CloudEye - Cloud Burst Prediction System

Asmi Rajbhar
Dept. Of Computer Engineering
VESIT, Mumbai, India
d2022.asmi.rajbhar@ves.ac.in

Ravina Vartak
Dept. Of Computer Engineering
VESIT, Mumbai, India
d2022.ravina.vartak@ves.ac.in

Dr. Gresha Bhatia
Dept. Of Computer Engineering
VESIT, Mumbai, India
gresha.bhatia@ves.ac.in

Om Patil
Dept. Of Computer Engineering
VESIT, Mumbai, India
d2022.om.patil@ves.ac.in

Madhura Golatkar
Dept. Of Computer Engineering
VESIT, Mumbai, India
2022.madhura.golatkar@ves.ac.in

***Abstract* - Cloud bursts are sudden and intense rainfall events that can lead to catastrophic flooding, causing severe damage to life and infrastructure. Accurate and timely prediction of cloud bursts is crucial for disaster preparedness and mitigation. This research presents an AI-driven Cloud Burst Prediction System that utilizes machine learning and meteorological data to forecast extreme rainfall events. The proposed system integrates atmospheric parameters and historical weather patterns to train predictive models for early warning alerts. Additionally, the system features an automated email notification framework that delivers personalized alerts to users two hours before predicted events, providing current weather conditions and nearby safe location details retrieved from a PostgreSQL database. Experimental evaluations demonstrate the system's effectiveness in identifying potential cloud burst scenarios with high precision while significantly improving evacuation response times. The findings highlight the significance of AI in enhancing weather forecasting capabilities and disaster risk reduction strategies.**

Keywords— Cloudburst prediction, CatBoost, Real-time weather data, disaster preparedness.

I. INTRODUCTION

Cloudbursts are extreme weather events characterized by short-duration, intense rainfall over a confined area, typically spanning 20-30 square kilometers. These events can lead to devastating flood-like situations within minutes. According to the India Meteorological Department, a rainfall event qualifies as a cloudburst when the precipitation exceeds 100 millimeters in an hour. While heavy rainfall is common during the monsoon season, not all instances meet this threshold.

From 1970 to December 2023, 32 cloudburst events have been recorded in India. The state of Uttarakhand is the most affected, with 11 recorded events, a consequence of its mountainous terrain and monsoon climate. Himachal Pradesh follows with 7 recorded incidents. Although cloudbursts are highly localized, their impacts can be far-reaching, often causing significant destruction in surrounding areas.

Cloudbursts predominantly occur in regions with significant orographic influence, such as the Himalayan states, Northeastern India, and the Western Ghats. These events result from moisture-laden air ascending mountainous areas, where it forms towering Cumulonimbus clouds. The upward movement of these clouds, combined with high relative humidity and dense cloud cover, provides the necessary conditions for cloudburst formation.

II. OBJECTIVES

System Development: Develop a comprehensive cloud burst prediction system for hilly and mountainous regions. This system will integrate advanced weather modeling techniques with machine learning algorithms, specifically utilizing the CatBoost model, to create a robust and accurate prediction tool tailored for challenging terrains.

Pattern Analysis: Analyze weather patterns to improve the accuracy of cloud burst predictions. By studying historical data and identifying key atmospheric conditions that precede cloudbursts, we aim to refine our predictive models and enhance their reliability in forecasting these extreme weather events.

Data Enhancement: Enhance data collection and integrate real-time monitoring for better forecasting. This involves deploying additional weather sensors in strategic locations and incorporating data from multiple sources, including satellite imagery and ground-based observations, to create a comprehensive, real-time picture of atmospheric conditions.

Alert System: Provide timely and accurate alerts to relevant authorities and the public. We will develop a multi-channel alert system, including a mobile application and integration with existing emergency communication networks, to ensure that warnings reach affected populations and emergency responders as quickly as possible.

Preparedness Planning: Enable effective preparedness and evacuation measures to mitigate risks. This objective focuses on working with local authorities to develop detailed evacuation plans, identify safe zones, and create clear guidelines for community action in the event of a predicted cloudburst.

Impact Reduction: Reduce the impact of cloud bursts on vulnerable communities through early warnings.

III. LITERATURE SURVEY

A literature survey is a comprehensive review of existing research, studies, and publications related to a specific topic or area of interest. Its primary purpose is to provide an overview of the current state of knowledge on the subject, identify gaps or inconsistencies in the literature, and highlight key findings, methodologies, and theoretical frameworks that have been previously established.

The paper titled **"Simulation of a Himalayan Cloudburst Event"** by S. Das, R. Ashrit, and M. W. Moncrieff explores the use of the MM5 mesoscale model to predict a cloudburst that occurred in Shillagarh, Himalayas, on July 16, 2003. The study identifies key atmospheric processes such as wind convergence, vertical shear, and orographic uplift, providing a conceptual model for cloudburst evolution. While the model successfully predicted rainfall 24 hours in advance, it missed the precise location by a few kilometers. Despite some limitations, including the overestimation of hydrometeor content, the research underscores the potential of mesoscale models for cloudburst forecasting and highlights the sensitivity of rainfall predictions to various cloud microphysics schemes.

The second paper, **"Cloudburst Prediction in India using Machine Learning"** by D. Karunanidy et al., focuses on cloudburst prediction during the South-West Monsoon season in regions like Himachal Pradesh, Uttarakhand, and Jammu and Kashmir. The study fills a data gap by creating a comprehensive dataset with meteorological factors such as temperature, wind gusts, humidity, and cloud density. The study utilizes machine learning algorithms, with Cat Boost achieving the highest prediction accuracy of 86.18%. The research emphasizes the critical role of humidity in cloudburst occurrences, while also aiming to improve disaster preparedness in cloudburst-prone areas.

In "**SkySentinel: Harnessing AI for Cloudburst Forecasting and Warning**," A. Sebastian et al. present an AI-driven early-warning system designed to forecast cloudbursts using data science techniques. By continuously monitoring atmospheric conditions such as pressure, humidity, and temperature, SkySentinel identifies patterns that could signal an impending cloudburst. The system integrates IoT devices for real-time monitoring, providing timely alerts to vulnerable communities. While the system enhances prediction accuracy, the study also acknowledges challenges such as false positives and the limited availability of historical weather data in remote areas.

The final paper, **"Sequence Model Based Cloudburst Prediction for the Indian State of Uttarakhand"** by M. Sivagami, P. Radha, and A. Balasundaram, proposes a deep learning-based prediction model using Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) models. The model is trained on cloudburst data from Uttarakhand, with GRU outperforming other modelsThe research suggests that this deep learning approach offers a cost-effective alternative to Doppler radar and highlights the need for further exploration of additional meteorological and geographical factors.

IV. LACUNAS IN THE EXISTING SYSTEM

Current cloudburst prediction systems encounter significant obstacles that hinder effective forecasting and response. A major challenge is the integration of real-time meteorological data, particularly in mountainous regions where weather stations are sparse and connectivity is limited. This results in gaps in data collection and delays in transmission, impacting the accuracy and timeliness of predictions. Cloudbursts are highly localized and intense, often covering areas as small as 20-30 square kilometers, which existing models struggle to capture due to their insufficient spatial and temporal resolution.

In many regions prone to cloudbursts, the lack of adequate evacuation facilities exacerbates the situation. Designated evacuation centers and pre-planned routes are often missing, especially in remote mountainous areas, making it difficult to respond effectively during an emergency. Limited resources, such as vehicles and supplies for rapid evacuations, further hinder response efforts. Even when cloudbursts are predicted, significant delays in disseminating warnings can occur due to inadequate communication infrastructure, which undermines evacuation effectiveness. Public unawareness regarding cloudbursts and appropriate response measures can lead to further delays in taking action. To improve community resilience, it's essential to develop high-resolution predictive models, enhance data integration, establish comprehensive evacuation plans, and optimize communication systems for timely alerts and public education.

V. DATASET CURATION

Parameter	Correlation	P-Value	Interpretation
Rain	0.772122	6.69×10^{-71}	Strong Positive (Significant)
Precipitation	0.768599	6.95×10^{-70}	Strong Positive (Significant)
Precipitation Hours (h)	0.461585	5.63×10^{-20}	Moderate Positive (Significant)
Wind Gusts (10m)	0.439552	4.62×10^{-18}	Moderate Positive (Significant)
Wind Speed (10m)	0.232855	1.01×10^{-5}	Weak Positive (Significant)
Pressure (MSL Mean)	-0.222249	2.58×10^{-5}	Weak Negative (Significant)
Cloud Cover Mean (%)	0.166446	1.73×10^{-3}	Weak Positive (Significant)
Elevation	-0.130774	1.41×10^{-2}	Weak Negative (Significant)
Evapotranspiration (ET0, mm)	-0.122281	2.18×10^{-2}	Weak Negative (Significant)
Temperature (2m)	0.096822	6.96×10^{-2}	Weak Positive (Significant)
Apparent Temperature	0.077670	1.46×10^{-1}	Weak Positive (Not Significant)
Wind Direction (10m)	-0.050502	3.45×10^{-1}	Weak Negative (Not Significant)
Relative Humidity (2m Mean, %)	0.001562	9.77×10^{-1}	Weak Positive (Not Significant)

Table 1. Correlation Analysis of Meteorological Parameters

To understand the impact of various meteorological parameters on cloudburst events, a correlation analysis was performed. The table above summarizes the correlation coefficients, p-values, and their interpretations:

- The correlation analysis reveals that Rain (0.774) and Precipitation (0.748) have a strong positive correlation with cloudbursts, making them key indicators for prediction models.
- Wind Gusts (0.49) and Precipitation Hours (0.49) exhibit moderate positive correlations, suggesting they contribute to cloudburst conditions but are not primary factors.
- Elevation (-0.16) and Pressure (-0.16) show a weak negative correlation, indicating that cloudbursts may be less frequent in higher-altitude areas.
- Temperature (0.08) and Relative Humidity (0.001) have very low correlations, implying minimal direct influence on cloudburst occurrences.
- The p-values suggest that only the parameters with strong correlations (Rain, Precipitation) are statistically significant.

This analysis highlights the most influential meteorological factors in cloudburst events and helps refine predictive models.

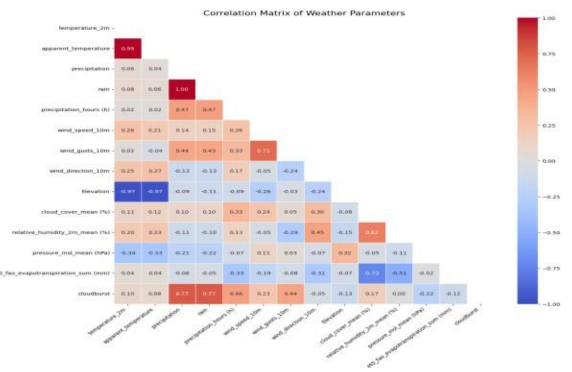


Figure 1. Correlation matrix of weather parameters (Pearson correlation coefficient)

VI. PROPOSED SYSTEM

To enhance the accuracy and timeliness of cloud burst predictions, we propose a comprehensive cloud burst prediction and alert system that integrates advanced weather modeling, machine learning algorithms, and real-time data from multiple sources. This system is anchored by a centralized monitoring center that processes incoming data and triggers automated alerts via a user-friendly website, which provides personalized notifications and evacuation guidance.

To address the challenges inherent in cloud burst prediction, several strategies can be implemented:

1. **Data Collection**: Collaborating with meteorological agencies and research institutions is essential for gathering pertinent data, both real-time and historical, which includes atmospheric conditions, rainfall patterns, and other critical variables

2. **Real-Time Monitoring**: The monitoring center will continuously analyze incoming weather data to identify emerging threats. By setting predefined thresholds, the system can issue alerts promptly when conditions indicative of a potential cloud burst are detected.

3. **Dynamic Evacuation Planning**: The system will integrate real-time traffic and weather data to provide dynamic evacuation routes. Comprehensive risk assessments will identify safe evacuation points in collaboration with local authorities, ensuring efficient response efforts.

4. **Machine Learning Models**: We will employ machine learning algorithms, particularly CatBoost, known for its ability to handle categorical and numerical data effectively. CatBoost will analyze historical data patterns alongside current weather conditions to predict the likelihood of cloud bursts. This model will be trained on relevant datasets, enabling it to recognize specific patterns indicative of impending cloud bursts.

5. **Pre-warning Messages**: Upon detecting a potential cloud burst, the system will generate pre-warning messages to alert relevant authorities and the public, allowing time for precautionary measures such as evacuation or flood preparation.

6. **Evaluation and Feedback Loop**: Post-prediction, the system will evaluate the accuracy of its alerts using performance metrics. This evaluation will inform continuous improvements to the predictive models, ensuring reliability and reducing false alarms.

By implementing this multifaceted approach, the proposed system aims to significantly mitigate the impact of cloud bursts on communities, enhancing overall preparedness and response capabilities.

VII. METHODOLOGY

This study presents the CloudEye system, designed to predict cloudburst events by integrating historical and real-time weather data with advanced machine learning techniques. The methodology encompasses several key components, including data collection, model training using the CatBoost algorithm, and system implementation through a RESTful API.

The data utilized in this study is categorized into two primary types. First, historical weather data was gathered from past cloudburst events, incorporating key meteorological variables such as rainfall, humidity, wind speed, and temperature. This dataset served as the foundation for training the CatBoost model, allowing it to learn patterns associated with cloudburst occurrences. Second, real-time weather data was obtained through the OpenWeather API, which provides current measurements of rainfall, humidity, and temperature. This real-time data is crucial for ongoing predictions and monitoring of cloudburst events.

The prediction model relies on the CatBoost algorithm, which is specifically optimized for handling categorical features and complex interactions within the data. In the training phase, the historical dataset underwent rigorous pre-processing, which included handling missing values and normalizing features to ensure the data was prepared for model training. The CatBoost model was then trained on this pre-processed historical data, which contained labels indicating the occurrence of cloudbursts. This training process enabled the model to establish meaningful relationships between various weather variables and cloudburst events.

In the inference phase, real-time data from the OpenWeather API was input into the trained CatBoost model. The model outputs a probability score that indicates the likelihood of a cloudburst event occurring. This score is subsequently compared to a predefined threshold to determine whether an alert should be issued.

The CloudEye system is implemented as a Flask web application, providing a user-friendly interface through two key routes. The /signup route facilitates user registration by collecting email addresses and city preferences for monitoring potential cloudbursts. The /predict route allows users to submit requests for cloudburst predictions based on their selected cities, querying the OpenWeather API and returning the corresponding prediction results.

Integration with the OpenWeather API is essential for retrieving meteorological parameters such as rainfall, humidity, wind speed, and temperature. These parameters are utilized as input features for the prediction model, enhancing its predictive capabilities.

To ensure timely alerts, the system employs Python's `smtplib` for sending email notifications. When the model predicts a high likelihood of a cloudburst, an alert email is generated and dispatched to registered users, providing critical information and safety recommendations.

The effectiveness of the CatBoost prediction model is evaluated using several metrics, including accuracy, precision, recall, and the F1 score. Accuracy measures the ratio of correct predictions to total predictions made. Precision assesses the proportion of true positive predictions among all positive predictions, while recall evaluates the model's capability to detect actual cloudburst events. The F1 score provides a harmonic mean of precision and recall, offering a balanced measure of model performance.

Through this comprehensive methodology, the CloudEye system aims to enhance cloudburst prediction capabilities, ultimately improving public safety and preparedness in regions vulnerable to such weather phenomena.

VIII. RESULTS

User Interface of our proposed solution

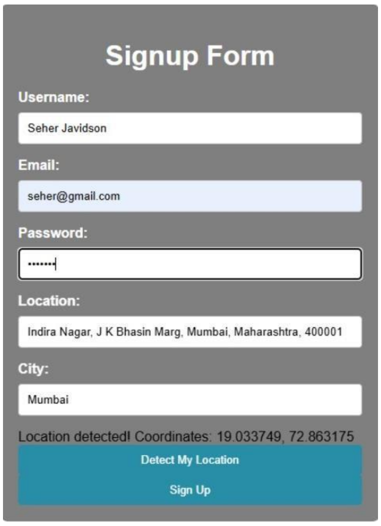


Figure 2. Signup Page

The Signup Form integrates a PostgreSQL database to store user details, including username, email, city, address, latitude, and longitude. Upon signup, the form captures the user's location using geolocation services and saves it in the database for future monitoring and alerts. This ensures that users can be notified of cloudburst conditions and directed to safe locations in real time.

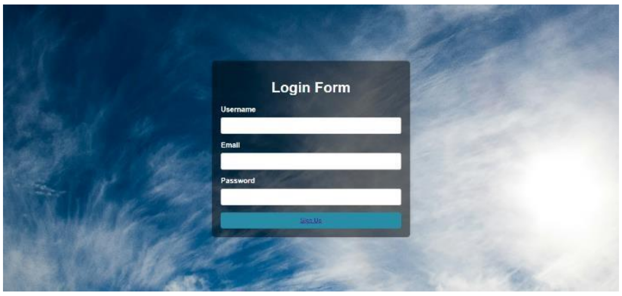


Figure 3. Login Page

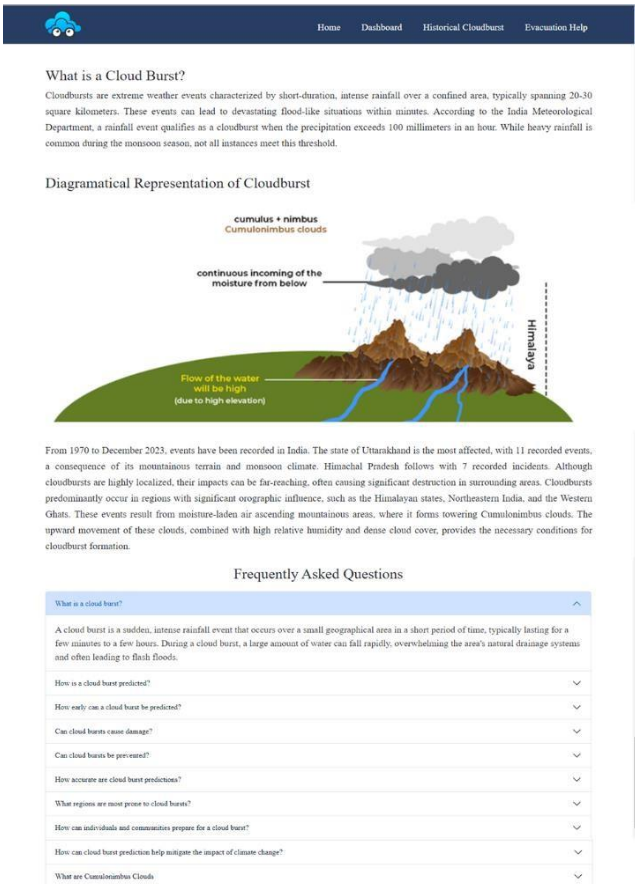


Figure 4. Home Page

The Home page provides a detailed overview of cloudbursts, including a clear explanation of what they are and their potential impacts. It features a visual diagram that illustrates the formation of cloudbursts, helping users understand the process. The page also offers historical context, highlighting regions in India most affected by cloudbursts, such as Uttarakhand and Himachal Pradesh. Additionally, there is an accordion-style FAQ section that answers common questions about cloudbursts, their prediction, impacts, and prevention. This layout allows users to gain a comprehensive understanding of cloudbursts through informative text, visual aids, and interactive elements.

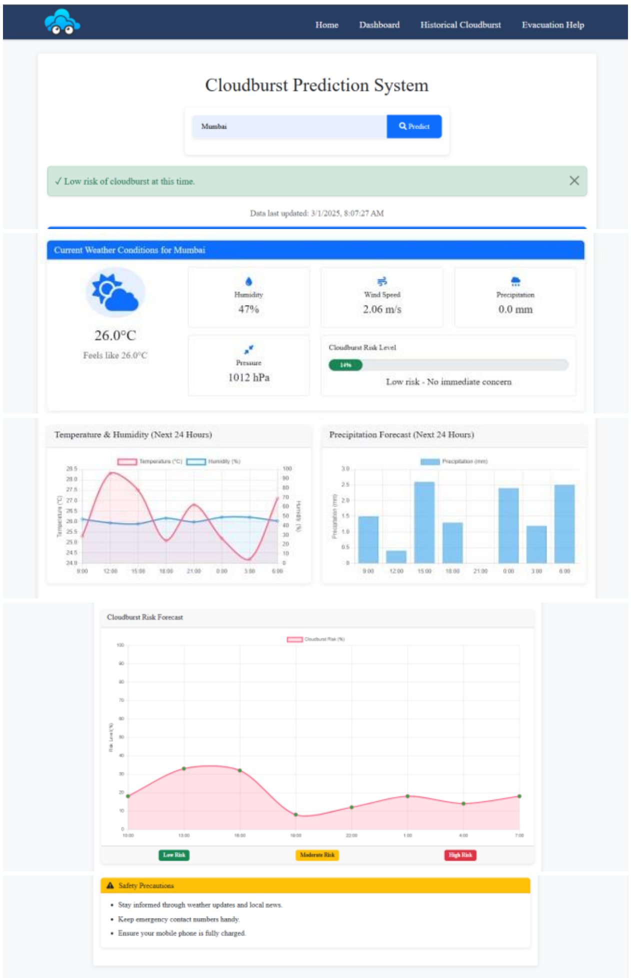


Figure 5. Dashboard

This Dashboard page provides a quick summary and realtime prediction of weather conditions, reassuring users if cloudburst is expected or not. The page highlights key parameters like apparent temperature, humidity, and precipitation through both text and a clear bar chart. Its simple design and concise messaging make it easy for users to grasp the forecast and understand the current weather conditions at a glance.

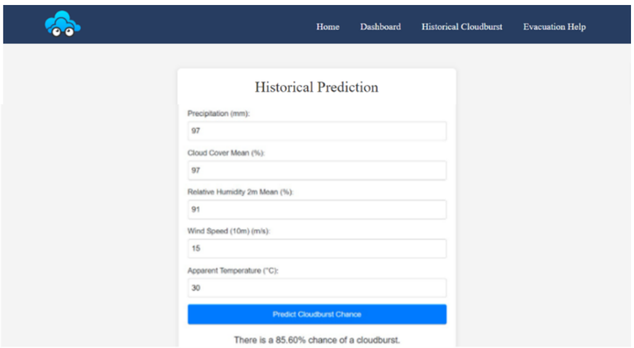


Figure 6. Historical Cloudburst Page

This Historical Cloudburst page allows users to input historical weather data to predict the likelihood of a cloudburst. The form includes fields for precipitation, cloud cover, relative humidity, wind speed, and apparent temperature. Once the user enters these values and clicks the "Predict Cloudburst Chance" button, the system calculates and displays the cloudburst probability—in this case, showing an 85.60% chance. The interface is straightforward, featuring a simple form and clear output for easy interpretation of cloudburst risks based on historical data.

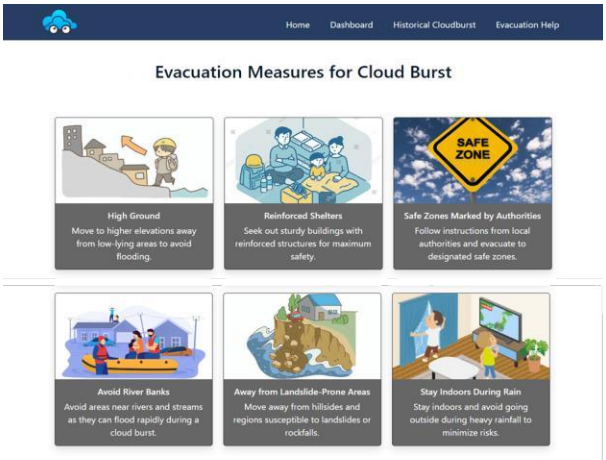
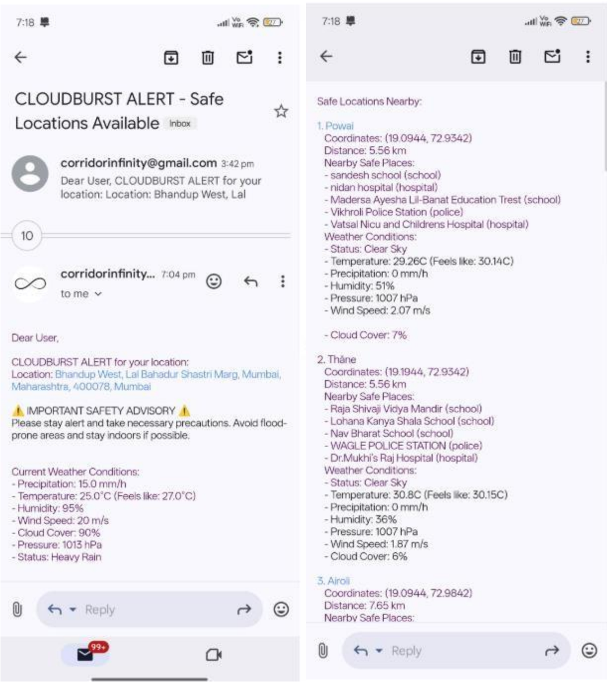


Figure 7. Evacuation Help Page

This is the Evacuation Help page offering essential guidelines for safety during a cloudburst. The page is designed with clear icons and concise text for each guideline, making it easy for users to quickly grasp and follow the evacuation measures.



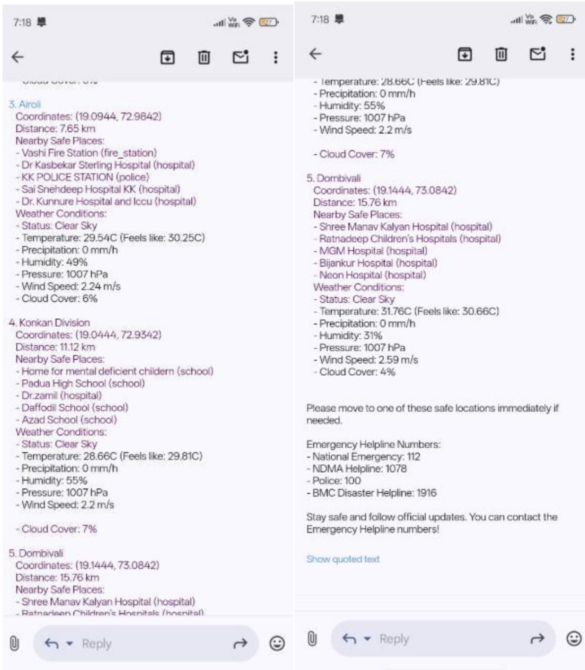


Figure 8. Alerts via Email

The Cloudburst Alert System is an innovative disaster mitigation solution that monitors meteorological conditions to predict cloudburst events and warn users approximately two hours before they occur. The system uses a PostgreSQL database hosted locally to store essential user information (email addresses, locations, and coordinates) through its fetch_users() function. When a cloudburst is predicted, the system identifies safe five locations near each user's registered coordinates - including hospitals, schools, and government buildings - and sends personalized email alerts via Gmail SMTP from **corridorinfinity@gmail.com** using the send_email() function. These alerts provide users with current weather conditions (temperature, humidity, wind speed, and pressure), detailed information about nearby safe locations organized by divisions, and emergency contact numbers. This geospatial approach to disaster management enables residents to evacuate to appropriate shelters before dangerous flash flooding occurs, potentially saving lives and reducing disaster impact in vulnerable regions.

IX. CONCLUSION

This study effectively demonstrates the potential of integrating advanced meteorological tools and machine learning techniques to enhance cloudburst prediction and monitoring. By utilizing historical and real-time weather data alongside the CatBoost algorithm, the CloudEye system significantly improves the accuracy and timeliness of cloudburst forecasts. This innovative methodology reduces dependence on conventional approaches, offering a scalable and cost-effective solution for predicting extreme weather events.

The implementation of a user-friendly Flask web application facilitates accessibility for stakeholders, including local authorities, researchers, and at-risk communities. By providing features such as real-time analysis, user registration for alerts, and detailed prediction reports, the platform empowers users to make informed decisions for disaster preparedness and community resilience. Ultimately, this research contributes to advancing both scientific understanding and practical strategies for mitigating the risks associated with cloudbursts.

X. REFERENCES

[1] D. Karunanidy, N. M., P. S. Rakshith, M. N., G. Sireesha, and M. Sreedevi, "Cloudburst Prediction in India Using Machine Learning," in Proc. 2023 IEEE Int. Conf. Big Data, 2023.

[2] A. Mittal and S. K. Khatri, "Environment Monitoring Using Internet of Things – Cloud Burst Prediction," in Proc. 2019 1st Int. Conf. Advances Inf. Technol. (AICAI), 2019, pp. 1-6, doi: 10.1109/AICAI.2019.8701364.

[3] M. Sivagami, "Sequence Model based Cloudburst Prediction for the Indian State of Uttarakhand," Disaster Advances, vol. 14, no. 7, June 2021.

[4] S. Das, R. Ashrit, and M. W. Moncrieff, "Simulation of a Himalayan cloudburst event," J. Earth Syst. Sci., vol. 115, no. 3, pp. 299–313, June 2006.

[5] A. Sebastian, J. B. Thomas, R. Mathew, R. Jose, and A. Manuel, "SkySentinel: Harnessing AI for Cloudburst Forecasting and Warning," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 1-10, doi: 10.1109/ICCPCT58313.2023.10246006.

[6] "Cloudburst," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Cloudburst.

[7] K. S. Thakur, "What are cloudbursts and why are they increasing in India?" Down to Earth, 25-Sep-2022. [Online].

[8] S. Singh and M. L. Kansal, "Cloudburst—A Major Disaster in The Indian Himalayan States," in Disaster Management and Risk Reduction in Climate Change, 2022, pp. 115-126. doi: 10.1007/978-981-16-5312-4_9.

[9] H. Chikoore, M.-J. Bopape, T. Ndarana, T. P. Muofhe, M. Gijben, R. B. Munyai, T. C. Manyanya, and R. Maisha, "Synoptic structure of a sub-daily extreme precipitation and flood event in Thohoyandou, north-eastern South Africa," Weather and Climate Extremes, vol. 33, p. 100327, 2021. doi: 10.1016/J.WACE.2021.100327.

[10] G. V. Kumar, P. Srinivas, K. Jain, A. Gairola, and R. K. Singh, "Hydrodynamic simulation of a cloudburst event in Asi Ganga Valley of Indian Himalayan region using MIKE11 and GIS techniques," Mausam, vol. 69, no. 4, pp. 523-534, 2021. doi: 10.54302/MAUSAM.V69I4.351.