187. There is no variation in forecast accuracy in the intervals with somewhat larger upper limits, such as 0.194, 0.195, and sometimes even up to 0.216. All things considered, the study shows a consistent capacity to forecast the characteristics under evaluation. Overall, the study demonstrates a consistent ability to anticipate the parameters being assessed. Its consistent performance with short prediction intervals demonstrates its accuracy and reliability. This type of model is very useful in domains like resource allocation, disaster management, and agricultural planning where precise and trustworthy forecasts are essential (Jaber *et al*., 2022). The model's predicted accuracy might be improved by additional refinement, and its resilience and adaptability could be validated by applying it to other datasets and parameters (Vairo *et al*., 2024).

3.7. *Future importance*

The relative importance of characteristics (L1 to L360) in the study is illustrated by the feature importance summary results (Fig. 11). The features with the highest significance, L43 (0.05), L41 (0.04), L332 (0.12), and L333 (0.06), exert the greatest influence on the prediction model. Additional impactful features include L44 (0.01), L12 (0.01), and L116 (0.01), which also contribute meaningfully to the model's predictions. In contrast, the remaining features, with lower importance scores, appear to have minimal impact. By identifying the most influential variables, these findings lay the groundwork for further research and model refinement. Feature importance plays a vital role in guiding data-driven decision-making and pinpointing the most critical components for accurate predictions (Sarker, 2021). Future studies should explore advanced feature selection techniques and validate their effectiveness across diverse datasets to enhance model performance (Ali *et al*., 2024).

4. Conclusions

This study offers a comprehensive climatological assessment of Southern Telangana, analyzing 43 years of gridded meteorological data (1981–2023) across 12 districts. Through descriptive statistics, correlation analysis, and predictive modeling, we examined the spatial and temporal dynamics of maximum temperature, minimum temperature, and rainfall. The temperature data revealed high consistency across the region, with mean maximum temperatures ranging from **33.17 °C to 33.83 °C** and minimum temperatures from **20.87 °C to 22.11 °C.** Correlation coefficients for both maximum and minimum temperatures exceeded **0.98**, indicating strong spatial synchrony and suggesting the influence of regional climatic drivers. In contrast, rainfall exhibited significant variability, with mean values ranging from **1.84 mm to 3.53 mm**, and correlation values between locations ranging from **0.15 to 0.77**, highlighting localized and erratic precipitation patterns. Among the six predictive models evaluated-ANN, SVR, RF, LSTM, ARIMA, and TBATS, the **Random Forest (RF)** model consistently outperformed others. It achieved the lowest Test RMSE and MAE across all parameters: **0.115 / 0.058** for maximum temperature, **0.112 / 0.055** for minimum temperature, and **0.120 / 0.062** for rainfall. These results underscore RF's robustness in handling both stable and volatile climate data. Feature importance analysis further identified key predictors, notably **L332 (importance score: 0.1209)** and **L333 (0.0557)**, which significantly influenced model accuracy. The prediction interval analysis demonstrated model reliability, with consistent upper bounds **(~0.187)** and zero lower bounds, confirming the stability of RF forecasts.The novelty of this research lies in its integration of long-term regional climate data with a comparative evaluation of machine learning and statistical models, enhanced by interpretability through feature importance and uncertainty quantification. These insights provide a foundation for developing localized, data-driven forecasting systems and climate adaptation strategies tailored to semi-arid regions like Southern Telangana.

Acknowledgements

This study was carried out as part of the Gramin Krishi Mausam Sewa (GKMS) initiative of the India Meteorological Department. The authors express their sincere gratitude to the Director General of Meteorology, IMD, for his support and deeply appreciate the constructive feedback from colleagues and reviewers.

Funding

No funding was received for this study.

Data Availability Statement

The dataset used in this study is publicly available on (https://imdpune.gov.in/cmpg/Griddata/). Additional processed data and code can be accessed upon request.

Authors's Contirbutions

Guhan Velusamy: Conceptualization, Methodology, Formal Analysis, Visualization, Validation, original draft, review & editing. (*email: guhanthiran@gmail.com*).
Dharma Raju Akasapu: Conceptualization, Methodology, Supervision, Visualization, review & editing. (*email: dharmaraju.met@gmail.com*).
Sravani Alanka: Supervision, review & editing. (*email: srisravani.a@gmail.com*).
Nagaratna Kopparthi: Supervision, review & editing. (*email: knratna@gmail.com).*

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

References

Ali, M. Z., Abdullah, A., Zaki, A. M., Rizk, F. H., Eid, M. M. and El-Kenway, E. M., 2024, "Advances and challenges in feature selection methods: a comprehensive review", *J. Artif. Intell. Metaheuristics*, **7**, 1, 67-77. doi: http://dx.doi.org/10.54216/JAIM.070105.

AlShafeey, M. 2024. Unraveling climate trends in the mediterranean: a hybrid machine learning and statistical approach. *Modeling Earth Systems and Environment*, **10,** 5, 6255-6277. doi:https://doi.org/10.1007/s40808-024-02117-w.

Alsharef, O. M., and Hassan, M. A. S. 2024. Climate change resilience strategies for sustainable development: Integrating weather forecasts and adaptation measures. *Journal for Research in Applied Sciences and Biotechnology*, **3**, 2, 59-65. doi: https://doi.org/10.55544/jrasb.3.2.13

Anand, N., Satheesh, S. K., and Moorthy, K. K. 2022. Land-atmosphere interactions at a semi-arid region in the Deccan Plateau. *Journal of Geophysical Research: Atmospheres*, **127**, 21, e2022JD037211. doi: https://doi.org/10.1029/2022JD037211.

Chandu, K., Raju, A. D., Kumar, S. V. J., Satyanarayana, G. C., and Dasari, M. (2025). Rainfall in Telangana (India): a decadal trend analysis. *Proceedings of the Indian National Science Academy*, 1-17. doi: https://doi.org/10.1007/s43538-025-00454-7.

Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., and Yang, Z. 2023. Machine learning methods in weather and climate applications: A survey. *Applied Sciences*, **13**, 21, 12019. doi: https://doi.org/10.3390/app132112019.

Gaur, P., Vadla, P. K., Sudarshan, P., Kedeelaya, P., and Narendra Babu, C. 2025. Discovering new relationships using multipoles through correlation and threshold analysis. *Intelligent Decision Technologies*, **19**, 3, 1344-1364. doi: https://doi.org/10.1177/18724981251318529.

Gbaguidi, G. J., Topanou, N., Filho, W. L., Agboka, K., and Ketoh, G. K. 2025. Unleashing the power of intelligence: revolutionizing malaria outbreak preparedness with an advanced warning system in Benin, West Africa. *Archives of Public Health*, **83**,1, 102. doi: https://doi.org/10.1186/s13690-025-01554-y.

Gbaguidi, G. J., Topanou, N., Walter Filho, L., and Ketoh, G. K. 2025. Assessing vulnerability to malaria in the context of climate change: a comprehensive study in northern Benin, West Africa. *Discover Social Science and Health*, **5**, 1, 1-17.

Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., ... and Prabhat. 2016. North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, **46**, 3, 1151-1184. doi: https://doi.org/10.1007/s00382-015-2638-6.

Hamdan, A., Ibekwe, K. I., Etukudoh, E. A., Umoh, A. A., and Ilojianya, V. I. 2024. AI and machine learning in climate change research: A review of predictive models and environmental impact. *World Journal of Advanced Research and Reviews*, **21**, 1, 1999-2008. doi: https://doi.org/10.30574/wjarr.2024.21.1.0257.

He, W., Lian, J., Zhang, J., Yu, X., and Chen, S. 2019. Impact of intra-annual runoff uniformity and global warming on the thermal regime of a large reservoir. *Science of The Total Environment*, **658**, 1085-1097. doi:https://doi.org/10.1016/j.scitotenv.2018.12.207.

He, Y., Ma, J., and Chen, G. 2023. Potential geographical distribution and its multi-factor analysis of Pinus massoniana in China based on the maxent model. *Ecological Indicators*, **154**, 110790. doi: https://doi.org/10.1016/j.ecolind.2023.110790.

Jaber, M. M., Ali, M. H., Abd, S. K., Jassim, M. M., Alkhayyat, A., Aziz, H. W., and Alkhuwaylidee, A. R. 2022. Predicting climate factors based on big data analytics based agricultural disaster management. *Physics and Chemistry of the Earth, Parts A/B/C*, **128**, 103243. doi: https://doi.org/10.1016/j.pce.2022.103243.

Li, H. 2022. SCADA data based wind power interval prediction using LUBE-based deep residual networks. *Frontiers in Energy Research*, **10**, 920837. doi:https://doi.org/10.3389/fenrg.2022.920837.

Makokha, J. W., Barasa, P. W., and Khamala, G. W. 2025. Enhancing climate resilience: A data-driven north rift weather prediction system for real-time forecasting and agricultural decision support. *Heliyon*, **11**,4. doi: https://doi.org/10.1016/j.heliyon.2025.e42549.

Mulla, S., Ahmed, R., Singh, K. K., Singh, S. K., Deshmukh, N., and Inamdar, F. K. 2023. Climate change effect on-climate parameters like temperature, rainfall and water resources sectors in India. In *Climate change impacts in India* (pp. 9-59). Cham: Springer International Publishing. doi: https://doi.org/10.1007/978-3-031-42056-6_2.

Muzammal, H., Zaman, M., Safdar, M., Adnan Shahid, M., Sabir, M. K., Khil, A., ... and Zaib, A. 2024. Climate change impacts on water resources and implications for agricultural management. In *Transforming agricultural management for a sustainable future: Climate change and machine learning perspectives* (pp. 21-45). Cham: Springer Nature Switzerland. doi: https://doi.org/10.1007/978-3-031-63430-7_2.

Patel, M., Patel, S. B., Swain, D., and Mallagundla, R. 2025. Enhancing Accuracy in Hourly Passenger Flow Forecasting for Urban Transit Using TBATS Boosting. *Modelling*, **6**, 2, 32. doi: https://doi.org/10.3390/modelling6020032.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., ... and Williams, S. E. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, **355**, 6332, eaai9214. doi: https://doi.org/10.1126/science.aai9214.

Rangelov, D., Boerger, M., Tcholtchev, N., Lämmel, P., and Hauswirth, M. 2023. Design and development of a short-term photovoltaic power output forecasting method based on Random Forest, Deep Neural Network and LSTM using readily available weather features. *IEEE access*, **11**, 41578-41595.

Renforth, P., Fischer, A. M., Koren, G., Brasseur, G., Brasseur, G., Stammer, D., ... and Stocker, V. 2025. Climate science for 2050. doi: https://doi.org/10.3389/fclim.2025.1554685.

Ries, D., Goode, K., McClernon, K., and Hillman, B. 2024. Using feature importance as exploratory data analysis tool on earth system models. *Geoscientific Model Development Discussions*, 1-35. doi:https://doi.org/10.5194/gmd-18-1041-2025.

Salman, H. A., Kalakech, A., and Steiti, A. 2024. Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 69-79. doi: https://doi.org/10.58496/BJML/2024/007.

Sarker, I. H. 2021. Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, **2**, 5, 377. doi: https://doi.org/10.1007/s42979-021-00765-8.

Shaikh, T. A., Rasool, T., Verma, P., and Mir, W. A. 2024. A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. *Annals of Operations Research*, 1-77. doi: https://doi.org/10.1007/s10479-024-06444-0.

Shetty, S., Umesh, P., and Shetty, A. 2022. Dependability of rainfall to topography and micro-climate: an observation using geographically weighted regression. *Theoretical and Applied Climatology*, **147**, 1, 217-237. doi: https://doi.org/10.1007/s00704-021-03811-w.

Sreelatha, K., and Anand Raj, P. 2021. Ranking of CMIP5-based global climate models using standard performance metrics for Telangana region in the southern part of India. *ISH Journal of Hydraulic Engineering*, **27**, sup1, 556-565. doi: https://doi.org/10.1080/09715010.2019.1634648.

Vairo, T., Pettinato, M., Reverberi, A. P., Milazzo, M. F., and Fabiano, B. 2023. An approach towards the implementation of a reliable resilience model based on machine learning. *Process Safety and Environmental Protection*, **172**, 632-641. doi: https://doi.org/10.1016/j.psep.2023.02.058.

Yadav, K., Malviya, S., and Tiwari, A. K. 2025. Improving Weather Forecasting in Remote Regions Through Machine Learning. *Atmosphere*, **16**, 5, 587. doi: https://doi.org/10.3390/atmos16050587.

Yaduvanshi, A., Nkemelang, T., Bendapudi, R., and New, M. 2021. Temperature and rainfall extremes change under current and future global warming levels across Indian climate zones. *Weather and Climate Extremes*, **31**, 100291. doi: https://doi.org/10.1016/j.wace.2020.100291.

Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., and Salehi, M. 2024. Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, **57**, 1, 1-42. doi: https://doi.org/10.1145/3691338.

Zhang, H., Liu, Y., Zhang, C., and Li, N. 2025. Machine learning methods for weather forecasting: A survey. *Atmosphere*, **16**, 1, 82. doi: https://doi.org/10.3390/atmos16010082.