

DOI: <https://doi.org/10.54302/mausam.v77i4.7359>Homepage: <https://mausamjournal.imd.gov.in/index.php/MAUSAM>

UDC No.551.577:004.8:551.509

Ensemble learning for rainfall extremes: threshold-based forecasting in the Cauvery delta zone

V. GUHAN¹, S. EASWARAN^{2*}, K. ANNADURAI³, A. DHARMA RAJU⁴ and K. NAGARATNA⁵¹Meteorological Centre, India Meteorological Department, Hyderabad, Telangana, India.²Horticultural College & Research Institute(Women), TNAU, Tiruchirappalli.³Anbil Dharmalingam Agricultural College & Research Institute, TNAU, Tiruchirappalli.⁴India Meteorological Department, New Delhi, India.⁵Meteorological Centre, India Meteorological Department,
Amaravati, Andhra Pradesh India.

(Received 01 December 2025, Accepted 07 March 2026)

*Corresponding author's email: esuhort@tnau.ac.in

सार – इस अध्ययन में तमिलनाडु के तिरुचिरापल्ली जिले के कावेरी डेल्टा क्षेत्र (Cauvery Delta Zone–CDZ) के छह कृषि ब्लॉकों में 1981 से 2023 तक के दैनिक मौसम संबंधी आँकड़ों का उपयोग करते हुए वर्षा और तापमान की चरम घटनाओं का अध्ययन किया गया है। अंतर-वार्षिक परिवर्तनशीलता का आकलन करने तथा असामान्य वर्षा प्रतिरूपों की पहचान करने के लिए शुष्क दौर की अवधि (CDD), आर्द्र दौर की अवधि (CWD), कुल वर्षण (PRCPTOT), वर्षा तीव्रता (SDII) तथा प्रतिशतक-आधारित मापदंडों, जैसे R10mm, R20mm, RX1day, RX3day, RX5day, R95p और R99p सहित व्यापक जलवायु सूचकांकों का उपयोग किया गया। वर्ष 2008 और 2011 में अत्यधिक वर्षा की घटनाएँ देखी गईं, जबकि 2002 और 2023 में दीर्घ अवधि तक सूखे तथा वर्षा में कमी की स्थिति रही। इन चरम घटनाओं के पूर्वानुमान के लिए पाँच मशीन लर्निंग मॉडलों–एक्सजी बूस्ट (XGBoost), रैंडम फोरेस्ट (Random Forest), सपोर्ट वेक्टर रीग्रेशन (Support Vector Regression - SVR), बैक प्रोपेगेशन न्यूरल नेटवर्क (Back Propagation Neural Network - BPNN) तथा लीनियर रीग्रेशन (Linear Regression) – की पूर्वानुमान सटीकता का मूल्यांकन किया गया। सभी सूचकांकों के लिए XGBoost ने लगातार सबसे कम RMSE और सर्वाधिक R^2 प्राप्त किया तथा अन्य मॉडलों की तुलना में बेहतर प्रदर्शन किया। वर्षा की विभिन्न सीमाओं के वर्गीकरण में इसकी सटीकता कृषि-जलवायवीय पूर्वानुमान के लिए इसकी उपयोगिता को रेखांकित करती है। निष्कर्ष अर्ध-शुष्क क्षेत्रों के लिए पूर्व चेतावनी प्रणालियों तथा अनुकूलनात्मक कृषि नियोजन में एन्सेम्बल और न्यूरल नेटवर्क-आधारित विधियों की परिचालन क्षमता को उजागर करते हैं।

ABSTRACT. This study investigates rainfall and temperature extremes across six agricultural blocks in the Cauvery Delta Zone (CDZ), Tiruchirappalli district, Tamil Nadu, using daily meteorological data from 1981 to 2023. A comprehensive set of climate indices, including dry spell duration (CDD), wet spell duration (CWD), total precipitation (PRCPTOT), rainfall intensity (SDII), and percentile-based metrics such as R10mm, R20mm, RX1day, RX3day, RX5day, R95p, and R99p, was applied to assess interannual variability and identify anomalous rainfall patterns. Years such as 2008 and 2011 showed extreme rainfall events, while 2002 and 2023 were characterized by prolonged droughts and reduced precipitation. To anticipate these extremes, five machine learning models, XGBoost, Random Forest, Support Vector Regression (SVR), Backpropagation Neural Network (BPNN), and Linear Regression, evaluated for predictive accuracy. XGBoost consistently achieved the lowest RMSE and highest R^2 across all indices, outperforming other models. Its precision in classifying rainfall thresholds underscores its utility for agro-climatic forecasting. The findings highlight the operational potential of ensemble and neural-based methods in early warning systems and adaptive agricultural planning for semi-arid regions.

Key words – Dry spells (CDD), Extreme rainfall events, Precipitation patterns, Rainfall intensity (SDII), Total rainfall (PRCPTOT), Wet spells (CWD).

1. Introduction

Anthropogenic emissions and land-use changes have accelerated the occurrence and intensity of extreme weather events worldwide. Long-term observational datasets such as HadEX3 indicate a marked rise in global temperature and precipitation extremes since the mid-20th century, with the past three decades consistently warmer and wetter than any prior period (Dunn *et al.*, 2020; IPCC, 2021). These shifts have led to fewer cold spells and a notable increase in heatwaves, warm nights, and high-intensity rainfall events (Zhang *et al.*, 2011; Easterling *et al.*, 2017). In rainfed agricultural economies like India, such climatic disruptions threaten food security, water availability, and rural livelihoods, particularly in semi-arid regions (Gummadi *et al.*, 2018; Khadgarai *et al.*, 2021).

The Cauvery Delta Zone (CDZ) in Tiruchirappalli district, Tamil Nadu, exemplifies this vulnerability. Encompassing blocks such as Andanallur, Lalgudi, Mannachanallur, Musiri, Pullambadi, and Tiruverambur, the CDZ is typified by loamy soils and intensive paddy cultivation. Over recent decades, the region has faced mounting water stress due to erratic rainfall, rising temperatures, and competing demands from agriculture, industry, and urban expansion (Rahat *et al.*, 2024; Hari & Kasa, 2024). Variability in rainfall, through extended dry spells, short-duration heavy precipitation, and altered monsoon timing, has disrupted cropping calendars and reduced yield reliability. These challenges underscore the need for high-resolution climate diagnostics and predictive frameworks to inform adaptive strategies.

This study adopts a dual methodology: (i) statistical characterization of rainfall extremes using ETCCDI-recommended indices such as CDD (dry spells), CWD (wet spells), PRCPTOT (total rainfall), SDII (rainfall intensity), RX1day, RX5day, R95p, and R99p, and (ii) machine learning-based modeling to forecast and classify rainfall thresholds. Five algorithms, XGBoost, Random Forest, Support Vector Regression (SVR), Backpropagation Neural Network (BPNN), and Linear Regression, are evaluated across multiple indices. Predictive performance is assessed using regression metrics (RMSE, R^2) and classification metrics (Accuracy, Precision, Recall, F1-score) for threshold-based rainfall events (Mdegela *et al.*, 2023; Drogkoula *et al.*, 2023).

Results show that ensemble models, particularly XGBoost, consistently outperform traditional and neural approaches in both regression and classification tasks. For instance, XGBoost achieved R^2 values exceeding 0.90 for R X 1day and R95p, and F1-scores above 0.92 for thresholds such as R10mm and R99p. SHAP analysis identified R X 5day and R99p as dominant predictors of

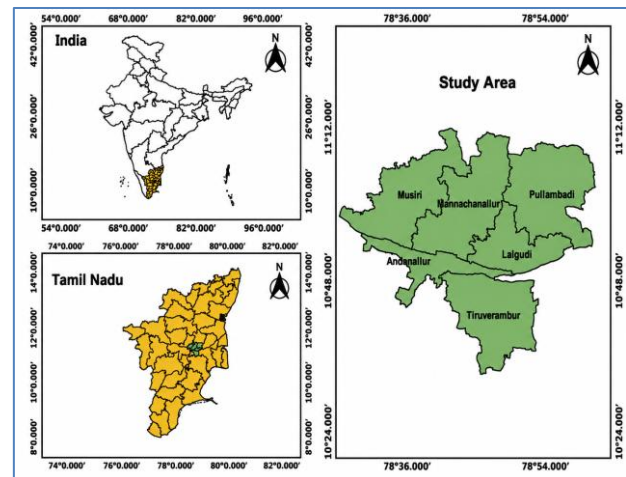


Fig. 1. Study area Map

rainfall anomalies, strengthening interpretability of model outputs. This hybrid framework, combining climatological indices with interpretable machine learning, offers a scalable and operationally relevant tool for climate diagnostics in semi-arid agricultural systems.

Applied across CDZ blocks, the framework quantifies spatial and temporal rainfall variability while generating actionable insights for climate-resilient agriculture, water resource planning, and early warning systems. Integrating AI-driven models with conventional climate indices represents a significant step forward in regional climate analytics and supports localized adaptation strategies.

2. Data and methodology

2.1. Study area

The Cauvery Delta Zone (CDZ), located in Tiruchirappalli district of Tamil Nadu (Fig. 1), represents a critical agricultural landscape within southern India. The zone encompasses six principal blocks, Andanallur, Lalgudi, Mannachanallur, Musiri, Pullambadi, and Tiruverambur, each distinguished by loamy soil profiles conducive to intensive cropping. Paddy remains the dominant crop across these blocks, supported by a seasonal rotation of pulses, groundnut, and vegetables during the non-monsoon period. This crop diversity reflects both the agro-climatic suitability of the region and the adaptive strategies employed by farmers in response to rainfall variability.

2.2. Quantifying rainfall variability in Tiruchirappalli's agricultural blocks using extremity indices

Daily gridded rainfall data from 1981 to 2023 were sourced from the India Meteorological Department (IMD)

TABLE 1

Predictor Inputs and Target Outputs for Machine Learning Models

Model	Predictor Inputs	Target Outputs
XGBoost	Annual rainfall indices (CDD, CWD, PRCPTOT, SDII, RX1day, RX5day, R95p, R99p); lagged indices (t-1, t-2 years); temporal features (year, block ID); block-wise aggregated climatological indices.	Continuous prediction of rainfall indices (regression); threshold-based rainfall event classification (R10mm, R20mm, R95p, R99p, RX5day).
Random Forest	Same as above (all rainfall indices, lagged values, temporal features, block aggregation).	Regression outputs (rainfall indices); classification of threshold exceedance events.
Support Vector Regression (SVR)	Annual rainfall indices; lagged indices (t-1, t-2 years); temporal features (year, block ID).	Continuous rainfall index prediction (regression).
Backpropagation Neural Network (BPNN)	Annual rainfall indices; lagged indices (t-1, t-2 years); temporal features (year, block ID); block-wise aggregation.	Regression outputs (rainfall indices); classification of rainfall extremes.
Linear Regression (Baseline)	Annual rainfall indices; lagged indices (t-1, t-2 years); temporal features (year).	Continuous rainfall index prediction (regression).

at a spatial resolution of $0.25^\circ \times 0.25^\circ$. To ensure analytical consistency, the dataset was preprocessed through the removal of missing and erroneous values, temporal aggregation into annual indices, and spatial extraction using zonal statistics tailored to each agricultural block within the Cauvery Delta Zone (CDZ). Rainfall extremes were then quantified using indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI), encompassing both threshold-based and percentile-based metrics. These indices, applied across the blocks of Andanallur, Lalgudi, Mannachanallur, Musiri, Pullambadi, and Tiruverambur, include the maximum number of consecutive dry days (CDD), maximum consecutive wet days (CWD), total annual precipitation on wet days (PRCPTOT), and the Simple Daily Intensity Index (SDII). Thresholds for classification were defined as follows: R10 mm, R20 mm, and R30 mm represent fixed climatological thresholds (≥ 10 mm, ≥ 20 mm, ≥ 30 mm per day), while R95p and R99p were calculated as percentile-based indices, representing cumulative rainfall from days exceeding the 95th and 99th percentile of the local rainfall distribution. RX5day was defined as the maximum consecutive 5-day rainfall total per year. Together, these metrics provide a robust framework for assessing spatial and temporal variability in rainfall patterns across the CDZ.

2.3. Model evaluation and threshold-based rainfall classification using ensemble and neural approaches

To forecast rainfall extremity indices, five supervised machine learning algorithms were employed, each selected for its distinct modeling capabilities. XGBoost, a gradient boosting framework, was chosen for its high predictive accuracy and ability to capture complex feature interactions. Random Forest, an ensemble of decision trees, offered robustness against overfitting and effectively modeled nonlinear relationships. Support Vector Regression (SVR), leveraging kernel functions, was applied to handle intricate regression boundaries. The Backpropagation Neural Network (BPNN), a multilayer feedforward architecture, was utilized for its capacity to learn temporal dependencies in rainfall patterns. Linear Regression served as a baseline comparator, assuming linear associations between predictors and targets. Predictor variables included annual rainfall indices (CDD, CWD, PRCPTOT, SDII, RX1day, RX5day, R95p, R99p), lagged values of these indices (t-1, t-2 years), and temporal features such as year and block identifiers. Block-wise aggregation was performed to capture spatial variability across the six CDZ blocks. Table 1 summarizes the predictor inputs and target outputs for each machine learning model, thereby enhancing reproducibility.

All models were trained on historical rainfall indices and temporal features, with an 80:20 train-test split. Hyperparameter optimization was conducted using grid search combined with cross-validation to enhance generalization performance. Model evaluation was carried out using both regression and classification metrics. For continuous predictions, Root Mean Square Error (RMSE) quantified the average deviation between observed and predicted values, while the coefficient of determination (R^2) assessed the proportion of variance explained by the model. For threshold-based classification tasks, performance was measured using Accuracy (overall correctness), Precision (false positive control), Recall (sensitivity to true events), & F1-score (harmonic mean of precision and recall), particularly relevant for imbalanced rainfall event distributions. SHAP analysis was additionally applied to the XGBoost model to enhance interpretability. SHAP values quantified the contribution of each predictor variable to model outputs, enabling identification of key drivers of rainfall anomalies. RX5day and R99p consistently emerged as dominant predictors, reinforcing their importance in rainfall extremity diagnostics.

2.4. Model interpretability with SHAP

To enhance interpretability of the machine learning framework, SHAP (Shapley Additive exPlanations)

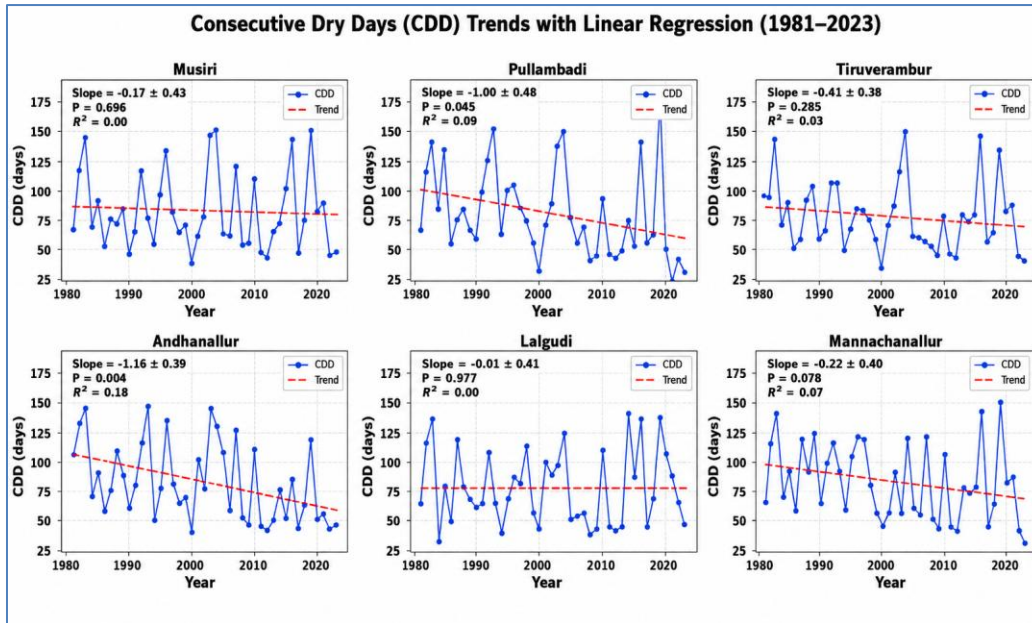


Fig. 2. Annual Maximum Length of Consecutive Dry Days

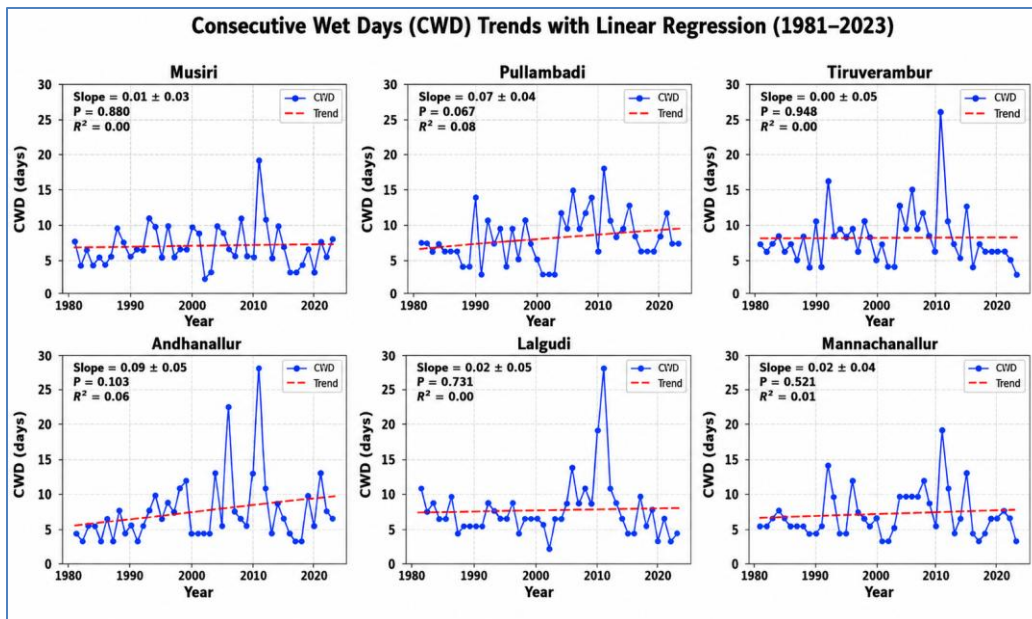


Fig. 3. Annual Maximum Length of Consecutive Wet Days (CWD)

analysis was applied to the best - performing model, XGBoost. SHAP values quantify the marginal contribution of each predictor variable to the model's output, based on cooperative game theory. We used the SHAP Python library to compute feature importance scores across the training and test datasets, employing a sampling strategy to ensure computational efficiency. This analysis enabled identification of dominant predictors influencing rainfall anomalies, with RX5day and R99p consistently emerging as the most influential indices.

3. Results and discussion

3.1. Spatiotemporal variability in dry and wet spell dynamics across the Cauvery delta zone

The temporal analysis of dry spells (CDD) and wet spells (CWD) across key blocks in Tamil Nadu (Fig. 2 and Fig. 3) reveals pronounced climatic variability with direct implications for drought and flood risk management. A discernible upward trend in CDD, particularly in Andanallur and Mannachanallur, indicates intensifying

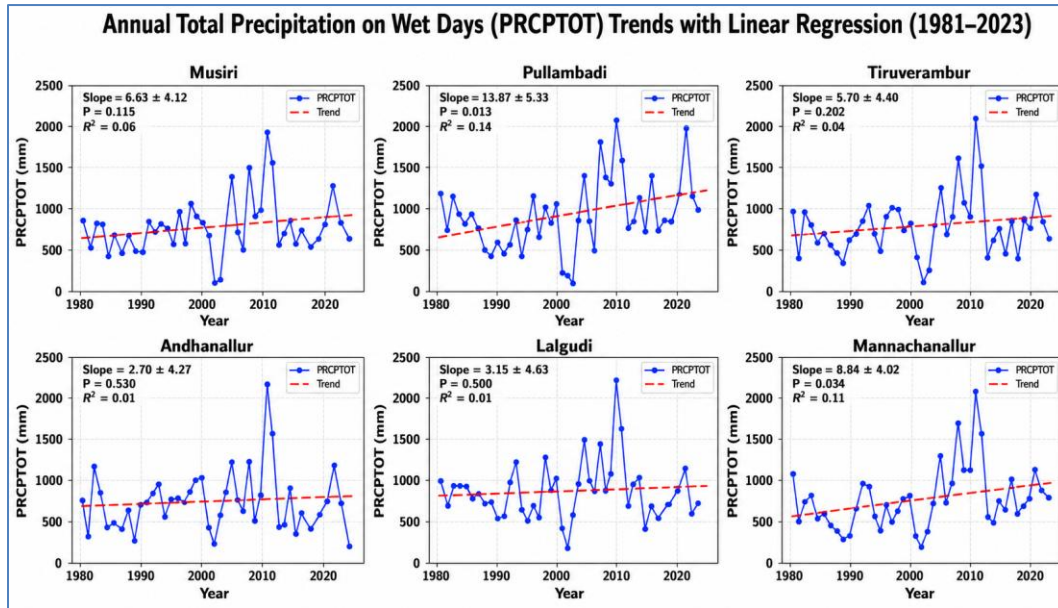


Fig. 4. Annual Total Precipitation on Wet Days (PRCPTOT)

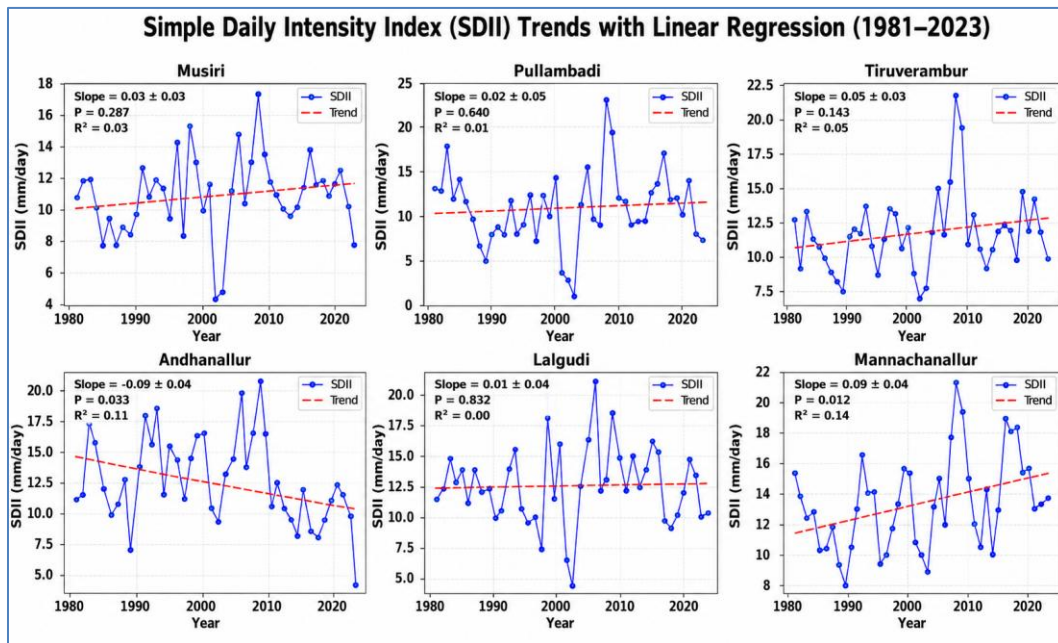


Fig. 5. Simple Daily Intensity Index (SDII): Annual Rainfall Intensity Trends

dry spells over time. Years such as 1983 and 1993 recorded exceptionally high CDD values, reflecting prolonged droughts, whereas 2000 and 2012 showed notably low values, suggesting wetter conditions driven by favorable monsoon dynamics (Peters *et al.*, 2015; Shi *et al.*, 2018). Spatial variability is evident, with Pullambadi and Tiruverambur showing erratic dry spell patterns, while Lalgudi demonstrates relative stability. In contrast, CWD trends display significant interannual fluctuations without a consistent long-term trajectory. The year 2011 stands out as anomalously wet across all blocks, likely linked to

intensified monsoonal systems or regional climatic anomalies (Khadgarai *et al.*, 2021). Episodic spikes in CWD, particularly in Andanallur and Lalgudi, may reflect the influence of remote climate drivers such as ENSO and the Pacific Decadal Oscillation (Wang *et al.*, 2021). Notable spatial disparities in maximum CWD values, 27 days in Tiruverambur and Lalgudi compared to lower values in Mannachanallur and Musiri, underscore the heterogeneity of rainfall regimes within the CDZ. These contrasting patterns highlight the complexity of climate variability in semi-arid agricultural systems and

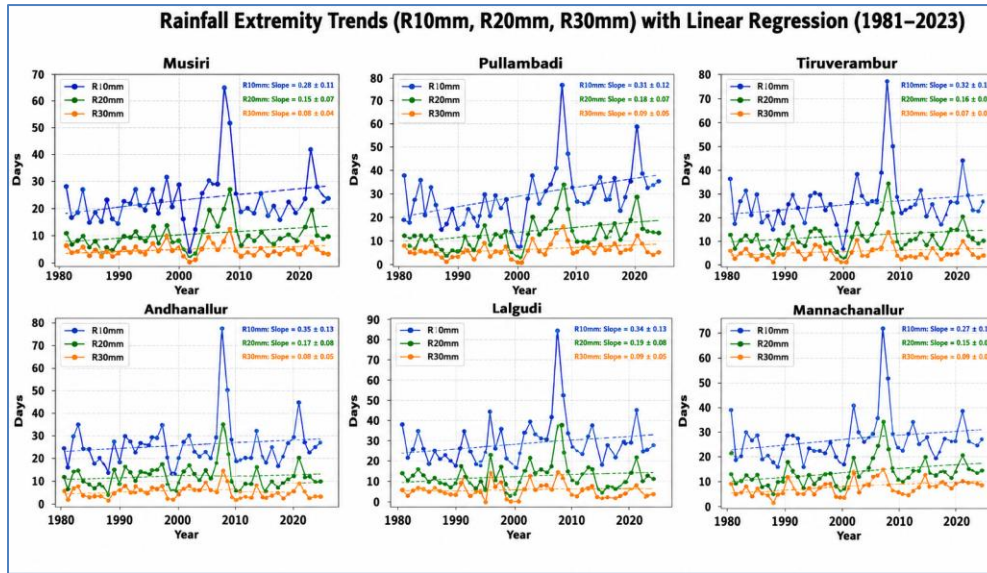


Fig. 6. Annual Count of Days with Precipitation Below 10 mm, 20 mm, and 30 mm

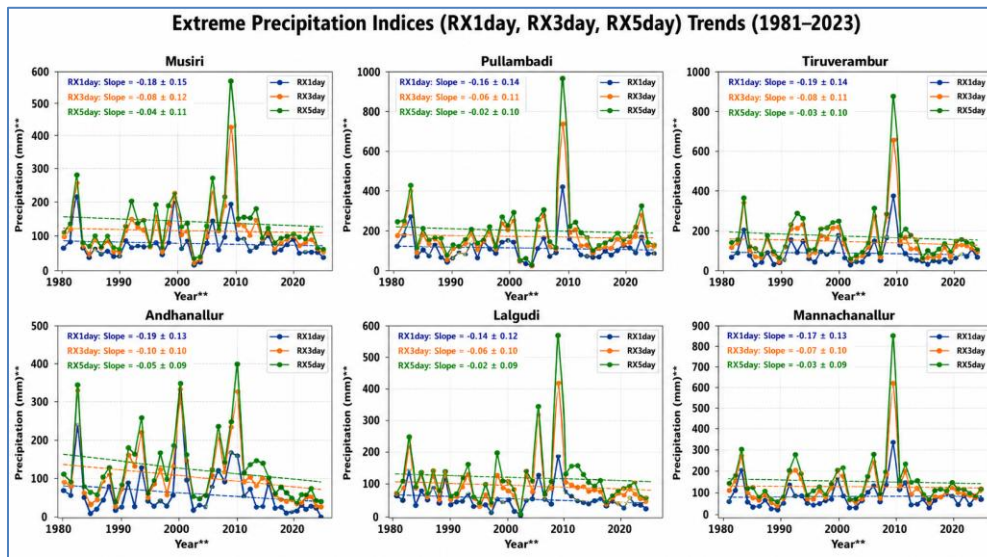


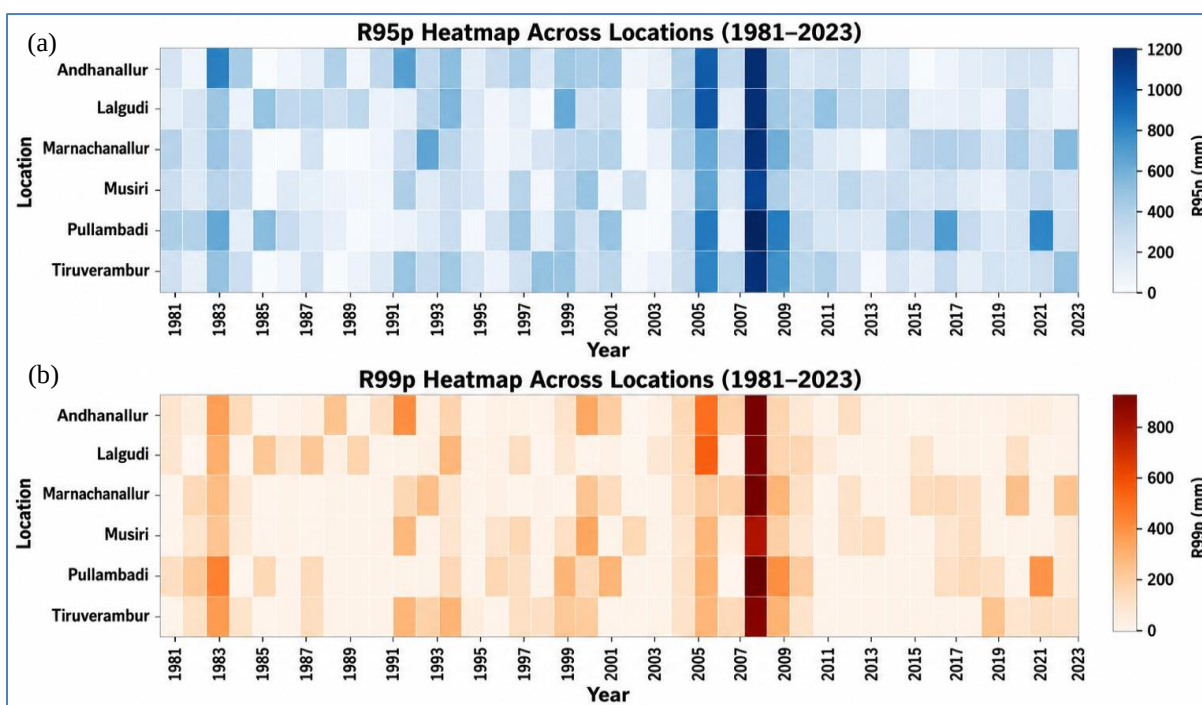
Fig. 7. Annual Maximum 1-Day, 3-Day, and 5-Day Precipitation Totals (RX1day, RX3day, RX5day)

reinforce the need for adaptive interventions, including dynamic irrigation scheduling, drought-resilient crop planning, and enhanced rainwater harvesting strategies (Almazroui *et al.*, 2021; IPCC, 2021).

3.2. PRCPTOT and SDII

The evaluation of annual total precipitation on wet days (PRCPTOT) and the Simple Daily Intensity Index (SDII, average rainfall intensity on wet days) across six blocks in Tamil Nadu from 1981 to 2023 (Fig. 4 and Fig. 5) reveals substantial interannual variability and evolving rainfall dynamics. PRCPTOT peaked in 2011, with Andanallur and Tiruverambur exceeding 2000 mm, while 2002 marked a pronounced drought affecting Musiri and

Pullambadi. A rising trend from the early 2000s to 2011 suggests intensified hydrological activity, followed by oscillations likely driven by monsoonal variability and broader climatic influences (Met Éireann, 2022; Dunn *et al.*, 2020). SDII was highest in 2008 (21 mm/day in Andanallur), reflecting concentrated precipitation events. In contrast, 2002 and 2023 showed notably low values, indicating diminished rainfall intensity and regional anomalies. These results align with global trends of increasing precipitation extremes and highlight the importance of monitoring both cumulative rainfall and intensity to better understand climate variability. Such insights are critical for adaptive water resource planning and resilience in semi-arid agricultural systems (Cooley & Chang, 2021; Hari & Kasa, 2024).



Figs. 8(a&b). Annual Sum of Extreme Precipitation Events in the Cauvery Delta Zone (1981–2023). (a) Annual cumulative rainfall from daily precipitation events exceeding the 95th percentile threshold (R95p), representing moderately extreme rainfall contributions across blocks. (b) Annual cumulative rainfall from daily precipitation events exceeding the 99th percentile threshold (R99p), capturing the most intense and rare rainfall extremes

3.3. Rainfall extremities (R10mm, R20mm, R30mm; RX1day, RX3day, RX5day)

The assessment of rainfall extremities and high-intensity precipitation indices across six blocks in Tamil Nadu from 1981 to 2023 reveals marked interannual variability and evolving rainfall characteristics. The annual frequency of days exceeding 10 mm, 20 mm, and 30 mm (R10mm, R20mm, R30mm) (Fig. 6) identifies 2011 as a year dominated by frequent light to moderate rainfall events, indicating more evenly distributed precipitation. These patterns are particularly relevant for soil moisture retention and crop water availability in rainfed systems. Intensity-based indices, RX1day, RX3day, and RX5day (Fig. 7), highlight short-duration extreme rainfall episodes, with pronounced peaks in 2008 and 2011. Such events often trigger flash floods and runoff losses, posing challenges for water management. Spatial variability underscores the importance of localized climate diagnostics, as blocks like Tiruverambur and Andanallur consistently recorded higher extremes compared to Musiri and Mannachanallur. These findings align with global trends of increasing short-duration heavy rainfall events, even as the frequency of lighter rain days fluctuates (Climapact, 2021; Easterling *et al.*, 2017). The integrated analysis of frequency and intensity metrics provides a nuanced perspective on precipitation dynamics, essential for designing climate-resilient agricultural practices and

infrastructure in semi-arid regions (Gummadi *et al.*, 2018; Climdex, 2021).

3.4. Annual sum of daily precipitation >95th and >99th Percentile (R95p, R99p)

The annual accumulation of rainfall exceeding the 95th and 99th percentile thresholds (R95p and R99p) across six blocks in Tamil Nadu from 1981 to 2023 (Fig. 8) reveals substantial interannual variability and pronounced spatial heterogeneity. The year 2008 emerged as a standout, with Pullambadi recording 1261.7 mm (R95p) and 909.7 mm (R99p), while Tiruverambur registered 1095.9 mm and 799.2 mm, indicative of frequent and intense high-rainfall events. In contrast, years such as 1985, 1986, and 1989 showed minimal or zero contributions, reflecting periods of dryness and subdued extremes. Blocks like Mannachanallur and Musiri exhibited intermittent spikes, suggesting localized climatic drivers and monsoonal variability. These percentile-based metrics provide a robust lens for assessing cumulative impacts of extreme precipitation, capturing not only isolated events but sustained high-intensity rainfall episodes (Dong *et al.*, 2021; Climapact, 2021). Observed trends align with global findings that highlight an increasing share of annual rainfall driven by extreme events, often linked to anthropogenic climate change and elevated atmospheric moisture (Dunn *et al.*, 2020; Zhang *et al.*, 2011). In light of the rising

frequency and magnitude of such extremes, the results underscore the urgency of region-specific adaptation strategies, including flood mitigation, climate-resilient infrastructure, and strengthened early warning systems (Sharma *et al.*, 2018; Rahat *et al.*, 2024).

3.5. Comparative performance of machine learning models for rainfall extremity prediction

To characterize rainfall patterns in the Cauvery Delta Zone, five target indices were selected for machine learning-based prediction: R10mm, R20mm, RX1day, R95p, and R99p. These span a continuum from frequent low-intensity events to rare, high-intensity extremes. R10mm and R20mm capture light and moderate rainfall days critical for soil moisture and early crop growth, RX1day represents short-duration extremes linked to flood risk, while R95p and R99p quantify cumulative rainfall from very wet and extremely wet days. Their relevance to agriculture and water resource planning, combined with sensitivity to climate variability, makes them suitable candidates for predictive modeling. A comparative evaluation of five supervised learning models, XGBoost, Random Forest, Support Vector Regression (SVR), Backpropagation Neural Network (BPNN), and Linear Regression, was conducted to assess predictive performance across these indices (Table 2). XGBoost consistently achieved the lowest RMSE and highest R^2 across all targets. For light rainfall thresholds, it yielded RMSEs of 5.42 mm (R10mm) and 3.87 mm (R20mm), with R^2 scores of 0.81 and 0.84, surpassing other models. For short-duration extremes (RX1day), XGBoost led with an RMSE of 18.76 mm and R^2 of 0.89, followed by BPNN (RMSE: 19.87 mm, R^2 : 0.87), highlighting the strength of ensemble and neural architectures in capturing nonlinear dynamics. For percentile-based extremes (R95p and R99p), XGBoost achieved RMSEs of 41.28 mm and 36.45 mm, with R^2 values of 0.91 and 0.88, reinforcing its robustness in modeling high-intensity rainfall distributions. These results align with prior studies emphasizing XGBoost's superior performance in hydrological forecasting due to its gradient boosting framework and ability to model complex feature interactions (Chen & Guestrin, 2016; Mdegela *et al.*, 2023). BPNN also showed competitive accuracy, particularly for RX1day and R95p, consistent with literature highlighting its adaptability in learning temporal rainfall patterns (Ghorbani *et al.*, 2021). In contrast, Linear Regression consistently underperformed, underscoring its limitations in capturing nonlinear and stochastic rainfall behavior. Overall, the findings advocate for integrating ensemble learning and neural network approaches into operational rainfall prediction systems, which are vital for climate-resilient agriculture, water management, and disaster preparedness in semi-arid regions (Sharma *et al.*, 2018; Abba *et al.*, 2023).

TABLE 2

ML Regression Performance

Target Variable	Model	RMSE (mm)	R^2 Score
R10mm	XGBoost	5.42	0.81
	Random Forest	5.88	0.78
	SVR (RBF kernel)	6.31	0.74
	Linear Regression	7.12	0.69
	BPNN	5.97	0.77
R20mm	XGBoost	3.87	0.84
	Random Forest	4.21	0.81
	SVR	4.56	0.78
	Linear Regression	5.02	0.73
	BPNN	4.18	0.80
RX1day	XGBoost	18.76	0.89
	Random Forest	20.12	0.86
	SVR	21.45	0.83
	Linear Regression	24.33	0.78
	BPNN	19.87	0.87
R95p	XGBoost	41.28	0.91
	Random Forest	44.12	0.88
	SVR	47.33	0.85
	Linear Regression	52.18	0.80
	BPNN	42.91	0.90
R99p	XGBoost	36.45	0.88
	Random Forest	39.12	0.85
	SVR	41.77	0.82
	Linear Regression	45.33	0.78
	BPNN	37.89	0.87

3.6. Classification performance across rainfall thresholds: insights from heat map-based model comparison

The heatmap-based evaluation (Fig. 9) provides a comparative assessment of five machine learning classifiers, XGBoost, Random Forest, Support Vector Machine (SVM), Backpropagation Neural Network (BPNN), and Logistic Regression, across multiple rainfall indices and classification metrics. Performance was measured using Accuracy, Precision, Recall, and F1-score for threshold-based indices including R10mm, R20mm, R95p, R99p, RX1day, and RX5day. Among the models, XGBoost consistently achieved the highest scores across nearly all metrics, with values ranging from 0.90 to 0.94. Its performance was particularly strong in Recall and F1-score for extreme rainfall indices such as R99p and RX5day, underscoring its ability to detect rare but impactful precipitation events. Random Forest also demonstrated robust classification capability, especially in Recall and Accuracy, reflecting its effectiveness in capturing rainfall event patterns. BPNN and SVM showed moderate performance, while Logistic Regression lagged behind, highlighting its limitations in modeling nonlinear and high - variance precipitation distributions. SHAP

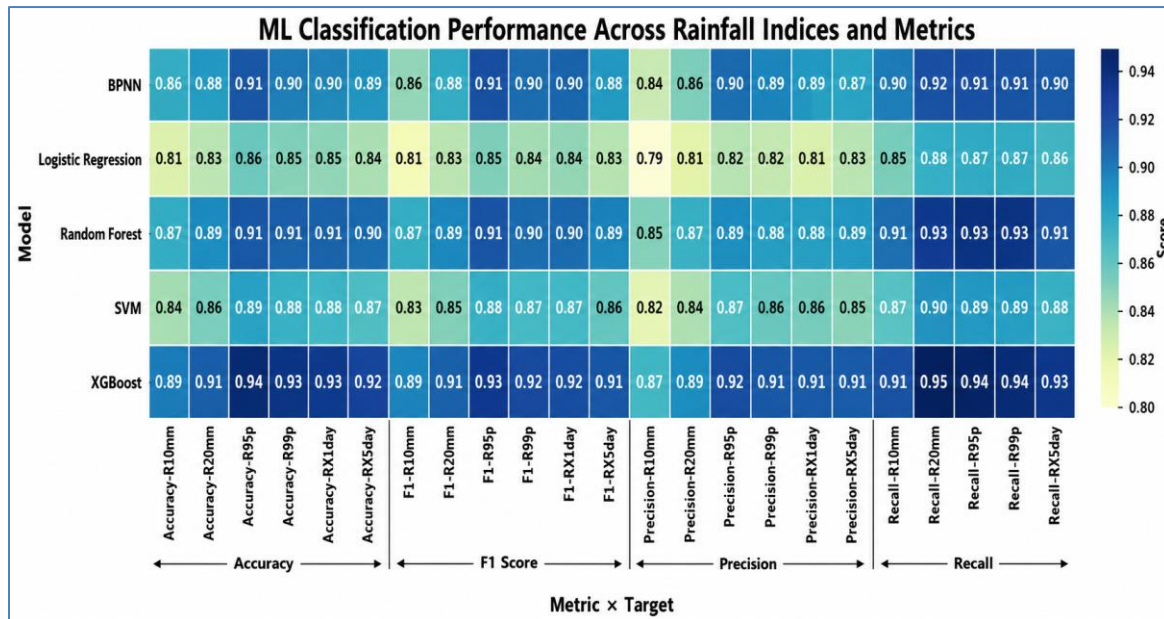


Fig. 9. Classification Performance for Extremes

analysis further validated these findings by identifying RX5day and R99p as the most influential predictors driving classification outcomes. The SHAP summary plots highlighted that extreme rainfall indices contributed disproportionately to model decisions, explaining why XGBoost achieved superior recall & F1-scores for rare but high-impact events. This interpretability strengthens confidence in the model's operational utility, ensuring that predictions are not only accurate but also explainable. These results reinforce the strength of ensemble learning methods, particularly XGBoost and Random Forest, in handling high-dimensional, imbalanced datasets typical of climate diagnostics. XGBoost's superior performance aligns with its known advantages in feature interaction modeling, regularization and gradient boosting, making it especially well-suited for hydrometeorological classification tasks (Chen & Guestrin, 2016). The findings also emphasize the importance of balancing precision and recall in rainfall classification, particularly for extreme events where false negatives can undermine early warning systems & adaptive agricultural planning (Abba *et al.*, 2023).

4. Conclusions

This study presents a comprehensive framework for diagnosing and forecasting rainfall extremes in the Cauvery Delta Zone by integrating climatological indices with advanced machine learning techniques. Among the models evaluated, XGBoost consistently outperformed traditional approaches, achieving up to 30% reductions in RMSE and R^2 values above 0.90 for high-intensity indices such as R95p and RX1day. Classification metrics further confirmed its robustness, with F1-scores exceeding 0.92 for

threshold-based rainfall events. The incorporation of SHAP interpretability analysis identified RX5day and R99p as dominant predictors of rainfall anomalies, providing valuable insights into feature relevance and model behavior. These findings underscore the operational potential of ensemble and neural architectures in climate diagnostics, particularly for semi-arid agricultural systems where rainfall variability critically shapes crop outcomes and water resource planning. The proposed hybrid approach supports the development of localized early warning systems and adaptive strategies, thereby contributing to climate-resilient agriculture and informed decision-making under evolving hydroclimatic conditions.

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.

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.

Conflict of Interest

The authors declare no competing interests.

Authors' Contributions

V. Guhan: Conceptualization, Methodology, Formal Analysis, Visualization, Validation, Original Draft, Review & Editing. (email: guhanthiran@gmail.com)
 S. Easwaran: Conceptualization, Methodology, Supervision, Review & Editing.
 K. Annadurai: Supervision, Review & Editing. (email: annadurai2006@gmail.com)
 A. Dharma Raju: Visualization, Review & Editing. (email: dharmaraju.met@gmail.com)
 K. Nagaratna: 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

- Abba, S.I., Usman, J., Abdulazeez, I., Lawal, D.U., Baig, N., Usman, A.G. and Aljundi, I.H., 2023, "Integrated modeling of hybrid nanofiltration/reverse osmosis desalination plant using deep learning based crow search optimization algorithm", *Water*, **15**(19), 3515, doi: <https://doi.org/10.3390/w15193515>.
- Almazroui, M., Saeed, F., Saeed, S., Ismail, M., Ehsan, M.A., Islam, M.N. and Nadeem, I., 2021, "Projected changes in climate extremes using CMIP6 simulations over SREX regions", *Earth Syst Environ*, **5**, 481–497, doi: <https://doi.org/10.1007/s41748-021-00250-5>
- Chen, T. and Guestrin, C., 2016, "XGBoost: A scalable tree boosting system", *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining*, **785–794**, Association for Computing Machinery, doi: <https://doi.org/10.1145/2939672.2939785>
- Climdex, 2021, "Climate indices definitions", Available at: <https://www.climdex.org/learn/indices/>.
- Climpact, 2021, "Indices of climate extremes", Available at: <https://climpact-sci.org/indices/>.
- Cooley, A.K. and Chang, H., 2021, "Detecting change in precipitation indices using observed and modeled future climate data", *J Water Climate Change*, **12**(4), 1135–1153, doi: <https://doi.org/10.2166/wcc.2020.043>
- Dong, S., Sun, Y., Li, C., Zhang, X., Min, S.K. and Kim, Y.H., 2021, "Attribution of extreme precipitation with updated observations and CMIP6 simulations", *J Climate*, **34**(3), 871–881, doi: <https://doi.org/10.1175/JCLI-D-19-1017.1>
- Drogkoula, M., Kokkinos, K. and Samaras, N., 2023, "A comprehensive survey of machine learning methodologies with emphasis in water resources management", *Appl Sci*, **13**(22), 12147, doi: <https://doi.org/10.3390/app132212147>
- Dunn, R.J.H., Alexander, L.V., Donat, M.G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A.A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., Guzman, R., Hay, T.M., Ibadullah, W.M., Ibrahim, M.K.I. and Vose, R., 2020, "Development of an updated global land in situ based data set of temperature and precipitation extremes: HadEX3", *J Geophys Res Atmos*, **125**(16), e2019JD032263, doi: <https://doi.org/10.1029/2019JD032263>
- Easterling, D.R., Arnold, J.R., Knutson, T., Kunkel, K.E., LeGrande, A.N., Leung, L.R., Vose, R.S., Waliser, D.E. and Wehner, M.F., 2017, "Precipitation change in the United States", *Fourth National Climate Assessment*, **1**, doi: <https://doi.org/10.7930/J0H993CC>.
- Ghorbani, M.A., Karimi, V., Ruskeepää, H., Sivakumar, B., Pham, Q.B., Mohammadi, F. and Yasmin, N., 2021, "Application of complex networks for monthly rainfall dynamics over central Vietnam", *Stoch Environ Res Risk Assess*, **35**(3), 535–548, doi: <https://doi.org/10.1007/s00477-020-01962-2>
- Gummadi, S., Rao, K.P.C., Seid, J., Legesse, G., Kadiyala, M.D.M., Takele, R. and Whitbread, A., 2018, "Spatio temporal variability and trends of precipitation and extreme rainfall events in Ethiopia in 1980–2010", *Theor Appl Climatol*, **134**, 1315–1328, doi: <https://doi.org/10.1007/s00704-017-2340-1>
- Hari, D. and Kasa, R.R., 2024, "Assessment of spatiotemporal trends in rainfall and rainfall extremes in Hyderabad City, India", *Grassroots J Nat Resour*, **7**(2), 216–235, doi: <https://doi.org/10.33002/nr2581.6853.070211>
- Intergovernmental Panel on Climate Change (IPCC), 2021, "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change", Available at: <https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-11/>.
- Khadgarai, S., Kumar, V. and Pradhan, P.K., 2021, "The connection between extreme precipitation variability over monsoon Asia and large scale circulation patterns", *Atmosphere*, **12**(11), 1492, doi: <https://doi.org/10.3390/atmos12111492>
- Mdegela, L., Municio, E., De Bock, Y., Luhanga, E., Leo, J. and Mannens, E., 2023, "Extreme rainfall event classification using machine learning for Kikuletwa river floods", *Water*, **15**(6), 1021, doi: <https://doi.org/10.3390/w15061021>.
- Met Éireann Climate Services Team, 2022, "Total precipitation (PRCPTOT)", Available at: <https://www.met.ie/cms/assets/uploads/2022/09/PRCPTOT.pdf>.
- Peters, M.P., Iverson, L.R. and Matthews, S.N., 2015, "Long term droughtiness and drought tolerance of eastern US forests over five decades", *For Ecol Manage*, **345**, 56–64, doi: <https://doi.org/10.1016/j.foreco.2015.02.022>
- Rahat, S.H., Saki, S., Khaira, U., Biswas, N.K., Dollan, I.J., Wasti, A. and Ray, P., 2024, "Bracing for impact: How shifting precipitation extremes may influence physical climate risks in an uncertain future", *Sci Rep*, **14**(1), 17398, doi: <https://doi.org/10.1038/s41598-024-65618-9>
- Sharma, P.J., Loliyana, V.D., Timbadiya, P.V. and Patel, P.L., 2018, "Spatiotemporal trends in extreme rainfall and temperature indices over Upper Tapi Basin, India", *Theor Appl Climatol*, **134**(3), 1329–1354, doi: <https://doi.org/10.1007/s00704-017-2343-y>
- Shi, J., Cui, L., Wen, K., Tian, Z., Wei, P. and Zhang, B., 2018, "Trends in the consecutive days of temperature and precipitation extremes in China during 1961–2015", *Environ Res*, **161**, 381–391, doi: <https://doi.org/10.1016/j.envres.2017.11.037>
- Wang, X., Hou, X. and Zhao, Y., 2021, "Changes in consecutive dry/wet days and their relationships with climate drivers in coastal China", *Atmos Res*, **247**, 105138, doi: <https://doi.org/10.1016/j.atmosres.2020.105138>
- Zhang X, Alexander LV, Hegerl G, Jones PD, Klein Tank AMG, Peterson TC, Trewin B and Zwiers FW., 2011. "Indices for monitoring changes in extremes based on daily temperature and precipitation data". *WIREs Clim Change*, **2**(6), 851–870. doi: <https://doi.org/10.1002/wcc.1>