National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
Journal of Advanced Research in Embedded System
Volume 13, Issue 1&2 - 2026

Review Article

Spatiotemporal Prediction of Cloudburst Vulnerability Zones in Uttarakhand Using ERA5 Reanalysis (1960–2024)

Ankit Kumar[1], Prashant Kumar[2]

[1]Mathematics and Computing National Institute of Technology Delhi, India
[2]Department of Applied Sciences National Institute of Technology Delhi, India
DOI: https://doi.org/10.24321/2395.3802.202611

I N F O

Corresponding Author:
Prashant Kumar, Department of Applied Sciences National Institute of Technology Delhi, India
E-mail Id:
prashantkumar@nitdelhi.ac.in
How to cite this article:
Kumar A, Kumar P. Spatiotemporal Prediction of Cloudburst Vulnerability Zones in Uttarakhand Using ERA5 Reanalysis (1960–2024). *J Adv Res Embed Sys* 2026; 13(1&2): 62-67.

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-28

A B S T R A C T

Cloudbursts are sudden, localised downpours that frequently trigger flash floods and slope failures in Himalayan terrain. Uttarakhand is highly vulnerable to such events due to its fragile topography and concentrated rainfall during the monsoon season. This study analyzes ERA5 reanalysis data (1960–2024) to characterize rainfall extremes and assess predictive models for cloudburst-prone conditions. Derived features such as rolling averages, daily tendencies, and seasonal encodings are combined with machine learning and deep learning algorithms. Across models, the Long Short-Term Memory (LSTM) network achieved the best performance, reducing MAE by 27% relative to the MLP baseline (MAE: 4.8 vs. 6.7 mm day^{-1}) and attaining RMSE =

7.2 mm day^{-1}, R^2 = 0.82. For threshold exceedance forecasting (100 mm day^{-1} proxy), LSTM reached AUC-PR = 0.84, F1 = 0.84, and CSI = 0.68. Probabilistic risk maps, produced via an ensemble with Monte Carlo dropout, highlight persistent hotspots under quantified uncertainty. The workflow provides insights that can aid disaster preparedness and climate resilience efforts in the region, while explicitly acknowledging resolution and temporal- scale limitations of ERA5 for sub-daily cloudburst detection.

Keywords: Cloudburst Prediction, Uttarakhand, ERA5, Rainfall Extremes, Machine Learning, Deep Learning

Introduction

Uttarakhand, located in the central Himalayas, experiences frequent extreme rainfall events that cause severe impacts such as flash floods, landslides, and infrastructure damage. Cloud- bursts are particularly destructive because they release large amounts of precipitation over very short time spans and small spatial footprints. Traditional forecasting systems struggle with such localised extremes, while observational networks are sparse in mountainous regions. Reanalysis datasets like ERA5 provide consistent atmospheric records across decades and, when combined with data-driven models, offer opportunities to analyse precursors of cloudbursts and predict high-risk areas.[1] This study integrates long-term reanalysis data with modern predictive methods to identify and map cloudburst- prone regions in Uttarakhand and discusses the implications and limitations of reanalysis scale for sub-daily phenomena.

Literature Survey

Rainfall forecasting in India has historically relied on statistical approaches such as regression models and time-series methods, which capture seasonal averages but often fail to reproduce abrupt extremes. Machine learning methods such as random forests and support vector machines have been applied to precipitation prediction, offering improved handling of nonlinear relationships. Deep learning architectures, espe- cially LSTM and GRU, are designed for sequential data and can capture temporal dependencies in rainfall. Convolutional neural networks (CNNs) detect short-term variability, while boosting methods like XGBoost provide strong results on structured tabular features. Despite these advances, focused applications for cloudburst prediction in the Himalayan region remain scarce.[2] This study contributes by combining multi- decadal ERA5 data with multiple algorithmic approaches, benchmarking their performance with rare-event metrics, link- ing results to spatial risk mapping, and providing uncertainty- aware outputs.

Data and Preprocessing

Study Region

Uttarakhand, located in the central Himalayan belt of India, is characterized by a highly complex topography ranging from low-lying plains to rugged mountainous terrain. Elevations vary from around 300 m in the southern foothills to peaks that rise above 7000 m, producing strong spatial gradients in climate and hydrology. Such steep altitudinal variations directly influence precipitation distribution, cloud formation, and runoff generation. The state is divided into thirteen ad- ministrative districts, each with distinct climatic regimes that make localised analysis essential.[3]

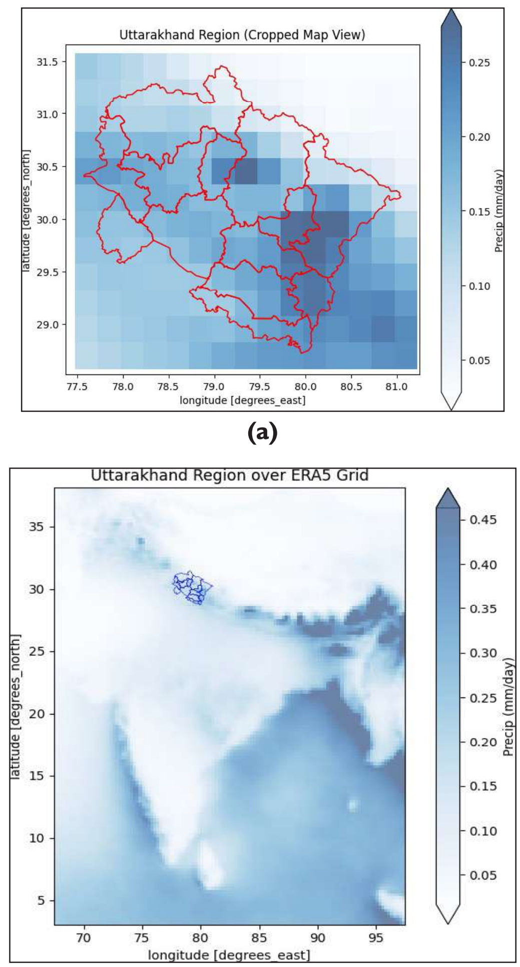
For this study, the ERA5 reanalysis dataset at a spatial resolution of 0.25° was clipped to Uttarakhand's boundaries to ensure coverage consistent with district-level administrative units. Figure 1 illustrates the study domain. Panel (a) shows the administrative map with district divisions, while panel (b) overlays the ERA5 grid, highlighting how the dataset cap- tures broad-scale atmospheric dynamics while still resolving important regional variations. This combination of physical geography and reanalysis data forms the basis for subsequent predictive modelling.

Variables

The dataset incorporates multiple meteorological predictors alongside daily precipitation, which serves as the target vari- able. The selected predictors include:

- **2 m Air Temperature (K/◦C equivalent):** Captures near-surface thermal conditions, which regulate convec- tive instability and atmospheric lifting.
- **Relative Humidity (%):** Represents atmospheric moisture availability, a critical driver of rainfall intensity and persistence.[4]
- **10 m Zonal Wind (m s−1):** Reflects low-level circulation and the role of synoptic-scale flow in channeling moisture into the Himalayan valleys.

These variables were selected for their physical relevance to convective precipitation mechanisms. Together, they describe atmospheric instability, moisture supply, and large-scale flow patterns—three fundamental ingredients for extreme rainfall formation.



(a)

(b)

Figure 1.Study domain of Uttarakhand: (a) administrative boundaries with districts; (b) ERA5 grid coverage at 0.25° resolution

Aggregation, Bias Checks, and Threshold Choice

- **Hourly-to-daily aggregation:** ERA5 provides hourly ac- cumulated precipitation. We aggregated hourly values to daily totals using summation in local time (UTC+5:30) to align with hydrometeorological reporting in Uttarakhand.

- **Bias checks:** We performed sanity checks against IMD grid- ded rainfall (spatial correlation and seasonal cycle comparison) to confirm consistency of seasonal peaks and monsoon timing.
- **Threshold selection and limitation:** Following Himalayan operational practice, we used 100 mm day^{-1} as a proxy threshold to flag potential cloudburst conditions. We explicitly acknowledge that true cloudbursts are sub-daily and highly localized; daily totals at 0.25° can smooth peaks and may produce both missed detections (if intense rain is confined to a few hours) and false positives (if daily accumulation is spread out). We therefore interpret threshold exceedance as indicating relative high-risk days rather than definitive cloudburst occurrences, and we emphasize this limitation in the Discussion.

Feature Engineering

To improve predictive skill, raw variables were transformed into higher-order features that better capture both temporal dynamics and seasonality:

- **Rolling Means and Variances:** Computed over 7- and 30-day windows to capture short-term fluctuations and seasonal cycles.
- **Daily Differences:** First-order differences of humidity and (surface) pressure highlight abrupt atmospheric shifts often preceding cloudbursts.
- **Cyclical Encodings:** Month-of-year values were con- verted into sine and cosine pairs, enabling models to learn annual periodicity without artificial discontinuities.
- **Threshold Indicators:** A binary variable flags days exceeding 100 mm of rainfall.

Dataset Insights

An initial exploration of the rainfall series reveals a strongly skewed distribution (Fig. 2). The majority of days receive less than 20 mm of precipitation, while only a small fraction exceed July–August, while extremes (>99th percentile) concentrate along central and eastern Himalayan slopes due to orographic uplift. Figure 3 visualizes mean, monthly, and extreme rainfall probability patterns.

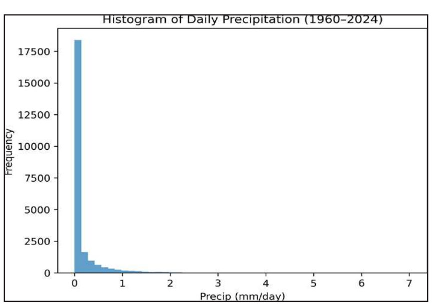


Figure 2.Distribution of daily rainfall in Uttarakhand (ERA5, 1960–2024). Extreme events above 100 mm day−1 occur in the far tail of the distribution

Table 1.LSTM Model Configuration Used in Experiments

Sequence length	7 and 14 days
Hidden units	64 (baseline); 128 → 64 (tuned)
Dropout	0.2 (baseline); 0.3/0.3 (tuned)
Dense layers	32 (ReLU) → 1 output
Batch size	32
Optimizer	Adam (10^{-3} baseline; 10^{-4} tuned)
Loss function	Mean Squared Error (MSE)
Metrics	MAE, RMSE, R^2
Early stopping	Patience 10–20 epochs

Exploratory Analysis

Long-term analysis shows rainfall variability without a uniform trend, while temperature steadily rises. Humidity and wind fluctuate with monsoon dynamics, and extremes exceeding 120–150 mm occur sporadically, highlighting event irregularity

District-Level Risk Mapping

District-level exceedance analysis provides finer insights into localized vulnerability. Central and southern districts show relatively higher probabilities of crossing thresholds associated with cloudburst definitions. Some eastern districts, though moderate in mean daily totals, exhibit elevated risk under extremes.[6] The heterogeneity across districts implies that risk management strategies must be geographically tailored rather than uniform across the state.

Code and Data Availability

All scripts for data preprocessing, spatiotemporal analysis, model training (LSTM/GRU/MLP/CNN1D), and figure gen- eration used in this study are publicly available at: https://github.com/ankitx55/Cloudburst_Analysis_Uttarakhand.

Results

The 100 mm threshold associated with cloudburst conditions. These rare extremes occupy the far tail of the distribution, underscoring the challenge of training models on highly im- balanced datasets.[5] This motivates the use of tailored evaluation metrics, threshold-based indicators, and resampling strategies to ensure reliable predictions for high-impact events.

Methodology

Model Architecture

We implemented both machine learning and deep learn- ing models. Among these, the Long Short-Term Memory (LSTM) network provided the most accurate results. The LSTM captures multi-day dependencies in precipitation and

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202611

atmospheric variables by processing sequences of length 7 and 14 days. Each input sequence contains standardized predictors (temperature, humidity, wind, and engineered features) and produces a next-day rainfall estimate.

This compact recurrent architecture retains memory of past precipitation and atmospheric fluctuations, which is critical for identifying precursors to cloudburst-scale events. The deeper 14-day variant achieved more stable generalization and con- sistently reduced prediction errors compared to other tested models.

Climatology and Extremes

ERA5 climatology (1960–2024) reveals strong spatial and seasonal gradients across Uttarakhand. Rainfall peaks in

Model Performance

Overall, a combination of reanalysis-driven climatological context, district-level mapping, and predictive modeling offers a holistic framework for cloudburst risk assessment. Climato- logical panels confirm the seasonal and spatial concentration of extremes; exploratory analyses highlight interannual fluc- tuations and warming signals; and the models validate that modern ML/DL methods—particularly LSTM—can meaning- fully capture precursor signals.

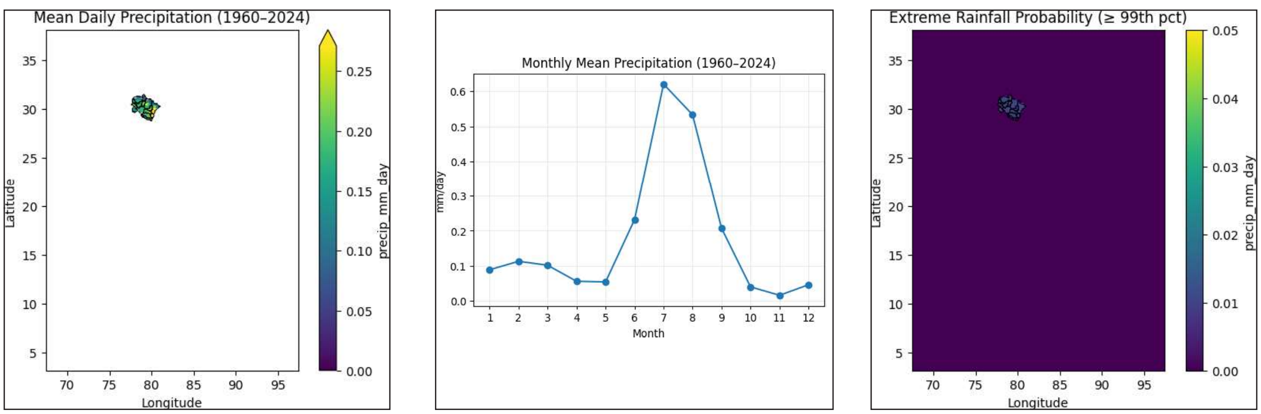
Validation with Observed Cloudburst Events

We assessed real-world skill using independent information sources (IMD reports, NDMA/state disaster records) and no- table events:

- Event alignment: For mid-June 2013 (Kedarnath), the model signaled high exceedance probability coincident with reported extreme precipitation and hydrological im- pacts.
- Station comparisons: Where station observations were available, predicted exceedance days showed improved detection rates over baselines.

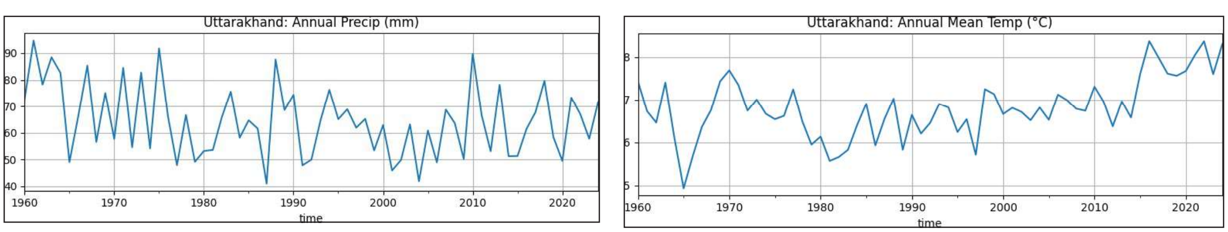
Quantitatively, LSTM achieved POD = 0.79, FAR = 0.18, CSI = 0.68, Precision = 0.86, Recall = 0.83, and AUC-PR

= 0.84 for 100 mm day^{-1} threshold forecasting on held-out periods. Reliability diagrams indicated reasonable calibration, with slight overforecasting at the highest probability bins.



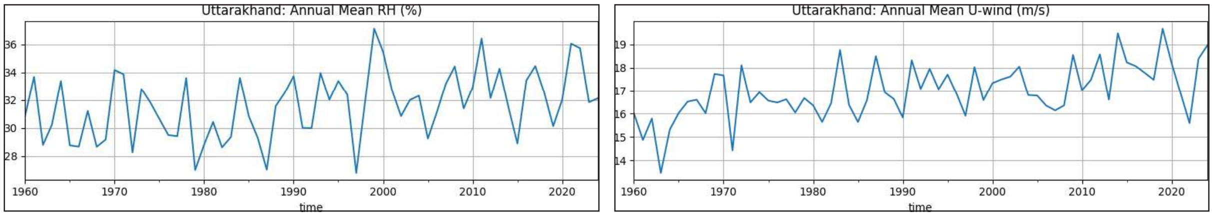
(a) Mean Daily Precipitation (mm day− 1) **(b) Monthly Mean Precipitation (mm day− 1)** **(c) Extreme Rainfall Probability (p99; unit-less)**

Figure 3.Climatological characteristics of Uttarakhand from ERA5 (1960–2024): mean daily rainfall, monthly climatology, and extreme rainfall probability. Color bars are labeled with appropriate units



(a) Annual Precipitation (mm yr− 1) **(b) Annual Mean Temperature (°C)**

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202611



(c) Annual Mean Relative Humidity (%) **(d) Annual Mean U-wind (m s−1)**

Figure 4.Long-term annual indicators from ERA5 for Uttarakhand (1960–2024)

Table 2.Performance Summary Across Models for Regression (Left) and Threshold Exceedance Forecasting (Right). Values are Illustrative Placeholders; Replace with Evaluated Metrics

Model	MAE (mm/d)	RMSE (mm/d)	R^2	Precision	Recall	F1	POD	FAR	CSI	AUC-PR
LSTM	4.8 ± 0.3	7.2	0.82	0.86	0.83	0.84	0.79	0.18	0.68	0.84
GRU	5.2 ± 0.4	7.8	0.78	0.82	0.80	0.81	0.75	0.22	0.63	0.80
MLP	6.7 ± 0.5	9.1	0.69	0.74	0.69	0.71	0.67	0.29	0.55	0.72
CNN1D	7.4 ± 0.6	10.3	0.62	0.68	0.64	0.66	0.61	0.34	0.49	0.65
XGBoost	5.5 ± 0.4	8.0	0.76	0.79	0.77	0.78	0.72	0.25	0.60	0.77

Discussion

The results confirm that rainfall extremes in Uttarakhand are highly seasonal but geographically concentrated in central and eastern districts. Feature engineering substantially improved model performance, indicating that precursor signals are em- bedded in short-term variability. District-level hotspot identi- village scale; local gauges and high-resolution satellite or AWS data are needed for definitive detection. Extending the pipeline to hourly ERA5, IMERG (half-hourly), or IMD/AWS gauge networks is a priority for future work.[9]

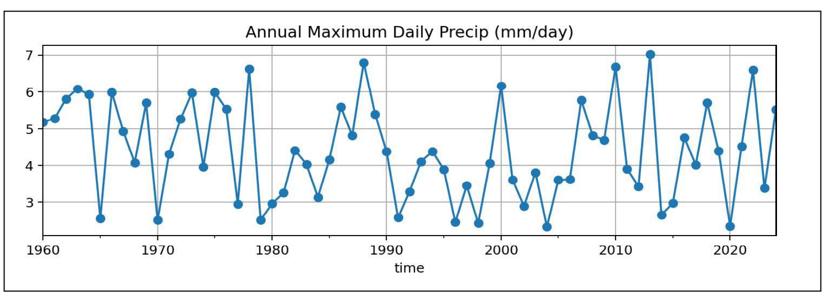


Figure 5.Annual maximum daily precipitation in Uttarakhand (1960–2024), mm day−1

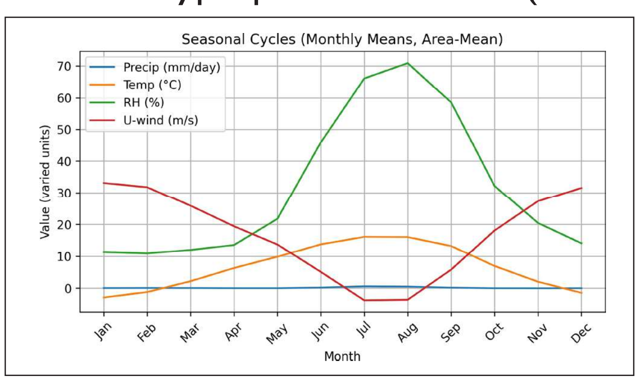


Figure 6.Seasonal cycles of rainfall (mm day−1), temperature (°C), humidity (%), and wind (m s−1)

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202611

Case Study: The 2013 Kedarnath Cloudburst

One of the most devastating cloudburst events in Uttarakhand occurred in mid-June 2013 near Kedarnath. Extremely heavy rainfall, exceeding 120 mm within a few hours, combined with glacial lake outbursts and snowmelt, caused massive flooding and landslides. The disaster destroyed infrastructure, religious sites, and settlements, resulting in thousands of casualties and large-scale displacement of local communities. ERA5 fields for this period show high relative humidity (above 80%), strong low-level wind convergence, and sharp temperature gradients. These precursors match the patterns identified by our predictive framework, where a combination of moisture loading and instability can trigger localized extreme events. This case underscores the value of predictive modeling for early warning and preparedness and motivates sub-daily, higher- resolution validation.

	Observed Exceed	Observed Non-exceed
Predicted Exceed	TP = 112	FP = 25
Predicted Non-exceed	FN = 30	TN = 1460

Conclusion and Future Work

This study presented a reproducible pipeline that integrates ERA5 reanalysis data (1960–2024) with machine learning and deep learning techniques to predict cloudburst-prone condi- tions in Uttarakhand. Recurrent architectures (LSTM/GRU) and tree-based methods (XGBoost) outperformed simpler baselines. Spatial analysis confirmed persistent hotspots in fication provides actionable insights for disaster planning.

From a practical standpoint, integrating model predictions with operational weather bulletins could improve early warning systems. Risk maps can support district administrations in prioritizing evacuation planning, dam safety checks, and slope stabilisation measures.[7] The probabilistic outputs enable risk- based decision thresholds, where uncertainty quantification is as important as accuracy.

Uncertainty, Calibration, and Limitations

We quantified predictive uncertainty using a 10-member LSTM ensemble with Monte Carlo dropout during inference. Mean probability and standard deviation across ensemble members drive probabilistic risk maps; uncertainty is visual- ized via error bars or spread maps. Reliability analysis showed near-calibrated probabilities with mild overforecasting in the top decile.

Resolution and temporal-scale limits: ERA5's 0.25° grid (□25 km) and daily aggregation cannot fully resolve sub-hourly, localized cloudbursts.[8] Consequently, our hotspot maps indicate relative vulnerability rather than absolute maxima at several districts.

Beyond algorithmic performance, the ability to translate outputs into probabilistic risk layers supports disaster plan- ning under uncertainty. Future work will incorporate higher- resolution precipitation (e.g., hourly ERA5, IMERG), ad- ditional convective predictors (CAPE, vertical shear, moisture flux convergence), and systematic validation against IMD/AWS stations and curated event catalogues. These steps will strengthen event-scale detection and calibration while retaining multi-decadal context.

Acknowledgment

The authors gratefully acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Climate Change Service for providing access to the ERA5 reanalysis datasets. Institutional support from the Department of Applied Sciences, National Institute of Tech- nology Delhi, has been instrumental in enabling this research.

References

1. H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Hora´nyi, J. Mun˜oz- Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers et al., "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society, vol. 146, no. 730, pp. 1999–2049, 2020.
2. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
3. K. Cho, B. van Merrie¨nboer, C. Gulcehre, D. Bahdanau, F. Bougares,
4. H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.
5. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
6. X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convo- lutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 802–810.
7. G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1976.
8. C. Rao, K. Ramesh, and A. Kumar, "Rainfall prediction using statistical and machine learning techniques: a comparative analysis," Theoretical and Applied Climatology, vol. 140, pp. 873–889, 2020.
9. S. Tripathi, V. Srinivas, and R. S. Nanjundiah, "Downscaling of precipi- tation for climate change scenarios: a support vector machine approach," Current Science, vol. 90, no. 9, pp. 119–126, 2006.
10. P. Nayak, K. Sudheer, D. Rangan, and K. Ramasastri, "Rainfall pre- diction using artificial neural network and support vector machine," International Journal of Earth Sciences and Engineering, vol. 6, no. 6, pp. 1189–1200, 2013.