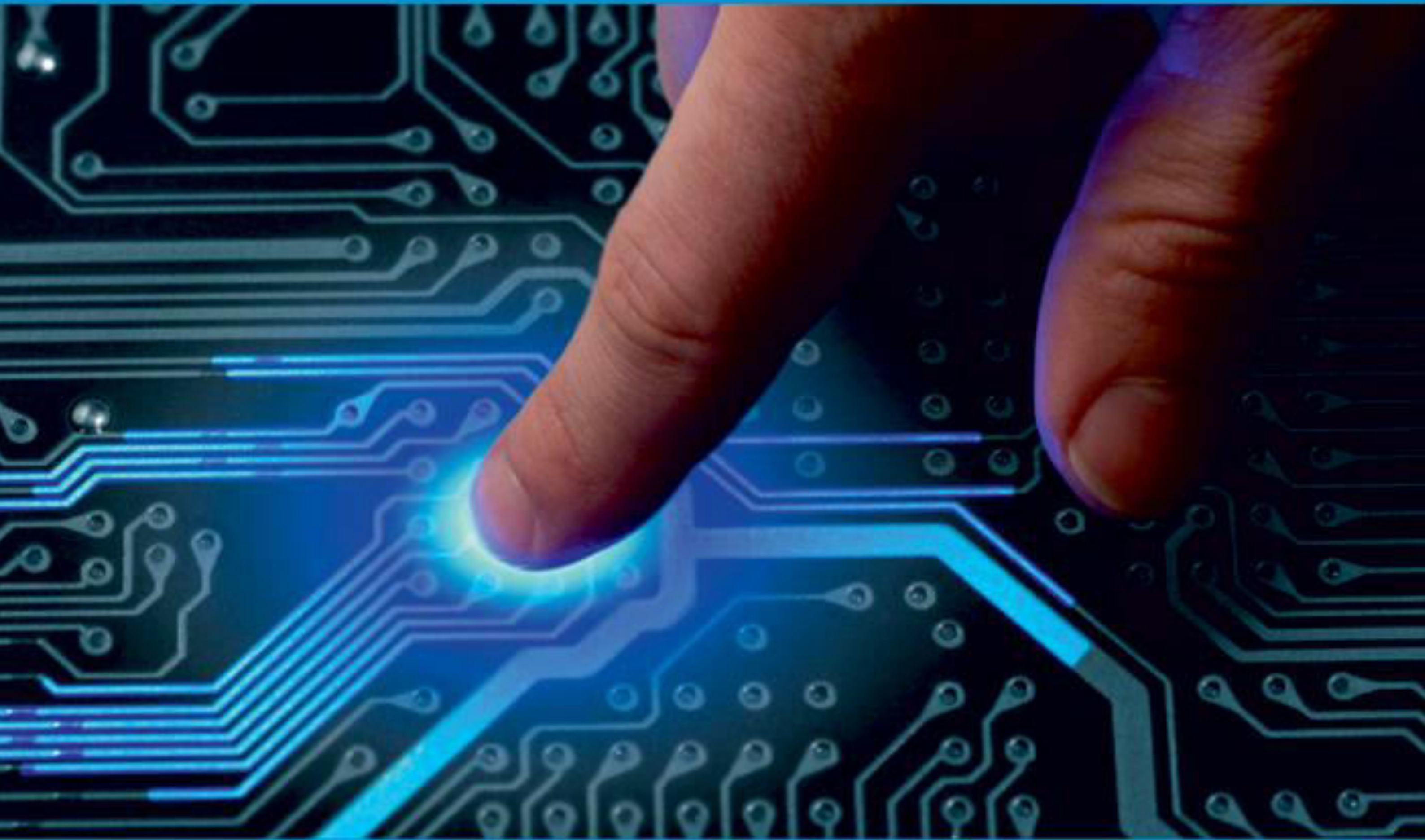
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Cloudburst Prediction in India Using Machine Learning

A.Tejaswini Reddy, B.Venkatesh, Sruthi Reddy, Dr.C.Dasthagiriah

Department of CSE, Anurag University, Hyderabad, Telangana, India

Department of CSE, Anurag University, Hyderabad, Telangana, India

Department of CSE, Anurag University, Hyderabad, Telangana, India

Associate Professor, Department of CSE, Anurag University, Hyderabad, Telangana, India

ABSTRACT: Cloudbursts pose a significant threat in India, especially during the South-West Monsoon season that commences in June. India's diverse climate regions, including the northern Himalayan region, Indo-Gangetic Plain, southern peninsula, and coastal areas, experience sporadic cloudbursts, with only 31 recorded instances, mainly in Himachal Pradesh, Uttarakhand, and Jammu and Kashmir. To address the lack of comprehensive Indian cloudburst data, we've curated a dataset, incorporating meteorological factors for cloudburst prediction. This dataset encompasses variables such as Temperature, Wind Gust, Wind Gust Speed, Humidity, Monsoon patterns, Air Pressure, and Cloud Density. Our goal is to improve preparedness and mitigation strategies, safeguarding lives, and property in cloudburstprone areas. Employing optimized machine learning algorithms, our model analyzes these parameters alongside prevailing weather conditions, facilitating cloudburst event prediction. We evaluate the prediction performance of machine learning algorithms, including KNN. The KNN algorithm outperformed others with an accuracy of 86.18%. Moreover, we provide graphical insights into the correlation between humidity and cloudburst occurrence, emphasizing the importance of weather variables in prediction models. This research contributes to cloudburst forecasting, even with limited Indian data, and highlights the potential of utilizing diverse machine learning techniques for improved accuracy.

I.INTRODUCTION

A cloudburst is a brief but intense precipitation event, often accompanied by hail and thunder, capable of causing flooding. These events can deposit a staggering 72,300 tons of water over a single area. Cloudbursts typically occur when the rainfall rate exceeds 100mm per hour. Traditional prediction methods include weather forecasting, data mining techniques for meteorological data modeling, and laser beam atmospheric extinction measurements from both manned and unmanned aerospace vehicles. Hailstorms and thunder can sometimes accompany these intense rain events [1].

Cloudburst events are commonly observed in mountainous regions, where warm air currents ascend, carrying raindrops upwards. This prevents spontaneous rainfall and leads to significant cloud condensation. As water accumulates at higher altitudes, the warmth below hinders its descent. The upward air currents weaken, resulting in a sudden downpour.

Cloudbursts typically occur at elevations ranging from 1000 to 2,500 meters above sea level [2].
India is recognized as a monsoon-driven nation on the global climate map, experiencing several cloudburst events in recent years, particularly in the western Himalayas and along the west coast. To enhance our understanding and prediction of cloudbursts, India has established the Cloud Observatories, a network of four high-altitude physics

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

observatories equipped with advanced technology. These observatories aim to investigate cloud and rain dynamics in high-altitude regions, focusing on cloud interactions, convection, circulation, and improving forecasting and monitoring of cloudburst incidents. Their ultimate goal is to mitigate the impact of such events in these areas [3]. In-depth research into cloudburst events through numerical modeling, as conducted by [1], revealed valuable insights into the dynamic structures and interactions with local topography. The author in [4], proposed that monsoonal low-pressure systems amplify low-level convergence and upper- level divergence, contributing to intense monsoonal heavy rainfall in orographic regions. Various studies have attributed the occurrence of heavy rainfall (ranging from 200 to 1000 mm/h) within a brief span to the presence of cumulonimbus clouds in the area. Gupta et al. [5] further emphasize that cloudburst events often result from convective systems trapped in enclosed valleys surrounded by mountainous terrain.

Sudden and intense short-term heavy rainfall events can be influenced by convective cloud feedbacks, whose connection to climate change remains uncertain due to their sensitivity to temperature stratification and largescale atmospheric circulation changes. Climate change may lead to variations in storm size, either increasing or decreasing. Nevertheless, it’s likely to result in higher rainfall intensity and expanded storm coverage, which can contribute to elevated rainfall levels. This, in turn, may lead to an increase in flash flooding, posing substantial regional concerns. Consequently, adapting to rapid climate change is essential to address the potential global impact of intensified flash floods [6].

Scientists examine historical data and climate models to understand the impact of rising global temperatures on heavy

rainfall patterns. The 2018 Kerala floods, triggered by exceptionally intense monsoon rains, resulted in widespread devastation, causing significant loss of life and displacement. Kerala, often the first to receive monsoon rains in India, faced additional challenges due to heavy rainfall events in 2018 and 2019, as confirmed by various observational and modeling studies, attributed in part to regional climate changes [7], [8]. Predicting cloudbursts in the Himalayan region is crucial to minimize potential damage and loss of life. Due to limited observations in this area, utilizing reanalysis data is necessary to understand the cloudburst formation process. Historical data is integrated through data assimilation techniques to create consistent gridded reanalysis data, offering insight into atmospheric conditions. Accurate forecasting of cloudburst events requires comprehensive data sources and analytical methods. Enhancing our understanding of cloudburst mechanisms is essential for proactive disaster management in the Himalayan region. Thus, the proposed approach combines historical data and modern machine learning methods to improve cloudburst prediction and reduce its impact.

This paper’s key contributions can be summarized as follows: 1) We have curated a dataset specifically for Indian cloudburst analysis. 2) We’ve utilized machine learning algorithms to forecast cloudburst events in the Indian subcontinent.

- Literature Survey

Numerical weather prediction (NWP) models offer moder- ate accuracy in large-scale medium-range weather forecasting, yet precipitation forecasting remains challenging. Mesoscale models rely on initial and boundary conditions from global models, which tend to oversimplify terrain, land cover, and vegetation. Both global and regional models neglect detailed geographical features for improved results. Operational forecasting centers utilize mesoscale models to deliver comprehensive weather forecasts for specific geographical regions with higher resolution [9]. Rising rainfall intensity and its correlation with temperature variations can lead to alterations in flood patterns, as demonstrated by [10]. Consequently, adjustments are required in flood forecasts to account for these evolving conditions. Back in 2008 [11], the cloudburst prediction model relied on an Arduino-connected rain gauge, but its primary drawback was the Arduino’s processing limitations.

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The cloudburst prediction model employed in 2010 was the WRF mesoscale Model [12], notable for its effectiveness. However, it presented challenges in terms of implementation and demanded substantial data resources. The work in [13] focuses on rainfall prediction through empirical statistical methods. We utilize various datasets, including variables like minimum and maximum temperature, pressure, wind direction, and relative humidity. The prediction model is based on Multiple Linear Regression. Certain predictors, such as wind direction, are excluded due to limitations in data collection, which could enhance predictive accuracy.

II.WORKING OF CLOUD BURST PREDICTION MECHANISM

Cloudburst Prediction using KNN:
Data Collection:
Weather Data: Collect historical weather data such as temperature, humidity, pressure, wind speed, precipitation, etc.
Cloudburst Incidents: Data on cloudburst events, including their locations and conditions before the event.
Data Preprocessing:
Data Cleaning: Handle missing or inconsistent data.
Feature Selection: Select important features that influence cloudbursts (e.g., temperature, humidity, etc.). Normalization: Normalize data to bring it to a comparable scale.
Training and Testing Split:
Split the dataset into a training set and a testing set (e.g., 80% for training, 20% for testing).KNN Algorithm:
Determine the Value of K: Choose an optimal value of K (number of neighbors).
Distance Calculation: Calculate the distance between the current weather conditions (input) and historical data points (e.g.,Euclidean distance).

Classification: The K nearest neighbors are identified and voted on whether a cloudburst is likely to occur based on their labels (cloudburst vs. no cloudburst).
Prediction:
Based on the majority vote of the K-nearest neighbors, classify the current weather condition as leading to a cloudburst ornot.
Evaluation:
Use metrics like accuracy, precision, recall, and F1-score to evaluate the prediction model on the testing dataset.

TRAIN DATASET

Day	Temperature	Humidity	wind	Cloudburst ?
1	28	89	16	Yes
2	30	78	15	No
3	29	96	19	No
4	31	87	20	No
5	27	83	19	No
6	30	92	13	No
7	29	75	16	No
8	26	88	18	Yes
9	27	98	10	No
10	28	76	17	No

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The KNN Classifier, is a versatile gradient boosting algorithm that excels in classification tasks, making it well-suited for cloudburst weather forecasting. It effectively combines multiple decision trees, reducing overfitting and improving the accuracy of predictions. This is crucial in the context of cloudburst prediction, where accurate and timely forecasts are of utmost importance. The XG Boost algorithm's focus on boosting performance and accuracy contributes significantly to more reliable forecasts of cloudburst events. It employs regularization techniques and optimized tree building processes, resultingin improved model generalization and predictive power.

TABLE I – DATASET CONSIDERED FOR EXPERIMENTATION.EXPERIMENTATION AND RESULT ANALYSIS

In experimentation, we employed a unique approach to address the scarcity of Indian cloudburst prediction datasets. We initially obtained a reference dataset from Kaggle, which was originally Australian based. Subsequently, we adapted this Australian datasetto create an Indian counterpart with matching attributes.

Our model's training process leveraged the Australian dataset, while the Indian dataset served as the testing dataset. We ensured that both datasets shared common attributes, which were categorized into numerical, discrete, continuous, and categorical features. We employed various supervised machine learning algorithms, including Cat Boost, Random Forest, Decision Tree, Logistic Regression, and XGBoost Classifier. These techniques were integrated into our proposed selforganized structure for prediction. Table I in our study illustrates the datasets used for training and testing. By aligning the attributes between the Australian and Indian datasets, we aimed to build a robust model for cloudburst prediction, despite the limited availability of dedicated Indian data. Cloudburst datasets typically consist of tabular weather data, where each column represents a specific variable, and each row corresponds to a data point relevant to cloudburst prediction.

TABLE II – EXPERIMENT RESULTS BASED ON F1- SCORE, PRECISION, RECALL, SUPPORT

Day	Temperature	Humidity	wind	Cloudburst ?	Accuracy
1	28	89	16	Yes	80.12
2	30	78	15	No	79.12
3	29	96	19	No	81
4	31	87	20	No	83
5	27	83	19	No	81.45
6	30	92	13	No	81.76
7	29	75	16	No	80.98
8	26	88	18	Yes	80
9	27	98	10	No	79.56
10	28	76	17	No	80.12

TABLE III – ACCURACY OF DIFFERENT DAYS

In Table 2 and 3, the algorithm performance metrics are presented. The Cat Boost algorithm achieved the highest accuracy at 86.18%. For value 0, it had a precision of 0.88, recall of 0.95, and an F1 score of 0.91. For value 1, it showed a precision of 0.75, recall of 0.56, and an F1 score of 0.64. The Random Forest algorithm attained an accuracy of 84.14%. For value 0, it had a precision of 0.89, recall of 0.91, and an F1 score of 0.90. For value 1, it exhibited a precision of 0.66, recall of 0.61, and an F1 score of 0.63.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In Figure 1, we depicted a graph illustrating the relationship between humidity and count. Relative humidity, which quantifies the moisture content in the air relative to its capacity, plays a pivotal role in weather analysis. You can measure humidity using a hygrometer or calculate it based on air temperature, dew point, and established equations. Alternatively, aDIY approach involves constructing a sling psychrometer using readily available materials and basic tools.

specific humidity = 6.11 x 10^ {7.5 x dew point} / {237.3 + dew point}

In Figure 2, we presented a correlation matrix, a tabular representation of correlation coefficients between variables.

Each cell within the table reflects the correlation between two specific variables. This matrix serves to summarize data, act as input for advanced analyses, and diagnose more complex analytical processes. Typically, a correlation matrix is square, featuring identical variables in both rows and columns. The main diagonal, displaying a line of 1.00s from the top-left to the bottom-right, signifies perfect self-correlation for each variable. This symmetrical matrix exhibits mirrored correlations above and below the main diagonal.

III.CONCLUSION

Our study demonstrates the feasibility of leveraging diverse data sources and advanced machine learning algorithms to address data limitations in cloudburst prediction We employed several machines learning algorithms, including KNN, Decision Tree, and Logistic Regression, in our experimentation. The results revealed that KNN outperformed other algorithms with an accuracy of 81.18%. This highlights the algorithm's ability to handle both categorical and numerical features effectively, a critical aspect of cloudburst prediction. Random Forest also demonstrated promising results with an accuracy of 79.14%, emphasizing its ensemble approach's advantage in capturing complex weather patterns.

REFERENCES

1. Das, S., Ashrit, R., and Moncrieff, M., “Simulation of a Himalayan cloudburst event,” Journal of earth system science, vol. 115, pp. 299– 313, 2006.
2. Chaudhuri, C., Tripathi, S., Srivastava, R., and Misra, A., “Observation and numerical-analysis-based dynamics of the Uttarkashi cloudburst,” in Annales Geophysical, vol. 33, no. 6. Copernicus GmbH Gottingen, ¨Germany, 2015, pp. 671–686.
3. Bhan, S., Devyani, A., and Sinha, V., “An analysis of monthly rainfall and the meteorological conditions associated with cloudburst over the dry region of leh (Ladakh), India,” Mausam, vol. 66, no. 1, pp. 107– 122, 2015.
4. Joseph, S., Sahai, A. K., Sharmila, S., Abhilash, S., Borah, N., Chattopadhyay, R., Pillai, P. A., Rajeevan, M., and Kumar, A., “North Indian heavy rainfall event during June 2013: diagnostics and extended range prediction,” Climate Dynamics, vol. 44, pp. 2049– 2065, 2015.
5. Gupta, V., Dobhal, D., and Vaideswaran, S., “August 2012 cloudburst and subsequent flash flood in the asi ganga, a tributary of the Bhagirathi River, Garhwal Himalaya, India,” Current Science, pp. 249–253, 2013.
6. Ali, H., Peleg, N., and Fowler, H. J., “Global scaling of rainfall with dewpoint temperature reveals considerable ocean-land difference,” Geophysical Research Letters, vol. 48, no. 15, p. e2021GL093798, 2021.
7. Dixit, A., Sahani, S., Rajagopalan, B., and Choubey, S., “Role of changing land use and land cover (lull) on the 2018 mega floodsover Kerala, India,” Climate Research, vol. 89, pp. 1–14, 2022.
8. 8 Dimi, A., Chegutu, A., Niyogi, D., Thayer, R. J., Ray, K., Tripathi, S., Pandey, A., and Mohanty, U., “Cloudbursts in Indian Himalayas: a review,” Earth-Science Reviews, vol. 168, pp. 1–23, 2017.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

9. Das, S., "Mountain weather forecasting using mm5 modelling system," Cur. Sci, vol. 88, no. 6, pp. 899–905, 2005.
10. Wasko, C., Westra, S., Nathan, R., Orr, H. G., Villarini, G., Villalobos Herrera, R., and Fowler, H. J., "Incorporating climate change in flood estimation guidance," Philosophical Transactions of the Royal Society A, vol. 379, no. 2195, p. 20190548, 2021.
11. Tiwari, A. and Verma, S., "Cloudburst predetermination system," SOR J. Compute. Eng, vol. 17, pp. 44–56, 2015.
12. Fodor, T., Beeharry, Y., Humbugs, V., Basso, V., and RamnarainSeetohul, V., "A framework for a real-time cloud-based weather forecasting system for Mauritius," International Journal of

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details