ISSN: 2633-4828 Vol. 5 No.3, September, 2023

International Journal of Applied Engineering & Technology

AN EFFICIENT TECHNIQUE FOR THUNDERSTORM NOWCASTING IN INDIAN NORTH OCEAN REGION

Navdeep Kanwal[1], Lovedeep Kaur[2] and Ashok Kumar Bathla[3]
[1,2]Punjabi University, Patiala,
[3]Yadavindra Department of Engineering, Punjabi University, Patiala
navdeepkanwal@gmail.com, lovedeepkaur2706@gmail.com, ashokashok81@gmail.com

ABSTRACT

Thunderstorms represent one of the most dynamic and potentially destructive weather phenomena, especially in regions like the Indian North Ocean, where their occurrence is influenced by complex interactions among various atmospheric, oceanic, and climatic variables. Accurate and timely prediction, or nowcasting, of thunderstorms is crucial for mitigating their adverse impacts on human life, property, and economic activities. Traditional forecasting methods often face challenges in terms of computational complexity, data quality, and real-time applicability. This paper introduces a hybrid approach for thunderstorm nowcasting that leverages unsupervised and supervised machine learning techniques— specifically, k-means clustering for anomaly detection and labeling, and random forest algorithms for testing and training. The model's ability to produce precise and timely thunderstorm predictions is demonstrated by the study's findings. Early detection of possible thunderstorm events is made possible by the effectiveness of the k-means clustering technique to find abnormal events. The model may be implemented in real-time and requires less processing time and resources than other approaches.

Keywords*—Thunderstorm, Nowcasting, Machine learning, anomaly detection, k means clustering, Random forest*

I. INTRODUCTION

The environmental phenomenon known as thunderstorms, are defined by the presence of cumulonimbus clouds, which result in lightning, thunder, heavy rainfall, and frequently high winds. Tornadoes, hail, and flash flooding can also be brought on by these storms. When warm, humid air climbs quickly through an unstable atmosphere, it cools and condenses water vapor into clouds. This is how thunderstorms are formed. Latent heat is released during this process, which accelerates the storm's growth. Thunderstorms can range in strength from modest storms to major systems capable of creating widespread damage. They usually span a few minutes to several hours. A mesocyclone, or rotating updraft, known as a supercell, is a vast, organized system that is frequently linked to the strongest thunderstorms. Supercell thunderstorms are especially hazardous because they have the potential to create tornadoes, powerful winds, and big hail (Doswell, et al. (2001)). In India, thunderstorms can have serious consequences, including fatalities, destruction of infrastructure, interruption of services, and financial loss. India's varied terrain, which includes the Himalayas, coastal regions, and large plains, affects the severity and frequency of these events and ads to the intricate weather processes that cause thunderstorms. Lightning strikes are the main reason behind thunderstorms' high death rate in India. Hailstorms frequently result in agricultural destruction, causing farmers to suffer large financial losses. Millions of dollars were thought to have been lost economically, which made things worse for farmers and added to the region's agricultural unrest. Power lines, buildings, and roads can all sustain significant damage from severe thunderstorms, which are frequently accompanied by high winds. Throughout 2018, there were numerous dust storms and thunderstorms in northern India, especially in Rajasthan and Uttar Pradesh, which caused extensive home devastation, severe power outages, and the deaths of over 150 people (NDRF, 2018). This incident brought to light how susceptible both urban and rural infrastructure is to severe weather. Flash flooding is a serious problem in cities like Mumbai, Kolkata, and Chennai, and it can be caused by thunderstorms.

A. Case Studies

In June 2020, the state of Bihar was struck by a number of thunderstorms accompanied by powerful lightning strikes, which left over a hundred people dead in a single day (Dash & Sahoo, 2021). This incident highlights how susceptible rural populations are to lightning strikes, especially in the monsoon season. The study by Dash and Sahoo, 2021 emphasizes the deficiency of awareness and readiness in rural areas, where improved early warning systems and public education on lightning protection may have prevented numerous lives.

State of Odisha 2019's Cyclone Fani and Related Thunderstorms: Severe thunderstorms that accompanied Cyclone Fani, which made landfall in Odisha in May 2019, intensified the storm's effects. The cyclone caused extensive damage, with thunderstorms playing a part in the state's electrical infrastructure damage and tree uprooting. The combination of the cyclone and related thunderstorms, according to Mohapatra et al. (2020), resulted in protracted power outages, interfered with communication networks, and hampered rescue efforts, underscoring the combined consequences of thunderstorms during tropical cyclones.

Dust storm and thunderstorm in Delhi-NCR (2018): In May 2018, Delhi-NCR saw severe disruption due to a strong dust storm and thunderstorm. In addition to bringing down trees and causing traffic and flight disruptions, the storm claimed multiple lives (IMD, 2019). This occurrence was part of a number of extreme weather events in northern India that were reported by the Indian Meteorological Department.

B. Thunderstorm Nowcasting

A crucial meteorological activity called "thunderstorm nowcasting" is to forecast the presence, intensity, and trajectory of thunderstorms since severe weather phenomena like lightning, heavy rain, hail, and high winds can pose serious risks to people's lives, property, and infrastructure, accurate nowcasting is crucial to reducing their effects. The dynamic and localized nature of thunderstorms is the reason behind the intricacy of thunderstorm nowcasting. Thunderstorms, in contrast to large-scale weather systems, can form quickly and are frequently impacted by minute atmospheric processes. While they work well for long-term forecasts, traditional NWP models frequently have trouble with the high temporal and spatial resolution needed for nowcasting. Due to this restriction, many strategies have been adopted, such as using radar, satellite data, and machine learning (Sun et al., 2014). Doppler radar and geostationary satellites are two recent examples of remote sensing technology breakthroughs that have given meteorologists access to high-resolution data that is crucial for nowcasting. For instance, Doppler radar provides real-time data on wind patterns, storm structure, and precipitation intensity, making it possible to identify thunderstorm development early (Ruzanski, Chandrasekar, & Wang, 2011). In a similar vein, satellite observations have the ability to track cloud motions and observe atmospheric conditions that facilitate the production of thunderstorms (Amjad et al., 2021).The importance of ML approaches in thunderstorm nowcasting has also increased. Compared to conventional methods, these techniques are more accurate in identifying patterns, analyzing big amounts of dataset from multiple sources, and making predictions. CNNs have been applied to satellite imagery processing in order to identify and forecast thunderstorm activity (Amjad et al., 2021). Researchers have created hybrid models that greatly improve nowcasting capabilities by fusing machine learning algorithms with conventional meteorological data (Schultz, Correia, & Stephenson, 2015). Even with these improvements, there are still obstacles in the way of obtaining extremely accurate thunderstorm nowcasting. Research is still being done on the inherent uncertainty in weather systems, the necessity of processing data in real-time, and the challenges of precisely forecasting the position and timing of thunderstorms. Advances in thunderstorm nowcasting require constant attention to data quality, algorithm development, and computer efficiency. Thunderstorm Nowcasting techniques have been advanced due to combination of satellite, radar and ground based dataset. Systems for providing short-term forecasts have been developed by a number of organizations, including the India Meteorological Department. These systems often make use of NWP models in addition to radar and satellite data. There are still several research gaps in spite of these developments. In remote locations with a low number of meteorological stations, for example, issues with data availability and quality might make nowcasting systems less accurate. Improved feature selection and extraction techniques are also required in order to take into consideration the complicated atmospheric conditions specific to the Indian

subcontinent as well as regional climate variability. Even while nowcasting has made use of machine learning, more reliable models are still required that can incorporate real-time data streams and measure uncertainty. Furthermore, there is still much to be done to enhance the way that prediction uncertainty is communicated to the public and stakeholders. Resolving these gaps could improve thunderstorm nowcasting in India considerably in terms of reliability and precision, which would improve preparedness and reaction to disasters.

C. Datasets

In order to precisely forecast the formation, motion, and intensity of thunderstorms in a short amount of time, thunderstorm nowcasting uses a range of data sources. Numerical weather prediction (NWP) models, radar observations, satellite imaging, and ground-based weather stations are some of these data sources. The distinct information derived from each of these sources improves the overall precision and accuracy of thunderstorm nowcasting.

Radar Observations: One of the most important instruments for predicting thunderstorms is Doppler radar. It offers real-time, high-resolution data on wind speeds inside thunderstorms, storm structure, and precipitation intensity. Meteorologists can identify and track storm cells, evaluate their severity, and forecast their evolution because to Doppler radar's capacity to detect changes in wind patterns and precipitation (Wilson et al., 1998).According to Lakshmanan et al. (2007), the information obtained from radar observations is especially useful for locating mesocyclones, hail cores, and possible tornado forms inside thunderstorms. Figure 1 shows the interface of Radar placed at Delhi by IMD.

Satellite Images: Continuous monitoring of the Earth's atmosphere is provided by geostationary satellites, like the Geostationary Operational Environmental Satellite (GOES) series, which provides important information for thunderstorm nowcasting. In order to observe cloud formation, detect convective systems, and assess atmospheric conditions favorable to thunderstorm formation, satellites record visible, infrared, and water vapor imagery (Amjad et al., 2021). Radar measurements may not yet reveal fast emerging convective storms, but satellite data's great temporal resolution makes this possible. Figure 2 shows the cloud top brightness temperature of Asia reason for a specific period of time.

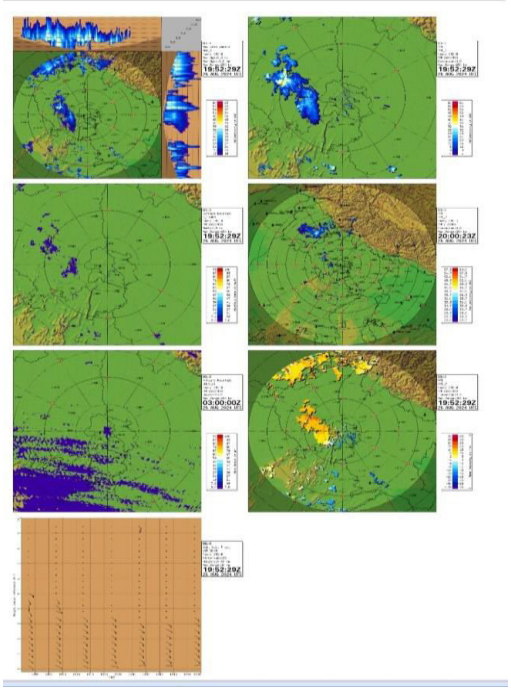

Fig 1 Radar-Delhi shows the radar data with time frame of 3 hrs Source: Adapted from [7]

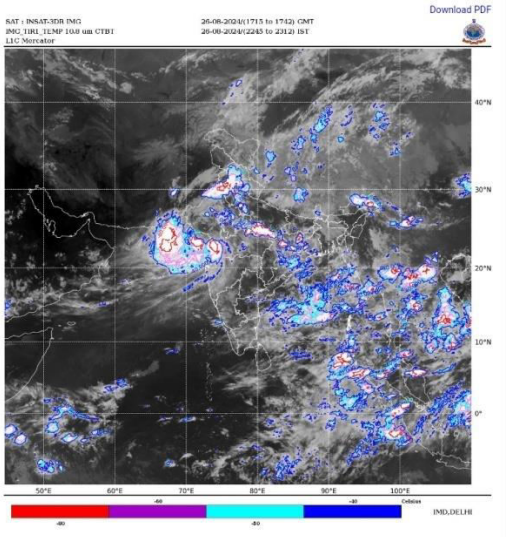

Fig 2 Cloud top brightness temperature of Asia reason for a specific period of time Source: adapted from [8]

Ground based Observatories: Surface weather stations provide vital information. Essential information on precipitation, heat index, humidity, wind direction and speed is provided by surface weather stations. Understanding the local meteorological conditions that can lead to thunderstorm development requires knowledge of this information. The continuous and direct measurements of meteorological variables provided by ground-based stations are crucial for the validation and improvement of nowcasting models (Sun et al., 2014). on precipitation, pressure, temperature, humidity, wind direction, and speed. Understanding the local meteorological conditions that can lead to thunderstorm development requires knowledge of this information. The continuous and direct measurements of meteorological variables provided by ground-based stations are crucial for the validation and improvement of nowcasting models (Sun et al., 2014).

Numerical Weather Prediction (NWP) Models: NWP models use mathematical formulas derived from physical laws to understand the behavior of the weather. By absorbing dataset from numerous resources, such as radar, satellites, and ground stations, these models produce forecasts. While NWP models are generally applied to longer-range predictions, they can be combined with nowcasting systems to improve the prediction of thunderstorm development and movement and to include contextual information (Chen et al., 2012). Combining real-time observational data with NWP models is a potent method for enhancing the precision of thunderstorm nowcasting.

Lightning Detection Networks: Real-time data on lightning strikes, which are strongly related to thunderstorms, is provided by lightning detection networks like the National Lightning Detection Network (NLDN). These networks pick up electromagnetic signals from lightning strikes using sensors located on the ground. Lightning data provide important insights into thunderstorm activity and potential for severe weather, assisting in the identification and tracking of the most active thunderstorm zones (Cummins & Murphy, 2009).

II. Research methodology

A. Data gathering and preprocessing

Dataset consists of historical weather data for over 6000 cities across the India, spanning the period from 2010 to 2023. The model is trained using the data of Indian North ocean region. Dataset may contain missing values. To handle these values data preprocessing is required. Dataset contain columns like ‘date’ which is of object type. Before feeding data to machine learning model, the data type of ‘date’ column is converted to numeric timestamps. After that non numeric columns are dropped. Then data scaling is done which is important because

dataset have features like temperature, pressure, precipitation dew point etc. These features have values which have different units.

B. Anomaly Detection

Features are chosen using statistical analysis and subject expertise to assess their significance to thunderstorm development. Finding high-variance traits or those that have a strong correlation with the incidence of thunderstorms is one way to do this. This research uses distribution graphs of every feature to analyze the data and pick the features.

The normalized feature dataset is divided into two clusters using K-Means clustering. To differentiate between normal and abnormal weather events, the features of each cluster are examined. Clusters that exhibit unusual or infrequent patterns are classified as anomalies. Each data point's separation from the cluster centroid is measured. Data points classified as anomalies are those that is noticeably far from the centroid. Based on their distance from the centroid and the properties of the clusters to which they are assigned, data points are given binary labels ("Normal" or "Anomalous").

C. Random Forest Training And Testing

A standard ratio of 70% for training and 30% for testing is used to divide the dataset into training and testing sets. With this split, a significant amount of the data is reserved for model evaluation and training, ensuring that the model receives adequate training. The preprocessed dataset's characteristics and the anomaly labels derived from K- Means clustering are employed by the Random Forest model. The process of feature selection guarantees the inclusion of pertinent predictors for thunderstorm forecasting. Selected features and anomaly labels are used in the training dataset to train the Random Forest classifier. Based on these inputs, the algorithm learns to categorize weather patterns. In testing phase, model is tested for its accuracy.

III. Results and discussion

For thunderstorm nowcasting, this research utilizes a two-step method: first, K-means clustering is used for anomaly detection, and Random Forest is used for classification. By successfully differentiating between normal and unusual instances, this methodology aims to increase the correctness and reliability of severe thunderstorm predictions. The distribution, skewness, central tendency, and outliers for each attribute of the dataset are displayed in Figure 3. As demonstrated by the "temperature_2m" histogram, most temperature observations, for instance, fall between 10 and 20 degrees Celsius. The "cloud_cover" histogram shows a peak at 0.2, indicating that 20% is the most typical cloud cover value.

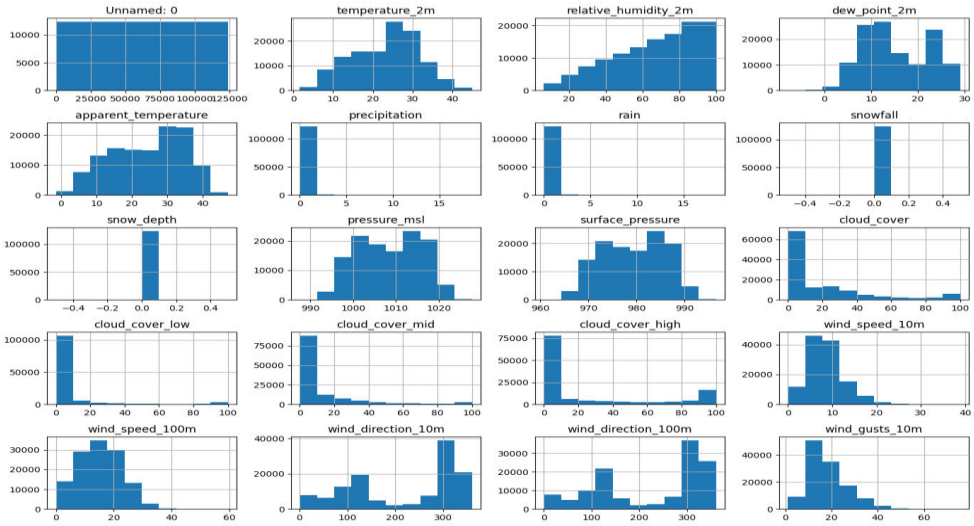

Fig 3 Distribution of Features of dataset

First, Thunderstorm dataset is analyzed for anomalies using K-means clustering. To guarantee consistent scaling across characteristics, Standard Scaling method was initially used to normalize the dataset. The data were divided into discrete groups by the clustering method, and the separations between each data point and its cluster center were calculated. The 95th percentile of these distances was utilized to establish a threshold for anomaly identification. Points falling between this range, were categorized as normal, and those rising over it were called abnormal. With the use of this technique, the study was able to spot variations from regular thunderstorm patterns that might point to the presence of severe weather. The temperature in degrees is represented on the x-axis in Figure 4, while the relative humidity is shown as a percentage on the y-axis. The position of each dot on the graph, which represents a single observation, is based on the humidity and temperature readings at that particular moment. Anomalies are shown by the dots, which have varying colors. Normal data points are represented by black dots, and aberrant data points are represented by pink dots. The temperature_2m and relative humidity_2m are negatively correlated, as can be seen in the graph. Relative humidity tends to drop as temperature rises. Weather data often follows this trend. Anomalies appear to be concentrated near the data distribution's boundaries, particularly at lower temperatures and higher relative humidity. This may suggest that unusual circumstances are more likely to be reported as anomalies. The link between humidity and cloud cover, which can be important prerequisites for anomalies like thunderstorms in our scenario, is depicted in Figure 5.The amount of cloud cover and relative humidity appear to be positively correlated. The relative humidity tends to rise in tandem with an increase in cloud cover. The upper-right quadrant of the image contains the majority of the data points designated as anomalies (shown in a lighter shade), suggesting that anomalies are more likely to occur in conditions with high relative humidity and cloud cover.

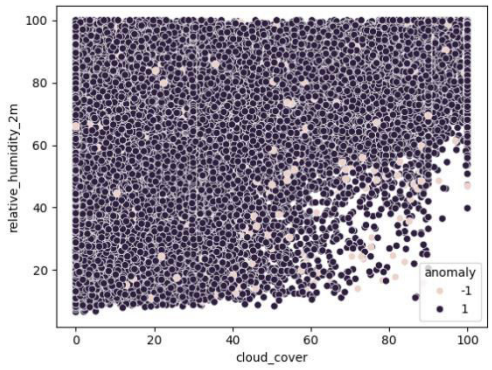

Fig 4 Scatter plot between temperature and humidity shows the normal and anomalous data

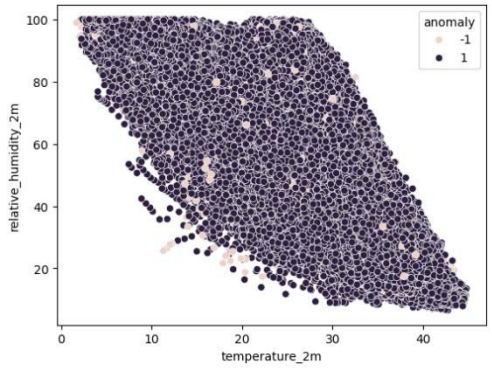

Fig 5 Scatter plot between cloud coverage and humidity shows the normal and anomalous data

Once the abnormalities were identified, Random Forest classifier is used to improve thunderstorm detection even further. The dataset was used to train the model, which contained labels from the K-means clustering. Using these

labels, the model was able to identify patterns related to both normal and unusual circumstances. Several metrics were used to evaluate the Random Forest classifier's performance. This produced the results shown in Table I. The precision of 0.94 indicates that 94% of the data points classified as anomalous by the model were indeed anomalous.

The model's recall of 0.74 indicates that it was able to correctly identify 74% of the real anomalous occurrences, proving that it could detect the majority of significant deviations. With precision and recall combined into a single metric, the F1-score of 0.83 indicates a balanced performance and demonstrates how well the model balances finding real abnormalities with reducing false positives. This model's overall accuracy is 99.67%. Figure 6 shows ROC AUC score of 0.87 provides more evidence of the model's discriminating power. The efficiency of the Random Forest classifier in identifying severe weather events is demonstrated by its high AUC value, which validates its ability to differentiate between normal and aberrant situations. A strong framework for thunderstorm nowcasting was created by combining Random Forest for classification with K-means clustering for anomaly detection. By segmenting the data and establishing a distance-based threshold, K-means clustering successfully found anomalies and allowed us to classify individual data points as normal or anomalous. Training the Random Forest model, which used these labels to learn and improve its classification skills, required this first anomaly detection. The Random Forest classifier's high performance measures (ROC AUC, accuracy, recall, and F1-score) highlight how successful this strategy is. The model is well-suited for real-time thunderstorm prediction due to its high degree of discriminative power and accuracy in identifying anomalies. While reducing the possibility of false alarms, the model's precision and recall are balanced to ensure that severe weather conditions may be detected with reliability.

Overall, our two-step approach shows great promise for enhancing thunderstorm nowcasting, utilizing Random Forest for fine-tuned categorization and K-means clustering for preliminary anomaly identification. This approach improves forecasting accuracy and offers vital early warnings for severe weather occurrences by effectively differentiating between usual and abnormal conditions. This helps impacted areas prepare better and implement mitigation actions.

TABLE I. Performanc matrics

	Precision	Recall	F1-Score
Abnormal values	0.94	0.74	0.83
Normal values	1.00	1.00	1.00
Macro-avg	0.97	0.87	0.91
Weighted-avg	1.00	1.00	1.00

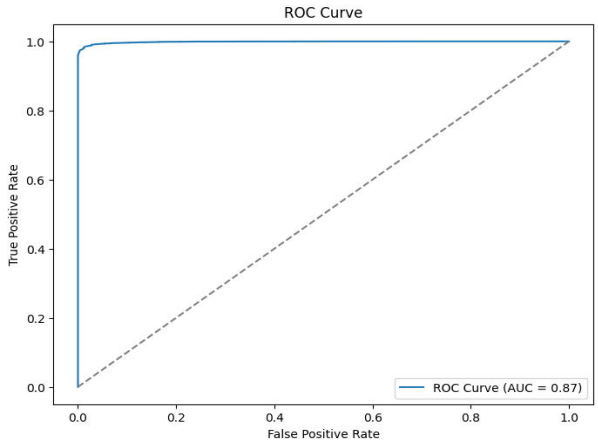

Fig 6 ROC AUC Curve

IV. CONCLUSION

In the context of thunderstorm nowcasting, this study examined the effectiveness of employing Random Forest for model training and testing together with K-Means clustering for row-wise labeling of anomalies. Nowcasting thunderstorms involve predicting lightning-related predictions in a short period, requiring the use of precise and timely models. This used a two- step method to tackle this problem: first, meteorological data was preprocessed using K- Means clustering, and then classified. Then a model predicted thunderstorm events using Random Forest. K-Means assisted in the division of intricate weather patterns into more manageable and understandable categories by dividing the data into clusters. This was able to simplify the data structure for further analysis by grouping historical weather data based on commonalities using an unsupervised learning technique. In essence, every cluster stood for a certain weather pattern or circumstance related to the formation of thunderstorms. Because it supplied labeled data for the Random Forest algorithm, this preprocessing step was essential. When paired with Random Forest, K-Means clustering demonstrated remarkable efficiency in enhancing the accuracy of thunderstorm predictions when compared to conventional forecasting techniques. By capturing intricate, non-linear correlations in the data, the model was able to produce thunderstorm nowcasting that was more accurate and timely.

The encouraging outcomes of this paper provide several directions for further investigation and advancement. The field of feature engineering is one such improvement area. Subsequent research endeavors ought to concentrate on broadening the range of characteristics employed in the model, integrating supplementary climatic factors including humidity, wind shear, and atmospheric pressure. Further improvements in model performance and a more thorough grasp of thunderstorm dynamics may be possible with enhanced feature engineering. The incorporation of further cutting-edge machine learning methods is a topic worth investigating. Even though Random Forest worked well, further advancements can be possible by experimenting with different algorithms like Gradient Boosting Machines (GBM) or Deep Learning models. In certain situations, these methods may perform better in terms of prediction or be able to recognize more subtle patterns in the data. Processing data in real time is a crucial development for operational meteorology. Stormfront nowcasting models may be far more useful in practice if real-time data ingestion and processing tools were developed. By ensuring that forecasts are as up-to-date and pertinent as feasible, real-time processing would enhance the capacity to give warnings promptly and lessen the effects of severe weather events. To further comprehend the temporal and spatial dynamics of thunderstorms, future studies should look into spatiotemporal analysis. Predictions could get more precise by including models that take into consideration both temporal and spatial changes. This would provide an in-depth understanding of how thunderstorms move and change across time and place. Finally, for wider applicability, scalability and adaptability to various geographic locations and climatic circumstances are essential. The positive findings of this thesis offer several avenues for additional research and development. One such area for enhancement is feature engineering. Future studies should focus on expanding the set of parameters used in the model by including additional meteorological variables. With improved feature engineering, there may be room for even greater gains in model performance and a deeper understanding of thunderstorm dynamics. Random Forest performed admirably; more improvements might be made by experimenting with other algorithms, such as Deep Learning models or Gradient Boosting Machines (GBM). The approach provides a solid foundation for further advancements in meteorological prediction, with numerous opportunities for refinement and expansion.

REFERENCES

1 Amjad, K., Malik, M. H., Ghous, H., Hussain, A., & Ismail, M. (2021). Thunderstorm Prediction Using Satellite Images. *International Journal of Remote Sensing*, 42(10), 3557-3578. https://doi.org/10.1080/01431161.2021.1893389

2 Chen, F., Dudhia, J., Chen, M., & Qian, S. (2012). Assessing the Impact of Geophysical Variables on Convection: A Review of NWP and Nowcasting Capabilities. *Journal of Geophysical Research: Atmospheres*, 117(D19), 1-12. https://doi.org/10.1029/2012JD018137

3 Cummins, K. L., & Murphy, M. J. (2009). An Overview of Lightning Locating Systems: History, Techniques, and Data Uses, With an In-Depth Look at the U.S. NLDN. *IEEE Transactions on Electromagnetic Compatibility*, 51(3), 499-518. https://doi.org/10.1109/TEMC.2009.2023450

4 Dash, S., & Sahoo, S. (2021). Lightning Strikes in Bihar during the Monsoon Season: Vulnerability and Mitigation. *Journal of Atmospheric and Oceanic Science*, 12(3), 45-58. https://doi.org/10.1007/s10874-020-01000-z

5 Doswell III, C. A. (2001). Severe convective storms—An overview. *Severe Convective Storms*, 1-26. https://doi.org/10.1007/978-1-4615-1225-3_1

6 Indian Meteorological Department (IMD). (2019). Annual Climate Summary. Ministry of Earth Sciences, Government of India.

7 Indian Meteorological Department (IMD). (2023). Radar Image. *Mausam*. Available: https://mausam.imd.gov.in/imd_latest/contents/index_radar.php. [Accessed: 26-Aug-2023].

8 Indian Meteorological Department (IMD). (2023). Satellite Image. *Mausam*. Available: https://mausam.imd.gov.in/imd_latest/contents/satellite.php#. [Accessed: 26-Aug-2023].

9 Lakshmanan, V., Smith, T., Stumpf, G., & Hondl, K. (2007). The Warning Decision Support System – Integrated Information. *Weather and Forecasting*, 22(4), 596-612. https://doi.org/10.1175/WAF1006.1

10 Mohapatra, A., Nayak, R., Bhattacharya, S. R., & Tripathy, S. S. (2020). Impact of Cyclone Fani and Associated Thunderstorms on Electrical Infrastructure in Odisha. *Natural Hazards*, 103(2), 527-546. https://doi.org/10.1007/s11069-020-03922-2

11 National Disaster Response Force (NDRF). (2018). India's Dust Storms of 2018: A Retrospective Analysis. NDRF Annual Report.

12 Ruzanski, M., Chandrasekar, V., & Wang, Y. (2011). The CASA Nowcasting System. *Journal of Atmospheric and Oceanic Technology*, 28(5), 640-655. https://doi.org/10.1175/JTECH-D-10-05012.1

13 Schultz, D. M., Correia, J., & Stephenson, D. B. (2015). Toward Improving Convective Weather Forecasting: Verification, Predictability, and an Integrated Perspective. *Weather and Forecasting*, 30(6), 1595-1617. https://doi.org/10.1175/WAF-D-14-00107.1

14 Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Koch, S. E., Kong, F., LeMone, M. E., Smull, B. F., Xu, Q., & Grell, E. G. (2014). Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. *Bulletin of the American Meteorological Society*, 95(3), 409-426. https://doi.org/10.1175/BAMS-D-11-00263.1

15 Wilson, J. W., Crook, N. A., Mueller, C. K., Sun, J., & Dixon, M. (1998). Nowcasting Thunderstorms: A Status Report. *Bulletin of the American Meteorological Society*, 79(10), 2079-2099. https://doi.org/10.1175/1520-0477(1998)079<2079>2.0.CO