Navigating predictive landscapes of cloud burst prediction approaches: insights from comparative research

Anil Hingmire[1], Sunayana Jadhav[2], Megha Trivedi[1], Karan Sankhe[1], Omkar Khanolkar[1], Yukta Patil[1]
[1]Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, University of Mumbai, Vasai, India
[2]Department of Electronics and Telecommunication Engineering, Vidyavardhini's College of Engineering and Technology, University of Mumbai, Vasai, India

Article Info

Article history:

Received Sep 26, 2024
Revised Jun 19, 2025
Accepted Jul 1, 2025

Keywords:

Cloud burst
Linear regression
Machine learning
Random forest
Support vector regression

ABSTRACT

Cloud burst forecasting remains an evolving field that grapples with the complexities of atmospheric phenomena and their impact on local environments. Cloud bursts in hilly regions demand robust predictive models to mitigate risks. This study addresses the challenge of imbalanced cloud burst occurrences, emphasizing the need for accurate predictions to minimize damage. It develops and evaluates a machine learning-based forecasting approach that includes several weather factors such as temperature, humidity, wind speed, and atmospheric pressure. The study also tackles the imbalance in cloud burst data. A dual-axis chart visually merges cloud burst occurrences with weather parameters, providing insights into their relationships over time. The model's overall accuracy is 0.68, with precision and recall for cloud burst events at 0.25 and 0.07, respectively, and an F1-score of 0.11. However, when it comes to forecasting non-cloud burst occurrences, it shows a high precision of 0.72. This study evaluates machine learning models for cloud burst prediction, highlighting random forest as the top performer with an accuracy of 85.43%, effectively balancing true positives and true negatives while minimizing misclassifications. This research contributes to cloud burst prediction, offering performance insights and suggesting avenues for future exploration.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Anil Hingmire
Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology
University of Mumbai
Vasai, Maharashtra, India
Email: anil.hingmire@vcet.edu.in

1. INTRODUCTION

Unexpected catastrophes such as earthquakes, cloud bursts, landslides, and floods are unforeseen events that inflict considerable amount of destruction and loss of life, costing billions of dollars each year. Changing weather patterns make it difficult to anticipate meteorological phenomena, and worldwide studies show that glaciers are retreating at an average pace of 10-15 meters per year as a result of climate change [1].

In a nutshell, Tiwari and Verma [2], outline how it begins with building a strong foundation by methodically gathering and analyzing past weather data. The complex study is conducted using a methodical approach intended to fully address the issues raised by cloud burst events. This entails a thorough analysis of significant factors to provide a detailed dataset that captures the varied and dynamic character of weather patterns in the area in question. To improve the system's forecasting skills, the research attempts to use a variety of characteristics, such as meteorological conditions, geographic variables, and climatic patterns [3]. This stage is essential to matching the intricate details of the unique conditions that define the region with the prediction model.

Additional endeavors comprise crucial stages of preprocessing data and training the model, where exacting methods are employed to guarantee the calibre and dependability of the dataset, augmenting the resilience of the analysis that follows. To maximize forecast accuracy, the machine learning model is carefully trained using cutting-edge techniques [4]. Dimri *et al.* [5], the last group of tasks focuses on comparison analyses and assessment measures. Robust evaluation measures are used to thoroughly assess the developed effectiveness of system. A crucial part of the research is comparing the model's predictions to past cloud burst events through a comparative analysis. Assessing the systems efficacy in practical scenarios and improving its predicting skills need careful consideration of this comparison element.

India's unique topography and climate render it susceptible to frequent cloud bursts, particularly in mountainous areas. Notable instances include the May 26, 2017, cloud burst in Tatalgaon and Bijrani, which demolished a home and killed eight domestic animals, and the August 14, 2017 event in Mangti and Malpa, which resulted in nine deaths, 18 missing people, and the loss of 51 animals. The following cloud burst on July 17, 2018, near Yamnotri, Uttarkashi, swept away a footbridge and seriously destroyed the Kali Kamli Dharamshala, demonstrating the terrible impact of natural disasters and the critical need for preparedness and mitigation techniques in susceptible places [6].

On July 19, 2018, Malari hamlet in Chamoli saw a devastating cloud burst that killed two people and damaged the Joshimath-Malari road. Padmalla-Faldiya and Mori tehsils in Uttarkashi saw even more devastation in August 2019, with 60 people killed, countless animals lost, and extensive property damage. By September, cloud bursts in Govindghat had caused considerable road damage, highlighting the critical need for comprehensive disaster planning in susceptible places [7].

The research on "Cloud burst Prediction" is essentially multifaceted, ranging from fundamental data collection and model development to rigorous training and evaluation, all of which are aimed at improving the overarching goal of improving the precision and relevance of cloud burst predictions within the designated geographic area.

2. RELATED WORK

The overview of current studies on cloud burst prediction systems may be found in the literature review that follows. The research included in this discussion concentrates on various approaches-from deep learning techniques to hardware-based systems-that aid in the comprehension and forecasting of cloud burst events in diverse geographical areas.

Tiwari and Verma [2] explored the detrimental impact of cloud bursts in the Himalayan regions and underscore the inadequacies of traditional prediction methods. The authors propose a novel cloud burst predetermination system leveraging Arduino technology, incorporating a rain gauge, float switch, and submersible pump for real-time calculation of rainfall intensity. This cost-effective solution aims to overcome the limitations of conventional methods, providing a practical and efficient monitoring approach.

Sivagami *et al.* [3] developed a cloud burst prediction model using deep learning, specifically gated recurrent unit (GRU) and long short-term memory (LSTM) networks. They applied predictive power score (PPS) to identify key features for these models. The dataset exhibited class imbalance, with 16 of 20 events being cloud bursts, a typical challenge in extreme event datasets. Sunil *et al.* [4] introduces "Predister," designed for cloud burst prediction in hilly areas, emphasizing the critical importance of early warnings in preventing the loss of life and property. The system integrates environmental sensors, data science, and artificial intelligence to monitor atmospheric conditions. Timely alerts are issued based on abnormal conditions, showcasing the potential to save lives and property in remote and vulnerable areas.

Dimri *et al.* [5] focus on cloud burst events in the southern range of the Indian Himalayas. Utilizing diverse data sources, including NASA's MERRA dataset and IMD's Rain Gauge stations, the study explores extreme precipitation patterns and large-scale factors contributing to cloud bursts. The paper provides a conceptual model for understanding these events, covering aspects such as precipitation patterns, orographic influences, and societal consequences. Reddy *et al.* [6] introduce a rainfall prediction model employing multiple linear regression (MLR) for Indian meteorological data. Emphasizing the significance of accurate rainfall predictions for industries, particularly agriculture, the research showcases the efficacy of the MLR-based approach. By incorporating multiple meteorological parameters, the study enhances prediction accuracy, offering potential benefits for various industries reliant on weather forecasting. The integration of expert systems adds a layer of sophistication to the prediction model, contributing to its practical applicability in diverse sectors.

The National Centre for Medium Range Weather Forecasting used a high-resolution WRF model to study the 2010 Leh cloud burst, which caused over 200 deaths in Ladakh. The 3 km model aligned with

TRMM satellite data, showing peak rainfall of over 4 cm in three hours. Analysis revealed the event resulted from a humid northwest flow capped by cooler, drier air, with instability triggered by a cloud cluster from Nepal [8]. Pabreja and Datta [9] demonstrated the use of k-means clustering on numerical weather prediction data to detect cloud burst signals 3-4 days in advance, using a case study of a cloud burst in Uttarakhand.

Das *et al.* [10] analyzed the Shillagarh cloud burst of July 16, 2003, which lasted less than 30 minutes and caused significant damage. Using the MM5 mesoscalmodel with nested domains (81-3 km resolution), the study successfully predicted rainfall 24 hours in advance but exhibited a location error of several kilometers. Lakshmi and Karthikeyan [11] studied K-means and spectral clustering approaches for cloud burst prediction. Their examination of specific humidity and relative humidity at various atmospheric pressure levels found that cloud bursts are most common around 925 hPa, whereas temperature data suggested development at 400 hPa, allowing for early identification of cloud burst events.

Wang *et al.* [12] presented a study on the Zhujiatang Landslide in China that creates a framework for anticipating landslide deformation phases utilizing multisource data and machine learning. The findings show that the landslide's deformation, which is most strong in the front and decreases towards the back, is closely related to rainfall patterns. The model uses the C5.0 decision tree to extract criteria, a graph convolutional network to anticipate stages, and the Morgenstern-Price approach to perform critical sliding. Overfitting concerns in C5.0, challenges in crucial sliding prediction with the Morgenstern-Price approach, and complicated knowledge representation difficulties in the random forest algorithm are among the limitations.

The study by Chen [13] focuses on improvements in landslide prediction, specifically using cutting-edge modeling strategies that include knowledge graph embedding. It discusses the rising frequency of landslides, which is made worse by climate change, and criticizes conventional forecasting techniques, which are frequently expensive and dependent on specialized knowledge. In order to increase forecast accuracy and make these methods more widely available, particularly for remote sensing applications, the authors suggest a more effective method for assessing possible landslide situations. Data-driven models have issues including overfitting, high dimensionality, and difficult feature selection, and they run the danger of oversimplifying landslide dynamics.

Schmith *et al.* [14] uses historical daily precipitation data (1914-2010) and recent hourly records to investigate geographical variations in cloud burst frequency in Denmark. They define cloud burst days based on hourly thresholds and use a binary regression model to predict cloud burst probability from daily precipitation amounts, indicating greater frequency in western Jutland. Model validation demonstrates strong predictive capability, and the data indicates that regional frequency variations are due to spatial precipitation distribution rather than variances in the model's predictive connection.

Garg *et al.* [15] evaluated high-resolution datasets-Indian Monsoon Data Assimilation and Analysis (IMDAA) and IMERG-V06B-for identifying cloud burst occurrences in the Northwest Himalayas (NWH). IMDAA successfully detects 11 of 16 cloud bursts, exceeding standard India Meteorological Department (IMD) data in areas such as Jammu and Kashmir. IMERG-V06B identifies cloud bursts with a modest probability (33.33%-63.39%), but its performance increases with time-based modifications to 41.24%-68.25%. While both datasets could monitor severe occurrences in NWH, their performance in difficult terrain remains unknown, underscoring the need for additional validation of cloud burst detection in mountainous places.

Himalayan states like Uttarakhand, Himachal Pradesh, and Jammu and Kashmir are especially vulnerable to cloud bursts due to their rugged terrain and monsoonal patterns, leading to frequent loss of life and infrastructure damage. Sati [16] explores the impact of cloud burst-caused hazards in the Uttarakhand Himalaya, including flash floods and landslides, by examining processes, impacts, and mitigation options. The study uses field visits and case studies to illustrate severe human and property losses from occurrences such as the August 2017 cloud burst. The study underlines that, while natural disasters are unavoidable, proactive actions, such as avoiding development near rivers and streams and promoting reforestation, can lessen catastrophe severity.

Karunanidy *et al.* [17] address cloud burst prediction in India by using a specific dataset and using machine learning methods. They consider temperature, wind speed, humidity, and cloud density in places prone to cloud bursts, such as the Himalayas. It tested multiple models and discovered that Cat Boost outperformed them all with 86.18% accuracy, proving machine learning's ability to forecast extreme weather occurrences even with insufficient historical data. Saha and Bera [18] relates rainfall intensity-duration (I-D) thresholds to landslide occurrences in the Garhwal Himalaya, discovering that rainfall intensities of 0.45-0.50 mm/hour over 48 hours might cause landslides, particularly when antecedent rainfall is 80 mm in 15 days. The study emphasizes the significance of early warning systems, climate assessments, and local participation, while also recommending future research areas on climate-driven threshold changes and vulnerability [18].

Hunt and Dimri [19] explore how synoptic-scale circulation, especially extratropical western disturbances (WDs) and tropical depressions (TDs), affect landslide occurrences in the Upper Indus Basin (UIB) of the western Himalaya and Karakoram. Their research showed seasonal fluctuations in landslides, with a rate of 0.05 per day in winter, attributed mostly to WDs, which improve moisture movement and precipitation. Notably, WD intensity did not link with landslide likelihood, indicating that smaller-scale orographic precipitation may also be involved. In the summer, landslides increase to 0.11 per day, with TDs from central India accounting for 60%, illustrating the importance of monsoonal flows. This study advances our understanding of meteorological interactions and landslide hazards, assisting in the development of effective mitigation techniques.

Singh and Pandey [20] study flash flood risk in Uttarakhand’s Upper Ganga Basin, a region prone to severe floods. They use GIS to analyze morphometric characteristics from SRTM DEM data and use the weighted sum approach (WSA), principal component analysis (PCA), and an integrated approach (IA) to 29 sub-watersheds. Their findings demonstrate that PCA and IA generate equivalent vulnerability assessments when evaluated against historical flood occurrences from 2018 and 2019. The study reveals five densely inhabited sub-watersheds at high risk, confirming the IA method’s dependability for flash flood assessment and the efficacy of GIS and remote sensing in hilly areas.

Mobini *et al.* [21] uses geographically precise property-level data to investigate the costs of an intense rain event that occurred on August 31, 2014, in Malmö, Sweden, with an emphasis on the interaction between pluvial flood damage and the built environment. The findings show that homes connected to combined sewage systems have much higher flood damage claims than those with separate systems, despite equal total costs, emphasizing the necessity for direct access to insurance data to improve future study.

The “Cloud burst Prediction” research project employs advanced machine learning techniques, focusing on the analysis of historical meteorological data, including precipitation patterns, temperature variations, humidity, and air pressure changes. Vijaykumar *et al.* [22] examine the devastating floods in Kerala during the 2018 and 2019 monsoons, identifying anomalously warm sea temperatures and unstable atmospheric conditions as key factors in the unprecedented mesoscale cloud burst event of 2019. Chauhan *et al.* [23] apply extreme value distribution and the markov chain approach to forecast precipitation patterns in their study of hydrometeorological data from the Yamuna River Basin. While artificial intelligence approaches have limits, the markov chain model predicts rainfall with a 79.17% accuracy, emphasizing the need of understanding return times for drought and flood threats.

Knös *et al.* [24] explores the development of a cloud burst catastrophe model in Jönköping, Sweden, using rainfall intensity as a direct hazard measure to assess urban vulnerability and establish a time-sensitive vulnerability curve, suggesting potential applicability in various geographical contexts despite uncertainties due to limited loss data. From 2001 to 2018, more than 5,000 water-related disasters (WRDs), including floods and droughts, accounted for 73.9% of all natural disasters, killing over 300,000 people and causing $1.7 trillion in economic damage worldwide. The frequency and intensity of these disasters have increased in the twenty-first century, with approximately $600 billion lost owing to over 2,900 floods and 290 droughts, which have had a severe impact on the health of 2.8 billion people, including nearly 300,000 flood injuries [25]. Cloud bursts, which feature quick and strong rainfall, can cause flash floods in metropolitan areas due to insufficient drainage capacity. To solve this, Hingmire and Bhaladhare [26] created an IoT-based urban flood management system with fuzzy logic that adapts in real time to rainfall intensity, water level, and flow rate. Their method reduced water levels by up to 73.9% during extreme conditions, indicating that it is a potential technique for mitigating cloud burst-related flooding in smart cities.

3. MOTIVATIONS AND PROBLEM STATEMENT

Despite the growing frequency of cloud bursts due to climate change, there is a lack of effective forecasting and monitoring systems that can predict these events and mitigate their impacts. This study aims to resolve this gap by developing a predictive model that utilizes meteorological data and advanced analytics to forecast cloud bursts more accurately.

By focusing on understanding the underlying causes and patterns of cloud bursts, this study aims to create a framework that enhances early warning systems and informs disaster preparedness initiatives. Ultimately, the goal is to improve community resilience to extreme weather events, ensuring that vulnerable populations are better equipped to respond to the challenges posed by cloud bursts.

4. IMPLEMENTATION METHODOLOGY

This study investigates the predictive accuracy of four machine learning algorithms-linear regression, support vector machine (SVM), random forest and decision tree-in predicting cloud burst events. A comprehensive study in which meteorological data was systematically collected from a curated and authoritative from Kaggle. The research focuses on the rigorous acquisition and analysis of this dataset, exploring its potential applications in diverse domains. By leveraging the wealth of information available on this meteorological platform, we aim to contribute valuable insights to the scientific community and advance the understanding of climatic patterns, ultimately fostering innovations in weather-related applications and decision-making processes. Sudden atmospheric changes pose challenges, and the spatial-temporal resolution of the model requires region-specific adaptations for optimal performance. The features selected to train the model are minimum temperature, rainfall, wind-gust speed, humidity 9am, humidity 3pm, pressure 9am, pressure 3pm, cloud 9am, cloud 3pm.

The dataset was split into training (80%) and testing (20%) sets to evaluate generalizability. Each model was implemented using Python's scikit-learn library, and hyperparameters were optimized through grid search with five-fold cross-validation on the training data to ensure each algorithm's optimal performance.

a) Model selection and experimental setup

Algorithms:

- Logistic regression was selected as a baseline linear model to understand its predictive performance with a probabilistic approach to classification. In the realm of cloud burst prediction, this method assumes that there is a direct and proportional relationship between different weather variables (such as temperature, humidity, and wind speed) and the likelihood of cloud burst events occurring. When such a linear relationship does exist, linear regression can provide reasonably accurate predictions, with an achievable accuracy range typically falling between 50% to 60%.
- Random forest was used for its ensemble-based architecture, which enhances prediction accuracy through multiple decision trees and mitigates overfitting.
- SVM was implemented for its margin maximization, offering an optimal hyperplane for binary classification of cloud burst and non-cloud burst events.
- Decision tree provides an interpretable model to assess the impact of individual features on cloud burst prediction.

Implementation:

- The dataset was split into training (80%) and testing (20%) sets to evaluate generalizability.
- Each model was implemented using Python's scikit-learn library, and hyperparameters were optimized through grid search with five-fold cross-validation on the training data to ensure each algorithm's optimal performance.

b) Model training and hyperparameter tuning

Each algorithm was trained on the training set, with the following hyperparameters fine-tuned for optimal accuracy:

- Logistic regression: regularization strength was varied to find an optimal balance between underfitting and overfitting.
- Random forest: number of trees and maximum depth were tuned to ensure robust feature selection while avoiding excessive computational cost.
- SVM: the kernel function and regularization parameter C were optimized to find the best hyperplane for classification.
- Decision tree: maximum depth and minimum samples per split were adjusted to control model complexity and enhance generalizability.

The results of each model, including metric scores and classification accuracy, were compared to identify the most effective machine learning approach for cloud burst prediction. Additionally, each model's performance was analyzed in terms of computational efficiency and interpretability, providing insights into the practical feasibility of each approach in real-time prediction systems.

5. RESULTS AND DISCUSSION

We obtained the confusion matrix of trained model, as depicted in Figure 1, serves as a crucial tool for evaluating the performance of a cloud burst prediction model. Based on the Figure 1(a) confusion matric of random forest model, the model achieves a high True Positive rate and relatively low false positive and false negative rates. It has a good balance of correct predictions in both positive and negative classes, indicating it performs well on both. Figure 1(b) shows that the SVM has the highest true positive rate, which suggests it excels at identifying positive cases. However, it has a relatively high false negative rate, which might indicate it occasionally fails to identify actual positives correctly. It has the lowest false positive (FP), which suggests minimal over-prediction. Figure 1(c) shows that the logistic regression has significantly lower TP and TN values, likely indicating it struggles with overall prediction accuracy in this dataset. The low true positive (TP) suggests it fails to detect positives effectively, and a low true negative (TN) indicates similar issues with negatives. Figure 1(d) shows that the decision tree has a high TP, but the highest FP among the models. This suggests that while it captures positive cases reasonably well, it also misclassifies negatives as positives more frequently than others.

Training Set- Random Forest

TARGET / OUTPUT	Positive	Negative	SUM
Positive	8368 74.16%	431 3.82%	8799 95.10% 4.90%
Negative	1213 10.75%	1272 11.27%	2485 51.19% 48.81%
SUM	9581 87.34% 12.66%	1703 74.69% 25.31%	9640 / 11284 85.43% 14.57%

(a)

Training Set- Support Vector Machine

TARGET / OUTPUT	Positive	Negative	SUM
Positive	8593 76.15%	206 1.83%	8799 97.66% 2.34%
Negative	1588 14.07%	897 7.95%	2485 36.10% 63.90%
SUM	10181 84.40% 15.60%	1103 81.32% 18.68%	9490 / 11284 84.10% 15.90%

(b)

Training Set- Logistic Regression

TARGET / OUTPUT	Positive	Negative	SUM
Positive	238 59.50%	41 10.25%	279 85.30% 14.70%
Negative	103 25.75%	18 4.50%	121 14.88% 85.12%
SUM	341 69.79% 30.21%	59 30.51% 69.49%	256 / 400 64.00% 36.00%

(c)

Training Set- Decision Tree

TARGET / OUTPUT	Positive	Negative	SUM
Positive	7572 67.10%	1227 10.87%	8799 86.06% 13.94%
Negative	1151 10.20%	1334 11.82%	2485 53.68% 46.32%
SUM	8723 86.80% 13.20%	2561 52.09% 47.91%	8906 / 11284 78.93% 21.07%

(d)

Figure 1. Confusion matrix of models; (a) random forest model, (b) support vector machine model, (c) logistic regression model, and (d) decision tree model

Figure 2 shows a Winrose chart of wind speed illustrates the increasing magnitude of the wind speed. The Winrose chart provides a comprehensive and visually intuitive overview of the prevailing weather conditions, enabling users to quickly grasp the overall weather pattern.

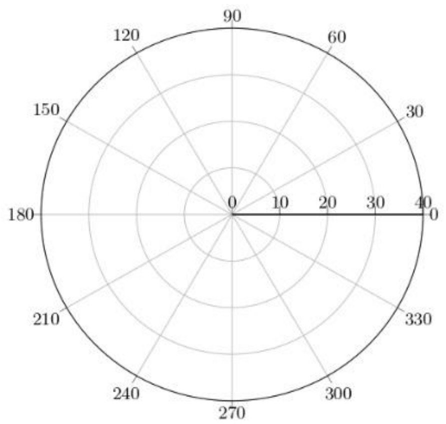

Figure 2. Winrose chart

Figure 3 shows a heat map that provides a visual snapshot of the correlations among key weather parameters, shedding light on their interconnectedness and potential implications for cloud burst occurrences. Temperature's influence is discerned through its correlations with other factors, indicating that higher temperatures may contribute to cloud burst events. Crucial to cloud burst likelihood, humidity levels are showcased in the heatmap, highlighting their interplay with various parameters.

Figure 4 shows a dual-axis chart that merges cloud burst occurrences with key weather parameters, offering a comprehensive view of their interrelationships over time. Temperature trends, humidity levels, wind speed patterns, cloud cover variations, atmospheric pressure, and precipitation intensity are all visually represented.

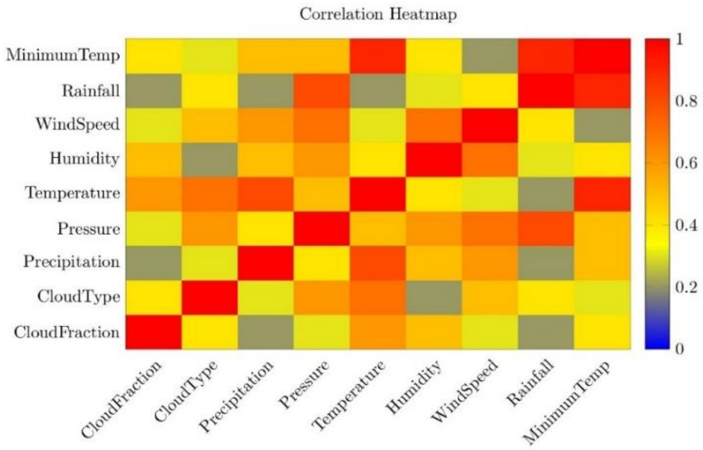

Figure 3. Heat map for correlations among key weather parameters

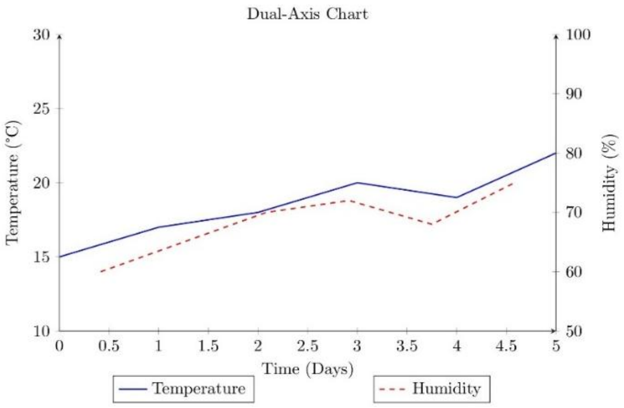

Figure 4. Cloud burst occurrences of random forest

Table 1 shows a comparison of models based on accuracy, recall, and F1-score. The random forest model performs best, with high accuracy (0.9510) and recall (0.8734) for the positive class, resulting in an F1-score of 0.9106. While the SVM has the best positive class accuracy (0.9766), its lower recall (0.8440) suggests occasional misses. Logistic regression scores poorly across all criteria, whereas decision tree achieves reasonable balance but falls short of random forest and SVM. Thus, random forest produces the most consistent results for this dataset.

Table 1. Comparative analysis of implemented models

Models	Precision		Recall		F1-values	
class	Positive	Negative	Positive	Negative	Positive	Negative
Random forest	0.9510	0.5119	0.8734	0.7469	0.9106	0.6074
Support vector machine	0.9766	0.3610	0.8440	0.8132	0.1560	0.9055
Logistic regression	0.8530	0.1488	0.6979	0.3051	0.7677	0.2000
Decision tree	0.8606	0.5368	0.8680	0.5209	0.8643	0.5287

Table 2 shows the comparison of the accuracy for each prediction model. The random forest model achieves the highest accuracy at 85.43%, closely followed by the SVM at 84.10%. The decision tree performs moderately well with an accuracy of 78.93%, while the logistic regression model has the lowest accuracy at 64.00%.

Table 2. Performance analysis of models

Model name	Accuracy
Random forest	85.43%
Support vector machine	84.10%
Logistic regression	64.00%
Decision tree	78.93%

6. CONCLUSION

The study demonstrates that the random forest model, with an accuracy of 85.43%, out performs other machine learning approaches (SVM, decision tree, and logistic regression) in predicting cloud burst events. The confusion matrix analysis highlights random forest’s effective balance of true positives and true negatives, minimizing both false positives and false negatives. Its strong precision and recall values in the positive class reinforce its ability to accurately detect potential cloud bursts while maintaining a moderate error rate for negative cases. SVM, with a similar high accuracy (84.10%), shows excellent true positive rates and low false positives, though it has a slightly higher false negative rate, suggesting some missed positives. Conversely, logistic regression struggles significantly with an accuracy of 64.00%, making it unsuitable for this application. The decision tree achieves moderate accuracy (78.93%) but suffers from high false positives, which can impact reliability. Overall, random forest emerges as the most balanced and reliable model for cloud burst prediction in this study, providing a robust tool for enhancing early warning systems. Future research could enhance prediction accuracy by exploring additional ensemble methods, hyperparameter tuning, and using larger, diverse datasets with real-time weather data for improved model generalization. Incorporating deep learning techniques and deploying models within real-time systems in collaboration with meteorological agencies could strengthen cloud burst prediction strategies and public safety.

ACKNOWLEDGEMENTS

The authors acknowledge the institutional support and resources that facilitated the successful completion of this research.

FUNDING INFORMATION

The authors declare that no funding or financial support was received from any external funding agency for the research reported in this article.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anil Hingmire	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Sunayana Jadhav				✓	✓	✓		✓		✓	✓			
Megha Trivedi				✓	✓	✓	✓			✓	✓		✓	
Karan Sankhe		✓	✓		✓		✓	✓						
Omkar Khanolkar		✓	✓			✓		✓			✓			
Yukta Patil		✓	✓				✓				✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

DATA AVAILABILITY

The meteorological data used in this study was sourced from Kaggle, a public data platform. The dataset was selected from a curated and authoritative collection relevant to the study objectives. All data used is available in the public domain and can be accessed freely for research and educational purposes.

REFERENCES

[1] A. M. Hingmire and P. R. Bhaladhare, "A review on urban flood management techniques for the smart city and future research," in *Intelligent Cyber Physical Systems and Internet of Things Engineering Cyber-Physical Systems and Critical Infrastructures*, 2023, pp. 303–317.
[2] A. Tiwari and S. K. Verma, "Cloud burst predetermination system," 2015.
[3] M. Sivagami, P. Radha, and A. Balasundaram, "Sequence model based cloudburst prediction for the Indian state of Uttarakhand," *Disaster Advances*, vol. 14, no. 7, pp. 1–9, Jun. 2021, doi: 10.25303/f2512105.
[4] A. Sunil, B. A. Binny, J. Benny, R. Rajeev, and M. A. Manuel, "Predister ('An Intelligent Device To Predict Cloud Burst')," *Predister: An Intelligent Device to Predict Cloud Burst*, vol. 7, no. 6, p. 2020, 2022, doi: 10.6084/m9.doione.IJRTI2206303.
[5] A. P. Dimri et al., "Cloud burst s in Indian Himalayas: A Review," *Earth-Science Reviews*, 2016.
[6] G. B. Reddy, J. Chethan, and M. Saravanamuthu, "Cloud burst forecast using expert systems," *International Research Journal of Engineering and Technology*, 2022, [Online]. Available: www.irjet.net.
[7] S. Khanduri, "Cloudbursts over Indian Sub-continent of Uttarakhand Himalaya: a traditional habitation input from Bansoli, District-Chamoli, India," *International Journal of Earth Sciences Knowledge and Applications*, vol. 2, no. 2, pp. 48–63, 2020.
[8] R. Ashrit, "Investigating the Leh 'Cloud burst," *National Centre For Medium Range Weather Forecasting Ministry Of Earth Sciences Government Of India*, 2010.
[9] K. Pabreja and R. K. Datta, "Clustering technique for interpretation of cloudburst over Uttarakhand," *Mausam*, vol. 67, no. 3, pp. 669–676, Jul. 2016, doi: 10.54302/mausam.v67i3.1386.
[10] S. Das, R. Ashrit, and M. W. Moncrieff, "Simulation of a Himalayan cloudburst event," *Journal of Earth System Science*, vol. 115, no. 3, pp. 299–313, Jun. 2006, doi: 10.1007/BF02702044.
[11] D. lakshmi and J. Karthikeyan, "Weather impact over Uttarakhand using k-means clustering technique for cloudburst prediction," *International Journal of Data Mining Techniques and Applications*, vol. 5, no. 2, pp. 148–149, Dec. 2016, doi: 10.20894/IJDMTA.102.005.002.010.
[12] X. Wang, M. Diao, H. Guo, L. Wang, H. Guo, and D. Li, "Landslide deformation prediction and automatic warning by coupling machine learning and physical models," *Earth and Space Science*, vol. 11, no. 6, Jun. 2024, doi: 10.1029/2023EA003238.
[13] L. Chen, L. Peng, and L. Yang, "Improving landslide prediction: innovative modeling and evaluation of landslide scenario with knowledge graph embedding," *Remote Sensing*, vol. 16, no. 1, p. 145, Dec. 2024, doi: 10.3390/rs16010145.
[14] T. Schmith, P. Thejll, F. Vejen, and B. Christiansen, "Regional variation of climatological cloudburst frequency estimated from historical observations of daily precipitation sums," *International Journal of Climatology*, vol. 43, no. 16, pp. 7761–7774, Dec. 2023, doi: 10.1002/joc.8291.
[15] S. Garg, P. Jena, U. Devi, and S. Azad, "Performance evaluation of high-resolution IMDAA and IMERG for detecting cloudburst events over the Northwest Himalayas," *International Journal of Climatology*, vol. 43, no. 8, pp. 3730–3748, Jun. 2023, doi: 10.1002/joc.8055.
[16] V. P. Sati, "Cloudburst-triggered natural hazards in Uttarakhand Himalaya: mechanism, prevention, and mitigation," *International Journal of Geological and Environmental Engineering*, vol. 12, no. 1, pp. 45–48, 2018.
[17] D. Karunanidy, P. S. Rakshith, G. Sireesha, M. Nandhini, N. MageshKumar, and M. Sreedevi, "Cloudburst prediction in India using machine learning," in *Proceedings of the 2023 6th International Conference on Recent Trends in Advance Computing, ICRTAC 2023*, Dec. 2023, pp. 645–650, doi: 10.1109/ICRTAC59277.2023.10480754.
[18] S. Saha and B. Bera, "Rainfall threshold for prediction of shallow landslides in the Garhwal Himalaya, India," *Geosystems and Geoenvironment*, vol. 3, no. 3, p. 100285, Aug. 2024, doi: 10.1016/j.geogeo.2024.100285.
[19] K. M. R. Hunt and A. P. Dimri, "Synoptic-scale precursors of landslides in the western Himalaya and Karakoram," *Science of the Total Environment*, vol. 776, p. 145895, Jul. 2021, doi: 10.1016/j.scitotenv.2021.145895.
[20] G. Singh and A. Pandey, "Flash flood vulnerability assessment and zonation through an integrated approach in the Upper Ganga Basin of the Northwest Himalayan region in Uttarakhand," *International Journal of Disaster Risk Reduction*, vol. 66, p. 102573, Dec. 2021, doi: 10.1016/j.ijdrr.2021.102573.
[21] S. Mobini, E. Nilsson, A. Persson, P. Becker, and R. Larsson, "Analysis of pluvial flood damage costs in residential buildings – A case study in Malmö," *International Journal of Disaster Risk Reduction*, vol. 62, p. 102407, Aug. 2021, doi: 10.1016/j.ijdrr.2021.102407.
[22] P. Vijaykumar *et al.*, "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*, vol. 33, p. 100339, Sep. 2021, doi: 10.1016/j.wace.2021.100339.
[23] P. Chauhan, M. E. Akiner, R. Shaw, and K. Sain, "Forecast future disasters using hydro-meteorological datasets in the Yamuna river basin, Western Himalaya: using markov chain and LSTM approaches," *Artificial Intelligence in Geosciences*, vol. 5, p. 100069, Dec. 2024, doi: 10.1016/j.aiig.2024.100069.
[24] D. Knös, K. Karagiorgos, J. Haas, B. Blumenthal, L. Nyberg, and S. Halldin, "Cloudburst-disaster modelling. A new open-source catastrophe model," *International Journal of Disaster Risk Reduction*, vol. 67, p. 102679, Jan. 2022, doi: 10.1016/j.ijdrr.2021.102679.
[25] J. Lee, D. Perera, T. Glickman, and L. Taing, "Water-related disasters and their health impacts: a global review," *Progress in Disaster Science*, vol. 8, p. 100123, Dec. 2020, doi: 10.1016/j.pdisas.2020.100123.
[26] A. M. Hingmire and P. R. Bhaladhare, "Advance urban flood control system using fuzzy logic and internet of things (IoT) for smart city," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 6 S, pp. 447–456, Jun. 2023, doi: 10.17762/ijritcc.v11i6s.6952.

BIOGRAPHIES OF AUTHORS

Dr. Anil M. Hingmire is currently working as, Assistant Professor in the Computer Engineering Department at Vidyavardhini's College of Engineering and Technology, affiliated with Mumbai University, India. He holds a Ph.D. in Computer Science and Engineering from Sandip University, Nashik, and has 23 years of teaching experience. His research focuses on artificial intelligence, computer security, cryptography, and the internet of things. He has over 26 research papers publications in national, international conferences and reputed journals. He also has 3 published Indian patent and 1 copyright to his credit. He can be contacted at email: anil.hingmire@vcet.edu.in.

Dr. Sunayana Jadhav is currently working as, Associate Professor, in Dept. of EXTC at VCET, Mumbai, India. She has completed her Doctoral degree in Electronics Engg. from Mumbai, India. She has over 22 years of teaching and research experience. She is a Chairperson and Member of syllabus setting committee for various subjects in Mumbai University for Electronics and Telecomm. Engg. and Electronics and Computer Engineering. Her area of research includes digital communication, IoT, VLSI and wireless sensor network. She has over 20 publications in international, national journals and conferences. She also has 2 Indian patents published to her credit. She can be contacted at email: sunayana.jadhav@vcet.edu.in.

Dr. Megha Trivedi is working as an Associate Professor, in Dept. of Computer Engineering at VCET, Mumbai, India. She has completed her Doctoral degree in Systems and Control Engineering from IIT Bombay, India. She has over 22 years of teaching experience. Her area of research includes control systems, optimization and machine learning. She has over 15 publications in national, international conferences and journals. She also has 1 published Indian patent and 7 copyrights to her credit. She is also a member of IEEE and ISTE. She can be contacted at email: megha.trivedi@vcet.edu.in.

Karan Sankhe is currently pursuing the Bachelor of Computer Engineering degree at Vidyavardhini's College of Engineering and Technology, Vasai, India. He has participated in and won various Hackthon software development competitions. He is Tech Enthusiast, and his area of interest is machine learning, and generative AI. He is an expertise in building dynamic, user-centric applications using modern frameworks and technologies. He can be contacted at email: karan.211533108@vcet.edu.in.

Omkar Khanolkar is currently pursuing the Bachelor of Computer Engineering degree at Vidyavardhini's College of Engineering and Technology, Vasai, India. He has participated in and won various Hackthon software development competitions. His area of interest is cyber security, machine learning, and web development. He is a founder member of WEBX.VCET club. He can be contacted at email: omkar.211313105@vcet.edu.in.

Yukta Patil is currently pursuing the Bachelor of Computer Engineering degree at Vidyavardhini's College of Engineering and Technology, Vasai, India. Her area of interest is enterprise web application development, AI-powered solutions, and mobile app development, with a knack for delivering user-centric and efficient solutions. She contributes to dynamic projects like real-time stock prediction chatbots, cross-platform mobile apps, and AIML-based predictive systems. She can be contacted at email: yukya.211463205@vcet.edu.in.