Cloud Burst Prediction System using Machine Learning

Herin Shani.S
Department of CSE
Saveetha engineering college(Autonomus)
Chennai, India
herinshani2311@gmail.com

Dr. G. Nagappan
Department of CSE
Saveetha engineering college(Autonomus)
Chennai, India
Nagappan.cse@saveetha.ac.in

***Abstract*—This abstract presents an innovative approach that leverages the Gramian Angular Field (GAF) in conjunction with Convolutional Neural Networks (CNN) to improve the accuracy and reliability of cloudburst prediction systems. The utilization of GAF and CNN represents a breakthrough in modeling the complexities inherent in meteorological data. GAF, with its generative capabilities, constructs synthetic data closely resembling the underlying meteorological distributions. CNN, renowned for its proficiency in spatial data analysis, is adept at recognizing intricate patterns within meteorological images. The application of GAF and CNN in cloudburst prediction systems signifies a significant advancement in the field. By effectively recognizing and generating synthetic data representative of meteorological complexities, this approach has the potential to significantly improve the precision and lead time of cloudburst predictions, thereby enhancing early warning systems and disaster preparedness.**

Keywords-Meteorological Data Analysis, Machine Learning in Meteorology, Early Warning Systems, Disaster Preparedness, Extreme Weather Forecasting, Data-driven Prediction ,Meteorological Image Analysis, Enhanced Prediction Accuracy.

I. INTRODUCTION

This project focuses on the innovative approach of using Gramian Angular Field (GAF) in combination with Convolutional Neural Networks (CNN) for cloudburst prediction.. By combining GAF and CNN, this approach aims to harness the strengths of generative modeling and deep learning to capture the intricate spatial and temporal patterns inherent in meteorological data. This synergistic coupling aims to improve the robustness and generalization of cloudburst prediction models. application of GAF and CNN in cloudburst prediction not only holds promise for increased accuracy in forecasting but also presents an opportunity to handle complex, high-dimensional meteorological data more effectively. The ability to recognize subtle patterns and relationships in meteorological data, combined with the capacity to generate synthetic yet representative datasets, has the potential to significantly enhance the precision and lead time in cloudburst predictions. This introduction sets the stage for exploring the application of GAF and CNN in cloudburst prediction systems, highlighting the potential advancements in early warning systems and disaster preparedness. The subsequent sections will delve deeper into the methodologies, algorithms, and outcomes of utilizing GAF and CNN for more accurate, timely, and reliable cloudburst predictions and recruiters are well-equipped to succeed in the digital recruitment era.

II. RELATED WORKS

Cloudburst revolutionizes serverless computing by enabling stateful Python programming through Anna for state sharing and caches, significantly reducing state-management issues and enhancing serverless consistency across applications[1]. Utilizing AI and data science, an advanced system aims to predict destructive cloudbursts in hilly areas by analyzing pressure, humidity, and temperature, providing early warnings to vulnerable regions, inspired by Koottickal village events[2]. Study assesses NETRA model alerts' accuracy for Western Himalayan cloudbursts in Uttarakhand, notably successful in Chamoli, Rudra Prayag, and Uttarkashi districts. May emerges as the most critical month. Co-occurrence analysis emphasizes Pauri and Uttarkashi[3]. Cloudburst offers a deployable remedy using forward error correction (FEC) over multipath, reducing datacenter short flow latency. It spreads FEC-coded packets proactively, cutting message completion time significantly compared to DCTCP and PIAS[4]. Uttarakhand's sparse rain gauge networks heighten vulnerability to disasters. Comparing TRMM satellite rainfall with gauge data reveals good agreement, highlighting the necessity for enhanced satellite retrieval algorithms incorporating local factors and topography[5]. This study proposes a Convolutional Neural Network (CNN) for timely landslide detection. Using features like vegetation index, temperature, and precipitation, the research delves into model architecture, feature processing, performance evaluation, and future improvements[6].

Researchers have access to a dataset comprising 7,600+ disaster-news articles covering COVID-19, storms, floods, and natural disasters. Created datasets aid in sentence classification, summarization, and identifying event details, demonstrating success with Random Forest classification[7]. Rising cloudburst occurrences in the Himalayas' southern region due to warmer climates require detailed investigation. Analyzing the 2012 Uttarkashi cloudburst using high-res IMDAA and ERA5 datasets reveals IMDAA's superior representation of variables and mechanisms[8].The study explored constructing a Splash cloudburst-disaster model, relying on rainfall intensity, individual property values, and municipal gauge network data in

2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0 | 979-8-3503-7378-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/OTCON60325.2024.10687554

Jönköping. Results suggested potential for simplified, reliable cloudburst catastrophe models in various urban contexts[9]. Uttarakhand's Upper Ganga Basin faces flash flood risks from extreme rainfall and cloudbursts. Analysis of 57 events identifies vulnerable areas, notably in the July-August monsoon, impacting densely populated regions[10].Kerala faced devastating floods in 2018 and 2019. The 2019 event, displaying unusual convective nature fueled by warm sea temperatures, met mesoscale cloudburst criteria, uncommon for the region. Such events, influenced by global warming, may endanger Western Ghats ecosystems[11]. Climate-induced storms challenge urban stormwater systems. Passive urban blue-green infrastructure struggles during cloudbursts, causing downstream flood risks. Research proposes adaptive real-time control (RTC) in Tongzhou, Beijing, reducing peak outflow, proving scalable urban sustainability solutions[12].

III. ARCHITECTURE DIAGRAM

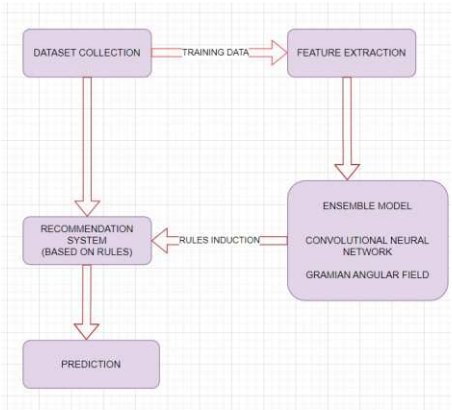

Fig.1. "Cloudburst Predicition"

IV. PROPOSED METHODOLOGY

A. **Feature Selection and Engineering:** Identify the most relevant features that may contribute to cloudburst prediction. Additionally, create new features derived from existing ones that could improve the model's performance.

B. **Model Selection**: Experiment with different machine learning models suitable for the prediction task. Models such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), or neural networks might be considered. Ensemble methods or hybrid models can also be tested for better accuracy.

C. Model Training : Split the data into training and testing sets. Train the selected models using the training data and validate their performance using the testing data. Techniques like cross-validation may be applied to prevent overfitting.

D. Hyperparameter Tuning: Fine-tune the models by adjusting their hyperparameters to improve performance. This process involves optimizing parameters that affect the model's learning.

E. Deployment and Integration: Once a satisfactory model is developed, integrate it into a system that can continuously receive and analyze real-time data to predict the likelihood of a cloudburst.

F. **Monitoring and Updates:** Continuous monitoring of the model's performance is crucial. Additionally, the model may need periodic updates to adapt to changing weather patterns and improve prediction accuracy.

G. **Validation and Improvement**: Periodically validate the model's predictions against real occurrences of cloudbursts and continuously work on improving the model's accuracy.

It's important to note that the success of the project heavily relies on the quality and quantity of data available for training, the choice of features, and the robustness of the chosen machine learning model. Moreover, collaboration with domain experts in meteorology can significantly enhance the accuracy and relevance of the predictions.

V. ALGORITHM

The Convolutional Neural Network (CNN) is a type of deep neural network optimized for processing structured grid-like data, such as two-dimensional layouts typical in visual imagery. This makes CNNs particularly effective for tasks involving photographs, videos, and other visual media. By leveraging their ability to detect and interpret spatial hierarchies in images—recognizing patterns from simple to complex—CNNs excel in areas where the identification of objects, scenes, and activities in visual content is crucial. They consist of various layers, including convolutional layers that filter inputs for useful information, pooling layers that reduce data dimensionality, and fully connected layers that classify the extracted features into outputs. This structure enables CNNs to focus on specific features without being overwhelmed by data size or complexity, making them highly suitable for image recognition, video analysis, and other automated tasks that require a sophisticated understanding of visual contexts.

Explanation of CNN Algorithm:

CNNs have transformed computer vision by empowering machines to interpret visual information, and they are equally influential in fields like natural language processing and medical image analysis where data often has a grid-like structure. These networks excel at automatically learning hierarchical representations, making them superb at extracting features and recognizing patterns. The process involves several key layers: convolutional layers detect features by applying filters to the input; pooling layers reduce dimensionality to simplify the information; and fully connected layers classify these features into meaningful outputs. This structured approach enables CNNs to efficiently process and analyze complex datasets across various domains.**Convolutional layers**: These layers apply convolution operations to input data using filters or kernels, scanning the input with these filters to extract specific features.

1. **Pooling layers**: Pooling, like max pooling or average pooling, reduces the dimensionality of the data, preserving the most important information and reducing computational requirements

2. **Activation functions:** Typically, CNNs use activation functions like ReLU (Rectified Linear Activation) to introduce non-linearity into the network, allowing it to learn complex patterns.

3. **Fully connected layers**: At the end of the network, fully connected layers combine the features extracted by earlier layers to make predictions or classifications.

4. **Output:** The final output should ideally provide actionable information for decision-makers or stakeholders to take preventive measures or plan responses in the case of an impending cloudburst.

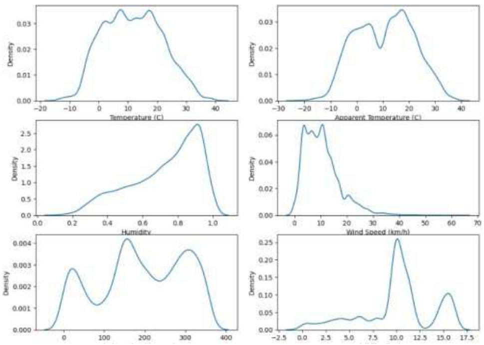

Fig. 2. Graphical representation

In the proposed system, the CNN algorithm can be employed for various purposes:

1. **Probability or Confidence Score:** The CNN could output a probability or confidence score indicating the likelihood of a cloudburst occurrence within a certain timeframe. This could be a continuous value representing the estimated probability of a cloudburst event.

2. **Binary Classification:** The output might be a binary classification result indicating the presence or absence of an imminent cloudburst. For instance, the CNN could output a binary decision - "cloudburst likely" or "cloudburst not likely" based on the input data and the learned patterns.

3. **Spatial or Temporal Prediction:** Depending on the design, the system might generate a spatial map indicating regions with higher potential for a cloudburst event. Alternatively, it could forecast the temporal aspect, predicting the time window in which a cloudburst is more likely to occur

4. **Risk Assessment or Severity Level**: The CNN could output an assessment of the severity or risk level associated with the potential cloudburst. It could indicate the potential impact or scale of the event based on learned features.

5. **Visualization of Weather Patterns:** The output might be visual representations, such as heat maps or other graphical representations, showing weather patterns or atmospheric conditions that contribute to a higher likelihood of a cloudburst.

In summary, These potential outputs will heavily rely on the data available, the features extracted, and the model's ability to learn and predict cloudburst occurrences based on the identified patterns and relationships within the input data.

VI. RESULTS

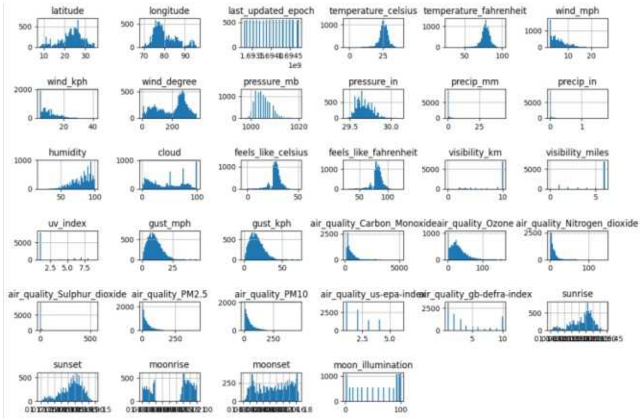

Fig. 3. Tight layout

Initially, the project showcases temperature and various weather conditions for a specific location, serving as a foundational example of how the CNN algorithm can be employed to predict cloudburst events. The displayed graph, which leverages this algorithm, focuses on forecasting the occurrence of cloudbursts in a particular area. Specifically, this graph highlights the safety conditions of the Northern Himalayas, providing critical insights into potential severe weather patterns. By utilizing CNN, the model processes and analyzes meteorological data to predict sudden, intense rainfall, thus aiding in disaster preparedness and risk management for the region.

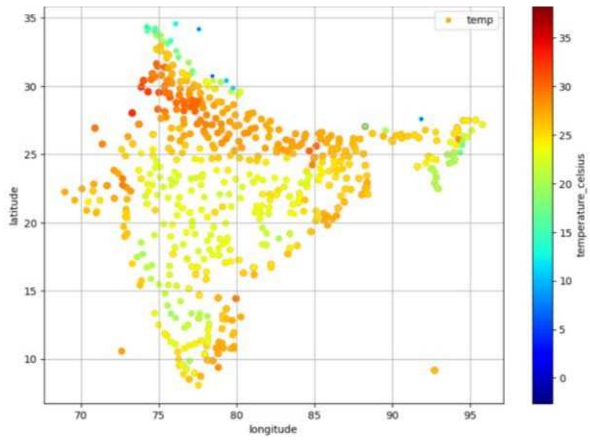

Fig. 4..Celsius of given location

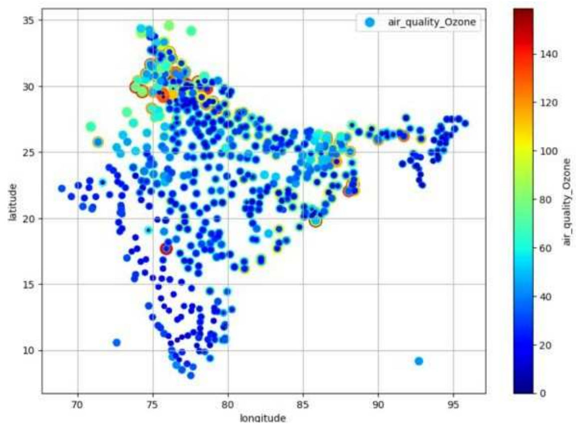

Fig. 5 Air quality of given location

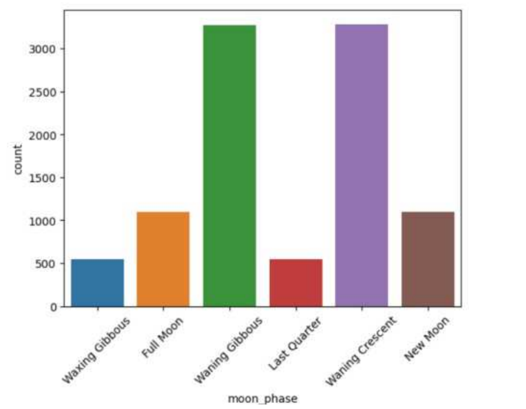

Fig. 6 Moon phase for the given Location

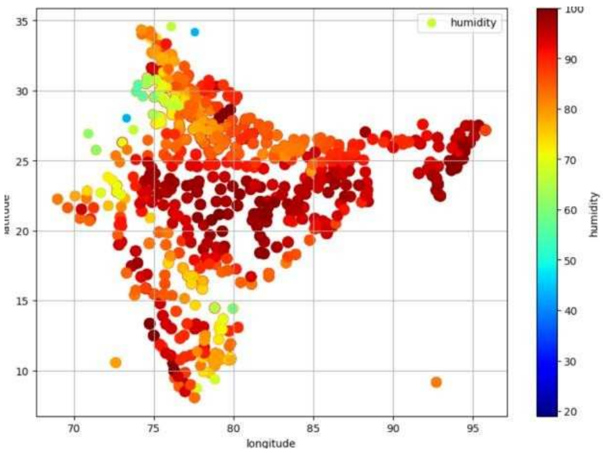

Fig. 7 Humidity of given location

The project utilizes data such as Celsius temperature, humidity, UV index, moon illumination, and wind speed to generate a graph. This graph, created through the application of the CNN algorithm, effectively illustrates the clustered components of the weather data. The CNN's capacity to analyze and visualize complex datasets allows for a comprehensive depiction of these variables in a coherent and interpretable format. By organizing and clustering the data, the graph provides valuable insights into the various factors that might influence weather conditions in a specific area, serving as a tool for better understanding and predicting environmental and atmospheric changes.

VII. PREDICTIVE ANALYSIS

Predictive analysis for cloudburst events, which are characterized by sudden, intense rainfalls that can lead to severe flooding, is crucial for disaster readiness and mitigation. To develop a predictive model utilizing Convolutional Neural Networks (CNNs) and Gramian Angular Fields (GAF), one must begin by gathering high-resolution meteorological data, including rainfall intensity, cloud coverage, temperature, and more. This data must then be prepared by transforming it into GAF images—a technique that encodes time series data into matrix formats using polar coordinates, which helps preserve the correlation between different times through Gramian Angular Summation Field (GASF) and relative timing of events through (GADF).

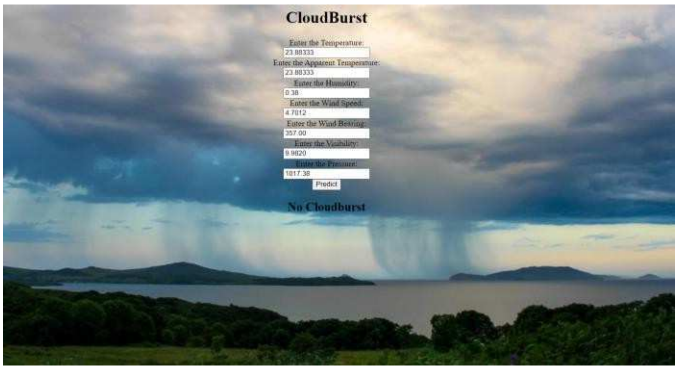

Fig. 8. No cloudburst Occurs

Following data preparation, a tailored CNN architecture is developed, comprising multiple layers designed to extract spatial hierarchies of features and reduce computational load. This model is trained using historical, labeled data, optimizing through techniques like backpropagation and employing metrics such as accuracy and precision for evaluation. Once validated, the model can be deployed in real-time systems to predict cloudbursts, necessitating continuous monitoring and periodic retraining with new data to maintain accuracy and adapt to changing patterns. This approach not only leverages the CNN's pattern recognition capabilities but also uses GAF's effective transformation of time-series data, providing a robust framework for predictive analysis in meteorological contexts.

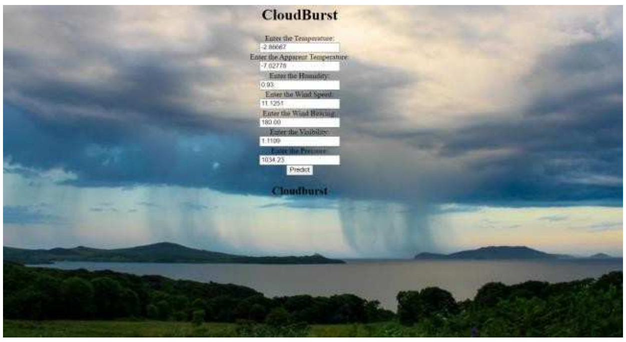

Fig. 9. Cloudburst Occurs

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

Algorithm	**Value**	**Precision**	**Recall**	**F1-score**	**support**
CNN	0	0.89	0.94	0.93	22717
	1	0.78	0.58	0.66	6375
Cat boost	0	0.88	0.95	0.91	22717
	1	0.75	0.56	0.64	6375
Random forest	0	0.89	0.91	0.90	22717
	1	0.66	0.61	0.63	6375
Logistic regression	0	0.92	0.77	0.84	22717
	1	0.48	0.76	0.59	6375
K nearest neighbour	0	0.90	0.77	0.83	22717
	1	0.46	0.71	0.56	6375
XGB classifier	0	0.88	0.94	0.91	22717
	1	0.72	0.55	0.62	6375
Decision tree	0	0.87	0.83	0.85	22717
	1	0.48	0.55	0.51	6375

TABLE II – ACCURACY OF DIFFERENT ALGORITHMS

Accuracy of CNN	86.42
Accuracy of catboost	86.18
Accuracy of randomforest	84.45
Accuracy of logistic regression	76.90
Accuracy of Knearestneighbour	75.53
Accuracy of XGBclassifier	85.49
Accuracy of decisiontree	76.97

In Table 1 and 2, the algorithm performance metrics are presented. The CNN algorithm achieved the highest accuracy at 86.42%. For value 0, it had a precision of 0.89, recall of 0.94, and an F1 score of 0.93. For value 1, it showed a precision of 0.78, recall of 0.58, and an F1 score of 0.66. The Cat Boost algorithm achieved an accuracy of 86.18%, with a precision of 0.88, recall of 0.95, and an F1 score of 0.91 for the category 0. For category 1, it recorded a precision of 0.75, recall of 0.56, and an F1 score of 0.64. On the other hand, the Random Forest algorithm reached an accuracy of 84.14%. It demonstrated a precision of 0.89, recall of 0.91, and an F1 score of 0.90 for category 0, while for category 1, the precision was 0.66, recall was 0.61, and the F1 score stood at 0.63.

Logistic Regression achieved an accuracy of 76.90%, with scores for category 0 including a precision of 0.92, recall of 0.77, and an F1 score of 0.84. For category 1, the scores were a precision of 0.48, recall of 0.76, and an F1 score of 0.59. The XGB classifier algorithm reached an accuracy of 85.49%, with precision, recall, and F1 scores of 0.88, 0.94, and 0.91 for category 0, and 0.72, 0.55, and 0.62 for category 1. The Decision tree algorithm showed an accuracy of 76.97%, with scores for category 0 of 0.87 in precision, 0.83 in recall, and 0.85 in F1 score, and for category 1, scores of 0.48 in precision, 0.55 in recall, and 0.51 in F1 score. Lastly, K-nearest Neighbors recorded an accuracy of 75.53%, with precision, recall, and F1 scores of 0.90, 0.77, and 0.83 for category 0, and 0.46, 0.71, and 0.56 for category 1.

VIII. CONCLUSION

The integration of these advanced techniques demonstrates the potential for more accurate and spatially nuanced predictions. The ability of GAF to represent complex data and CNN's proficiency in recognizing patterns within this data contribute to a robust predictive model. While showing promise, continued research and development are crucial to refine these systems, aiming for improved accuracy and reliability in cloudburst forecasting. The integration of GAF and CNN marks a significant step forward in enhancing our capability to anticipate and potentially mitigate the impact of these extreme weather events.

IX. FUTURE ENHANCEMENT

Implementing more sophisticated GAF-CNN models may improve accuracy by capturing complex spatial patterns in atmospheric data, potentially leading to more reliable and timely cloudburst forecasts. Additionally, exploring ensemble methods or incorporating real-time data streams for more dynamic and adaptable predictions could further enhance the system's capabilities.

REFERENCES

[1] Vikram Sreekanti, Chenggang Wu, Xiayue Charles Lin, Johann Schleier-Smith, Jose M. Faleiro, Joseph E. Gonzalez, Joseph M. Hellerstein, Alexey Tumanov “Cloudburst: Stateful Functions-as-a-Service” Cornell University 2020.

[2] Albin Sebastian; Jachin Bency Thomas; Rinu Mathew; Riya Jose; Anto Manuel “SkySentinel: Harnessing AI for Cloudburst Forecasting and Warning”, International Conference on Circuit Power and Computing Technologies (ICCPCT), August/2023

[3] Nita H Shah, Anupam Priamvada & Bipasha Paul Shukla, “Validation of satellite-based cloudburst alerts: An assessment of location precision over Uttarakhand, India”, Journal of Earth System Science, September 2023.

[4] Gaoxiong Zeng; Li Chen; Bairen Yi; Kai Chen “Cutting Tail Latency in Commodity Datacenters with Cloudburst”, IEEE Conference on Computer Communications 2022.

[5] Anoop Kumar Shukla; Satyavati Shukla,”Satellite Precipitation Estimates (SPEs) and Their Validation Using Ground-Based Measurments“,IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium 2020.

[6] Ruchika Malhotra; Abhinav Thapper; Aditya Kulraj Kunwar; Ashutosh Raturi,” Landslide Prediction using CNN with Data Modeling and Feature Reduction using Median Pooling”, 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)2021.

[7] Sumanta Banerjee; Shyamapada Mukherjee; Sivaji Bandyopadhyay “Disaster-news datasets for multi-label document classification, sentence classification, and abstractive document summarization tasks” International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) 2023.

[8] Shivaji Singh Patel, A. Routray , Devajyoti Dutta , Rajeev Bhatla , Vivek Singh , John P. George “Studying the evolution of Uttarkashi cloudburst

event from reanalysis datasets–A case study" Dynamics of Atmospheres and Oceans 2023.

[9] Daniel Knös , Konstantinos Karagiorgos , Jan Haas , Barbara Blumenthal , Lars Nyberg , Sven Halldin "Cloudburst-disaster modelling. A new open-source catastrophe model", International Journal of Disaster Risk Reduction 2022

[10] Prabhash K. Mishra, Renoj J. Thayyen, Hemant Singh, Swagatam Das, Manish K. Nema, Pradeep Kumar "Assessment of cloudbursts, extreme rainfall and vulnerable regions in the Upper Ganga basin, Uttarakhand, India" International Journal of Disaster Risk Reduction 2022.

[11] P. Vijaykumar , S. Abhilash , A.V. Sreenath , U.N. Athira , K. Mohanakumar , B.E. Mapes , B. Chakrapani , A.K. Sahai , T.N. Niyas , O.P. Sreejith "Kerala floods in consecutive years - Its association with mesoscale cloudburst and structural changes in monsoon clouds over the west coast of India" Weather and Climate Extremes 2021.

[12] Huaiyu Zhou , Ruidong Li , Hailong Liu , Guangheng Ni "Real-time control enhanced blue-green infrastructure towards torrential events: A smart predictive solution" , Urban Climate 2023.