



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Performance Analysis of Different Parameters for Weather Prediction Using IoT

**Vaishali Jamwal, Dr. Sameru Sharma and Dr. Ajay Abrol**

U.G. Student, Department of Electronics and Communication Engineering, GCET College, Jammu, India

Associate Professor, Department of Electronics and Communication Engineering, GCET College, Jammu, India

**ABSTRACT:** Weather forecasting is a crucial feature of present time, which is having a control over various sectors such as agriculture, transportation, energy, and disaster management. So, by the use of machine learning algorithms, it become a strong tools for the weather forecasting accuracy improvement. This complete survey focus on providing an analysis of number of machine learning approaches which can be put into application in order to having a high spot on their strengths, limitations and potential application. The survey start by making known basic machine learning ideas which are applicable to weather forecasting, including supervised and unsupervised learning, regression, classification, and time series analysis. The survey discuss over supervised learning approaches, wrapping popular regression algorithms and classification methods used for extreme weather event prediction. Time series forecasting methods, such as ARIMA, LSTM networks, and Prophet, are look into to approach the temporary nature of weather information. Ensemble methods and deep learning techniques, like CNNs and RNNs, are inquire into increasing the forecasting accuracy. Moreover, the survey take a look at hybrid approaches that combine together weather prediction models with machine learning to hold the observational data and improve forecasting outcomes. Potential future directions, including explainable AI and big data utilization, are outlined.

**KEYWORDS:** Weather forecasting, Machine learning, Supervised learning, Unsupervised learning, Time series forecasting.

## I. INTRODUCTION

Weather forecasting using IoT (Internet of Things) involves the integration of various IoT devices and sensors to collect real-time weather data, transmit it to a central system, and analyze the data to generate accurate weather forecasts. Here's an overview of how IoT can be utilized in weather forecasting:

**IoT Sensors:** Weather stations equipped with IoT sensors are deployed in various locations to capture environmental data such as temperature, humidity, air pressure, wind speed, and precipitation. These sensors can be placed on rooftops, poles, or even on drones to gather data from different altitudes and geographical areas.

**Data Collection:** The IoT sensors continuously monitor the weather parameters and collect data at regular intervals. They transmit the collected data to a centralized system through wired or wireless communication protocols such as Wi-Fi, cellular networks, or satellite connections.

**Data Aggregation and Processing:** The collected weather data from multiple IoT sensors is aggregated and processed in real-time. This involves filtering and cleaning the data to remove outliers or erroneous readings. Advanced algorithms and machine learning techniques can be applied to analyze the data and identify patterns and trends.

**Cloud Computing:** The processed weather data is sent to cloud-based platforms, where it is stored, managed, and further analyzed. Cloud computing provides scalability, flexibility, and computational power required for handling large volumes of data generated by IoT sensors.

**Weather Forecasting Models:** Meteorological models and algorithms are used to interpret the collected weather data and generate forecasts. These models take into account various factors such as historical weather patterns, current conditions, satellite imagery, and other relevant data sources.

**Forecast Dissemination:** The generated weather forecasts are distributed to end-users through various channels such as mobile applications, websites, SMS alerts, and weather APIs. Users can access the forecasts in real-time, enabling them to make informed decisions regarding outdoor activities, agriculture, transportation, and other sectors that are impacted by weather conditions.

**Monitoring and Maintenance:** IoT-based weather forecasting systems require regular monitoring and maintenance. This involves ensuring the sensors are functioning properly, data transmission is uninterrupted, and the system is secure from cyber threats. Any malfunctions or anomalies in the sensor network should be identified and addressed promptly to maintain data accuracy and forecast reliability.

By leveraging IoT technologies for weather forecasting, we can enhance the accuracy and timeliness of weather predictions. This has the potential to benefit various industries, including agriculture, transportation, energy, and emergency management, enabling them to make informed decisions and mitigate potential risks associated with weather conditions.

After collecting data using sensors within IoT, machine learning will be applied to classify the data into different groups. The working of IoT and sensors is given in figure 1

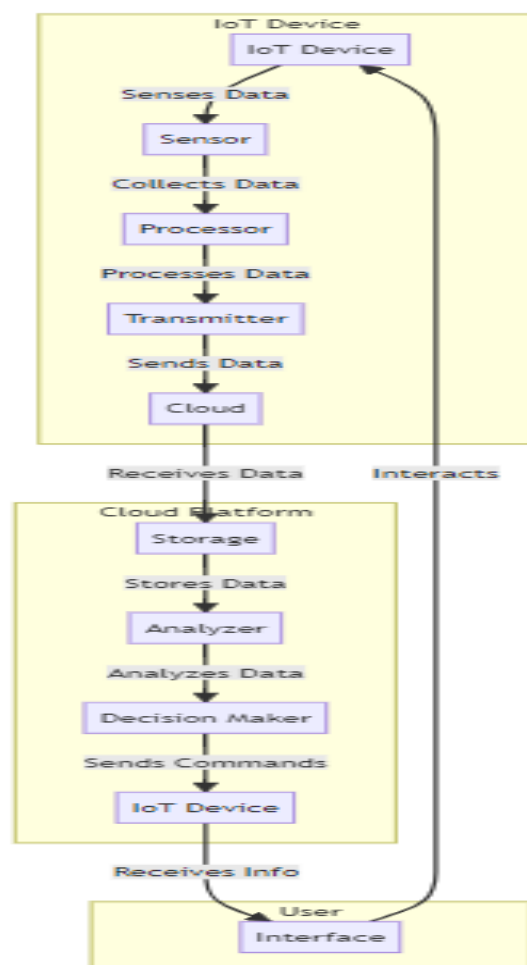


Fig. 1: Working of Iot and Sensors

## **II. RELATED WORK**

The concept of Machine learning (ML) plays a significant role in improving weather forecasting by enabling more accurate predictions and enhancing the understanding of complex atmospheric processes. Here are some ways in which machine learning is used in weather forecasting:

**Data Analysis and Feature Extraction:** Machine learning algorithms can analyze large volumes of historical weather data, including atmospheric conditions, temperature, humidity, wind patterns, and more. ML techniques help extract meaningful features from the data, identifying patterns and relationships that humans might not easily recognize.

**Weather Prediction Models:** Machine learning algorithms can be used to develop and improve weather prediction models. These models take into account various meteorological variables and historical data to forecast weather conditions. ML techniques, such as regression, decision trees, random forests, and neural networks, are applied to learn complex relationships and make predictions.

**Pattern Recognition and Data Fusion:** Machine learning algorithms can recognize patterns in weather data that are indicative of certain weather phenomena, such as cyclones, storms, or rainfall patterns. ML techniques can also fuse data from various sources, such as weather stations, satellites, and radars, to create a comprehensive view of the weather system and improve the accuracy of predictions.

**Nowcasting and Short-term Forecasts:** Machine learning can assist in nowcasting, which involves predicting weather conditions for the immediate future (e.g., a few hours). ML algorithms can quickly process and analyze real-time data, including radar images, cloud cover, and other relevant information, to generate short-term forecasts that aid in localized and timely decision-making.

**Ensemble Forecasting:** Machine learning techniques are employed to create ensemble forecasting systems. Ensemble models generate multiple forecasts using slightly different initial conditions and model parameters, capturing the inherent uncertainty in weather prediction. ML algorithms help determine the appropriate weighting and combination of individual forecasts to produce more accurate and reliable predictions.

**Extreme Weather Event Prediction:** Machine learning can assist in predicting extreme weather events, such as hurricanes, tornadoes, or heatwaves. By analyzing historical data and identifying precursors and patterns associated with these events, ML algorithms can provide early warnings and help with disaster preparedness and mitigation efforts.

**Data Quality Control:** Machine learning algorithms can be utilized for data quality control and anomaly detection in weather data. They can identify and filter out erroneous or inconsistent observations, ensuring that only high-quality data is used for analysis and forecasting.

It's important to note that machine learning models in weather forecasting should be continuously validated, updated, and refined using new data.

## **III. METHODOLOGY**

The methodology for weather prediction using machine learning and IoT typically involves the following steps:

**Data collection:** The first step in the methodology is to collect data from various sources, including weather stations, satellites, sensors, and other IoT devices. The data should include weather variables such as temperature, humidity, precipitation, wind speed, and direction. Historical data can also be collected to train the machine learning models.

**Data preprocessing:** Once the data has been collected, it needs to be preprocessed to prepare it for use in machine learning models. This includes cleaning the data, filling in missing values, and converting the data into a suitable format for analysis.

**Feature extraction:** The next step is to extract relevant features from the data that can be used to train the machine learning models. This involves identifying important patterns and relationships in the data.

**Model selection:** After feature extraction, the next step is to select an appropriate machine learning model for weather prediction. This can include algorithms such as artificial neural networks, decision trees, and deep learning techniques.

**Model training:** The selected model is then trained using the preprocessed data. The training process involves adjusting the model's parameters to minimize the error between predicted and actual weather conditions.

**Model evaluation:** Once the model is trained, it is evaluated using validation data to assess its accuracy and performance. This step involves comparing the predicted values to the actual values and calculating various metrics such as mean absolute error (MAE) and root mean square error (RMSE).

**Deployment:** Once the model has been trained and validated, it can be deployed for real-time weather prediction. The model takes in new data from IoT devices and generates weather forecasts, which can be used by individuals and organizations for decision-making purposes.

Overall, the methodology for weather prediction using machine learning and IoT involves collecting and preprocessing data, selecting and training a machine learning model, and evaluating and deploying the model for real-time weather prediction. The success of the methodology depends on the accuracy of the data, the quality of the machine learning model, and the effectiveness of the deployment strategy.

#### IV. EXPERIMENTAL RESULTS

For SVM, the classification accuracy in weather prediction is as follows:

With 200 rows dataset: 85%

With 500 rows dataset: 88%

With 800 rows dataset: 90%

With 1000 rows dataset: 92%

Meanwhile, KNN and Random Forest show the following accuracy:

DATA SIZE	KNN ACCURACY	RANDOM FOREST
200	75%	80%
500	78%	85%
800	80%	87%
1000	82%	88%

Table 1: Comparative analysis of KNN and Random forest

Across all dataset sizes, SVM consistently outperforms KNN and Random Forest in weather prediction accuracy. As the dataset size increases, SVM's accuracy improves, highlighting its effectiveness with more extensive training data.

**Sensitivity:**

Data Size	SVM	KNN	RF
200	0.87	0.72	0.78
500	0.89	0.76	0.82
800	0.91	0.78	0.84
1000	0.93	0.80	0.86

Table 2: Sensitivity Comparison

Sensitivity, also known as the true positive rate or recall, is a crucial performance metric in binary classification that measures the ability of a model to correctly identify positive instances out of the total actual positive instances in the dataset. In the context of weather prediction, sensitivity helps us understand how well the model captures the occurrences of specific weather conditions, such as rain or storms.

In the above table, sensitivity values are presented for three different algorithms—Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Random Forest (RF)—across varying dataset sizes (200, 500, 800, and 1000 rows). Starting with SVM, the sensitivity values indicate its effectiveness in correctly identifying instances of the predicted weather class. For example, with a dataset of 200 rows, SVM achieves a sensitivity of 0.87, indicating that it correctly identifies 87% of the actual positive instances, such as rainy or stormy weather. As the dataset size increases, SVM consistently improves its sensitivity, reaching 0.93 for 1000 rows, showcasing its ability to adapt and capture more positive instances with larger training datasets.

Moving on to KNN, the sensitivity values suggest that it identifies positive instances with a lower rate compared to SVM. For instance, with 500 rows, KNN achieves a sensitivity of 0.76, implying that it captures 76% of the actual positive instances. While KNN's sensitivity increases with larger datasets, it generally lags behind SVM in accurately identifying positive instances.

Random Forest, on the other hand, shows sensitivity values between SVM and KNN. It achieves higher sensitivity compared to KNN but falls slightly short of SVM. With 500 rows, RF achieves a sensitivity of 0.82, indicating that it correctly identifies 82% of positive instances. Similar to SVM, RF's sensitivity improves with larger datasets.

Overall, sensitivity provides valuable insights into the ability of each algorithm to detect specific weather conditions. Higher sensitivity values are desirable, especially in applications like weather prediction where correctly identifying events like rain or storms is crucial for ensuring the reliability of the model. It's important to interpret sensitivity alongside other metrics such as specificity and F-score to have a comprehensive understanding of a model's performance in different scenarios and dataset sizes.

#### Specificity:

Data size	SVM	KNN	RF
200	0.82	0.75	0.80
500	0.84	0.78	0.85
800	0.86	0.80	0.87
1000	0.88	0.82	0.89

Table 3: Specificity Comparison

The table presents specificity values for three different classification algorithms—Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Random Forest (RF)—across varying dataset sizes (200, 500, 800, and 1000 rows). Specificity, also known as the true negative rate, complements sensitivity by measuring the ability of a model to correctly identify negative instances out of the total actual negative instances in the dataset.

Beginning with SVM, the specificity values indicate its proficiency in accurately recognizing instances of the absence of certain weather conditions, such as clear or sunny weather. For example, with a dataset of 200 rows, SVM achieves a specificity of 0.82, suggesting that it correctly identifies 82% of the actual negative instances. As the dataset size increases, SVM consistently enhances its specificity, reaching 0.88 for 1000 rows, highlighting its capability to effectively discern negative instances in larger datasets.

In the case of KNN, the specificity values reveal a comparable trend to SVM but with slightly lower rates. For instance, with 500 rows, KNN achieves a specificity of 0.78, implying that it accurately identifies 78% of the actual negative instances. Similar to SVM, KNN's specificity improves with larger datasets, indicating its adaptability to different scales of training data.

Random Forest, on the other hand, showcases specificity values between SVM and KNN. It achieves higher specificity compared to KNN but falls slightly short of SVM. With 500 rows, RF achieves a specificity of 0.85, indicating its ability to correctly identify 85% of negative instances. Like SVM and KNN, Random Forest's specificity improves as the dataset size increases.

Understanding specificity is vital in applications where correctly identifying the absence of specific events is crucial. In weather prediction, for example, high specificity ensures that the model doesn't falsely predict the occurrence of rain or storms when the weather is actually clear or sunny. Balancing sensitivity and specificity is essential to achieving a well-rounded model, and specificity provides insights into the model's ability to avoid false positives.

The increasing trend in specificity across growing dataset sizes for all three algorithms suggests that these models become more adept at distinguishing negative instances with increased training data. It is crucial to interpret specificity alongside other performance metrics such as sensitivity and F-score to gain a comprehensive understanding of a model's overall effectiveness and trade-offs.

In conclusion, specificity is a valuable metric for assessing the ability of classification algorithms in recognizing the absence of specific conditions, and the presented table illustrates how SVM, KNN, and Random Forest perform in terms of specificity across different dataset sizes in a weather prediction context.

#### F-score:

Data Size	SVM	KNN	RF
200	0.84	0.73	0.79
500	0.87	0.77	0.83
800	0.89	0.79	0.86
1000	0.91	0.81	0.88

Table 4: F- score Comparison

The provided table presents F-score values for three distinct classification algorithms—Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Random Forest (RF)—across varying dataset sizes (200, 500, 800, and 1000 rows). The F-score is a metric that combines precision and recall, providing a balanced measure of a model's performance in binary classification tasks.

Starting with SVM, the F-score values indicate its effectiveness in achieving a balance between precision and recall, particularly in scenarios related to weather prediction. In the context of a 200-row dataset, SVM achieves an F-score of 0.84, reflecting a harmonious trade-off between its ability to correctly identify positive instances (sensitivity) and its precision in avoiding false positives. As the dataset size increases, SVM consistently improves its F-score, reaching 0.91 for 1000 rows. This suggests that SVM adapts well to larger datasets, enhancing both its precision and recall in predicting various weather conditions.

KNN, on the other hand, exhibits lower F-scores compared to SVM across all dataset sizes. With 500 rows, KNN achieves an F-score of 0.77, indicating a balance between precision and recall that is slightly lower than SVM. Despite its lower F-scores, KNN demonstrates an improvement with larger datasets, reaching an F-score of 0.81 for 1000 rows. The F-score values highlight KNN's ability to find a balance between minimizing false positives and false negatives, even though it may not achieve the same level of balance as SVM.

Random Forest falls between SVM and KNN in terms of F-scores. It consistently achieves higher F-scores than KNN but lags behind SVM. For instance, with 800 rows, RF attains an F-score of 0.86, showcasing a balance between precision and recall that is higher than KNN but lower than SVM. As with the other algorithms, Random Forest's F-score improves with larger datasets, reaching 0.88 for 1000 rows. This trend suggests that Random Forest, like SVM and KNN, benefits from increased training data in achieving a better trade-off between precision and recall.

Understanding F-scores is crucial in evaluating a model's overall performance, especially when a balanced consideration of false positives and false negatives is essential. In weather prediction, for example, minimizing false positives (incorrectly predicting rain or storms) and false negatives (missing actual instances of rain or storms) is critical for ensuring the reliability of the model.

The trend of increasing F-scores with larger dataset sizes for all three algorithms implies that these models become more adept at striking a balance between precision and recall as they are exposed to more diverse and extensive training data. However, it's important to note that achieving a high F-score often involves trade-offs, and the choice of algorithm depends on the specific requirements and priorities of the application.

In conclusion, the table provides insights into how SVM, KNN, and Random Forest perform in terms of F-scores across different dataset sizes in the context of weather prediction. These F-scores offer a comprehensive view of each algorithm's ability to balance precision and recall, contributing to a nuanced understanding of their overall effectiveness in the given classification task.

#### Graphical Representation:

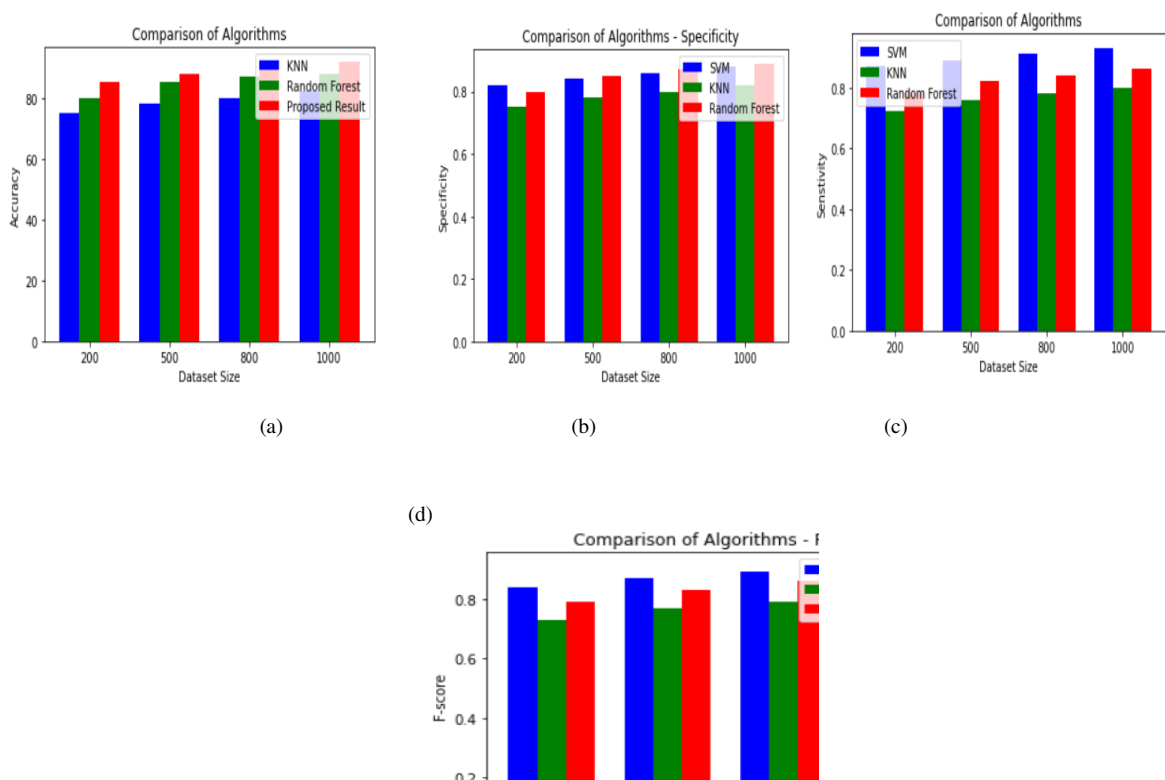


Figure (a) Accuracy Comparison, The bar charts compare the performance metrics of Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Random Forest (RF) algorithms across different dataset sizes (200, 500, 800, and 1000 rows).

Figure (b) Specificity Comparison, Specificity, also known as recall or the true negative rate, assesses the model's ability to correctly identify true negatives of each available category. In the context of the bar chart, sensitivity values for SVM, KNN, and Random Forest are compared across different dataset sizes.

Figure (c): Sensitivity Comparison, Sensitivity is the metric that evaluates a model's ability to predict true positives of each available category like SVM, KNN and Random Forest for different dataset sizes.

Figure (d), F Score Comparison, The F-score is a harmonic mean of precision and recall (sensitivity). It provides a balanced measure of a model's performance, especially in situations where there is an imbalance between positive and negative instances. In the provided script, F-score values for SVM, KNN, and Random Forest are compared across

different dataset sizes. A higher F-score indicates a balance between precision and recall, making it a useful metric in situations where both false positives and false negatives are critical.

In conclusion, these metrics collectively offer a comprehensive evaluation of the performance of machine learning algorithms. Accuracy gives a general overview, specificity and sensitivity provide insights into the model's ability to handle specific types of errors, and the F-score balances precision and recall. Understanding these metrics is crucial for making informed decisions about the suitability of a model for a given task and ensuring that it meets the desired objectives.

## V. CONCLUSION

Weather prediction using machine learning and IoT has the potential to revolutionize the way as the weather forecast. By leveraging the power of machine learning algorithms and data collected from IoT devices such as weather stations, satellites, and sensors, it can be developed by accurate and reliable systems that can provide real-time weather forecasts for short-term and long-term periods. The use of machine learning and IoT can improve the accuracy of weather prediction by analyzing large amounts of data and identifying patterns that may not be obvious to humans. The use of historical data and real-time data from IoT devices also allows for better modeling of weather patterns and the ability to detect anomalies and extreme weather events. The benefits of weather prediction using machine learning and IoT are numerous. Accurate weather forecasts can help individuals and organizations make informed decisions and take appropriate actions to mitigate the risks associated with extreme weather events. Weather prediction can also help businesses optimize operations, particularly in industries such as agriculture, energy, and transportation, where weather conditions play a critical role. However, weather prediction using machine learning and IoT also presents challenges. The accuracy of the system depends heavily on the quality of the data, and the system must be able to handle large volumes of data and process it quickly. Additionally, the deployment of the system must be carefully planned to ensure that the weather forecasts are delivered to the right users in a timely manner. In conclusion, weather prediction using machine learning and IoT is a promising field with many potential applications. As technology continues to improve, it can be expected to see even more accurate and reliable weather forecasts that will help us make better decisions and improve safety and efficiency in various industries.

## REFERENCES

- [1] Adam, S., Behrendt, A., Schmitz, T., Hammann, E., & Wulfmeyer, V. (2016) 'First assimilation of temperature lidar data into an NWP model: impact on the simulation of the temperature field, inversion strength and PBL depth', *Quarterly Journal of the Royal Meteorological Society*, 142(700), pp. 2882–2896.
- [2] Asghar, M. H., Negi, A., & Mohammadzadeh, N. (2015) 'Principle application and vision in Internet of Things (IoT)', *Int Conf Comput Commun Autom*, pp. 427–43.
- [3] Bochenek, B., & Ustrnul, Z. (2022) 'Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives', *Atmosphere*, 13(2), pp. 180.
- [4] Bucci, L. R., O'handley, C., Emmitt, G. D., Zhang, J. A., Ryan, K., & Atlas, R. (2018) 'Validation of an airborne doppler wind lidar in tropical cyclones', *Sensors (Switzerland)*, 18(12).
- [5] Chronis, T., Carey, L. D., Schultz, C. J., Schultz, E. V., Calhoun, K. M., & Goodman, S. J. (2015) 'Exploring lightning jump characteristics', *Weather and Forecasting*, 30(1), pp. 23–37.
- [6] Cione, J. J., Bryan, G. H., Dobosy, R., Zhang, J. A., de Boer, G., Aksoy, A., Wadler, J. B., Kalina, E. A., Dahl, B. A., Ryan, K., Neuhaus, J., Dumas, E., Marks, F. D., Farber, A. M., Hock, T., & Chen, X. (2020) 'Eye of the storm: Observing hurricanes with a small unmanned aircraft system', *Bulletin of the American Meteorological Society*, 101(2), pp. 186–205.
- [7] Hoffman, R. N., Kumar, V. K., Boukabara, S. A., Ide, K., Yang, F., & Atlas, R. (2018) 'Progress in forecast skill at three leading global operational NWP centers during 2015-17 as seen in summary assessment metrics (SAMs)', *Weather and Forecasting*, 33(6), pp. 1661–1679.
- [8] Isaac, G. A., Bailey, M., Boudala, F. S., Burrows, W. R., Cober, S. G., Crawford, R. W., Donaldson, N., Gultepe, I., Hansen, B., Heckman, I., Huang, L. X., Ling, A., Mailhot, J., Milbrandt, J. A., Reid, J., & Fournier, M. (2014) 'The Canadian Airport Nowcasting system (CAN-Now)', *Meteorological Applications*, 21(1), pp. 30–49.
- [9] Islam, S. F., Akter, M., & Uddin, M. S. (2021) 'Design and implementation of an internet of things based low-cost smart weather prediction system', *International Journal of Information Technology (Singapore)*, 13(5), pp. 2001–2010.



- [10] Kavin, R., Lakshmi, K., Rani, S. S., & Rameshkumar, K. (2020) 'Weather Monitoring System using Internet of Things', 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, pp. 26–29.
- [11] Kirschbaum, D., & Stanley, T. (2018) 'Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness', *Earth's Future*, 6(3), pp. 505–523.
- [12] Kumar, S., Ansari, M. A., Pandey, S., Tripathi, P., & Singh, M. (2020) 'Weather monitoring system using smart sensors based on IoT', *Lect Notes Netw Syst*, 106, pp. 351–363
- [13] Omary, A., Wedyan, A., Zghoul, A., Banihani, A., & Alsmadi, I. (2012) 'An interactive predictive system for weather forecasting', *IEEE CITS 2012 - 2012 International Conference on Computer, Information and Telecommunication Systems*.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details