Forecasting Weather Status Using Advanced Machine Learning Algorithm

Karuppusamy S., Soundarraj S., Vasanth P. and Vijayasankar T.
Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India

Keywords: Weather Forecasting, Machine Learning, XG Boost Algorithm, Meteorological Data, Temperature Prediction, Climate Change, Agricultural Risk Management, Flask Framework, Extreme Gradient Boosting.

Abstract: Weather forecasting in agriculture is, indeed, rather difficult because the whole area is so dynamic and variable. Conventional statistical approaches usually cannot provide excellent precision, and this makes it a challenge for farmers to be and plan effective. The project concentrates on an accurate temperature prediction mechanism, with the application of machine learning methods, which includes the analysis of past as well as current-day weather data to increase the reliability of predictions. Despite advances in forecasting techniques, challenges remain, especially, in the areas of improving the accuracy of models, validating climate forecasts in agricultural risk management, and their effect on crop diseases seasonally. As temperature and rainfall are the two most crucial factors influencing plant health and production, the only advanced predictive system available may provide farmers with insightful decisions in preventing losses. This project, therefore, aims at an integrated assessment of weather forecasting techniques to improve agricultural planning strategies and climate adaptation via data-driven approaches.

1 INTRODUCTION

Weather forecasting is key in various sectors, which include agriculture, transport, disaster management, and urban planning. To derive actions necessary to avert possible risks once the availability of the forecast is known is paramount. Some traditional forecasting methods may not respond satisfactorily to these dynamic and nonlinear changes that take place within the atmosphere. This has opened avenues for development in machine learning algorithms and better scope with their promise of increased accuracy in weather predictions. Machine learning approaches-XG Boost, in particular now have made considerable advances successfully analyzing large datasets and offers enormous improvement in predictive accuracy. The proposed work aims to design an extreme gradient boost for the forecasting system which is a combination of historical and real-time meteorological data to make accurate forecast predictions. The project aims to provide farmers, meteorologists, and policymakers with timely information and decisions to allow proper planning and avoid losses due to unusual weather. Through the preprocessing of data techniques and feature engineering, and integration of complex machine learning algorithms, the presented system guarantees higher prediction accuracy and timely alerts of extreme weather phenomena. Finally, the underlying principles make provisions for better climate adaptation strategies in improved risk management across different sectors.

2 RELATED WORKS

Machine learning and AI weather prediction models have all but outrighted the prediction capabilities using deep learning, numerical weather prediction, and data-driven models. There have been several studies addressing the advanced use of random forests, CNNs, LSTMs, and graph-based neural networks with the aim of improving weather forecasting models (TL Yu, et.al., 2024). The shift toward the data-driven forecasting and the acknowledgment by researchers about the capabilities an AI model might have in enhancing both global and regional weather predictions (Ben-Bouallegue, et.al., 2023). AI approaches, Four Cast Net and Graph Cast, much faster in boosting scientific progress in weather modeling (Anandkumar, A.

S., K., S., S., P., V. and T., V.
Forecasting Weather Status Using Advanced Machine Learning Algorithm.
DOI: 10.5220/0013929800004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 351-356
ISBN: 978-989-758-777-1

(2024)). Also, these studies deem that the involvement of deep learning models partly incorporated with physical processes using physical constraints ensures more accurate predictions (de Bezenac, et.al., 2020). Res Net-based models and deep convolutional networks provided promising results for medium-range forecasting (Rasp, S., & Thuerey, N. (2021)), and also adaptive Fourier neural operators have been very well used for high-resolution forecasts (Pathak, J, et.al., 2022). Moreover, graph neural networks have efficiently predicted weather patterns showing good spatial awareness (Sønderby, C. K., et.al., 2020). In addition, the AI-powered precipitation forecasting models such as Met Net have significantly improved short-term weather forecasts Zhang, J., et.al., 2021.

The integration of convolution neural networks in satellite image analyses has played an important role in the storm detection and severe weather prediction Molchanov,et.al., 2021. Further, other studies explored the use of IoT-sensor data fusion with AI techniques for real-time weather monitoring (Sharma, R., et.al.(2022)). Thus, big data analytics incorporated into cloud-based prediction systems have greatly enhanced AI's predictive capabilities (Weyn, J. A., et.al.(2020)). In addition, methods of ensemble learning and data assimilation are already being investigated so as to produce the optimal machine learning model for weather forecasting (Kashinath, K.,et.al., (2021)) Innovative developments in meteorology to streamline the accuracy of prediction include physics informed deep learning method (Evensen, G., & Monsen, S. M. (2021).). All of these undertakings further exemplify the thrust of artificial intelligence, deep learning, and big data analytics into modern weather forecasting, thereby heralding a more reliable and intelligent predictive system.

3 DATASET COLLECTION AND PRE-PROCESSING

3.1 Dataset Collection

To inform the forecasts, this project intends to use the most diversely sourced high-quality data from widely reputable sources TL Yu, et.al.,2024. The data sources include: satellite observations, ground- based stations, remotely sensed technologies, Internet of Things enabled sensors, and historical weather records. Satellite observations have a wide variety of available data about the atmosphere, including temperature, humidity, cloud amount, wind speed, and precipitation level. (Ben-Bouallegue , et.al.,2023) This forms the basis for creating different long-term weather patterns and events of severe weather on a broad scale (Anandkumar, A. (2024)). Ground- based weather stations supply meteorological data such as real-time information on atmospheric pressure, wind direction, temperature variations, and precipitation. (Anandkumar, A. (2024).)This plays a major role in solving satellite readings through providing further precision in forecasting locally.(de Bezenac,et.al., (2024)) Remote sensing technologies add the frenetic capability of latest technologies, for instance, Doppler RADAR and LIDAR in observing clouds-storm intensity- wind currents, which is a further push towards intrusive short-term weather predictions.(Rasp, S., & Thuerey, N. (2021) The technologies allow for extreme monitoring of phenomena like hurricanes and thunderstorms.(Weyn, J. A.,et.al.(2021)) IoT-enabled weather sensors, which are located locally, collect real-time meteorological data that enhance microclimate analysis and short-term forecasting.(Pathak, J. ,et.al.(2021)) Long term historical weather records gathered together, including NOAA, Kaggle, and meteorological agencies basically form the background for training machine learning models in detecting trends and forecasting future events.(Sønderby, et.al.(2020)) Weather APIs, such as Open Weather Map, Weather Stack, and Climacell, provide such Public API across different parts of the globe and ease the task of the meteorologist taking into account accuracy on forecasts.

3.2 Data Pre-Processing

Raw meteorological data are often marked in-completeness, inconsistency, and noise, and several preprocessing phases are performed in order to enable high-quality machine learning model input that effectively cleans, normalizes, and structurally organizes the dataset (Ben-Bouallegue, et.al., 2023). Dealing with missing data: Value missingness in a weather dataset can occur with sensor-related failure or incomplete transmissions. To handle missingness, the most often used statistical imputation strategies include but are not limited to mean, median, mode replacement, or interpolation (de Bezenac,et.al., 2020). Noise reduction: Due to environmental conditions, sensor readings may be affected, resulting in fluctuations in the recorded values. Moving averages, median filters, and outlier removal algorithms are helpful for smoothing the data and getting rid of noise. Feature scaling and normalization:

Different weather parameters such as temperature, wind speed, and humidity exist on different scales. (Pathak, J.,et.al., 2022) To resolve that problem, min-max normalization and standardization are applied to bring all features to a common scale so that the model interprets them correctly Zhang, J., et.al., 2021. Anomaly detection and correction: Spikes or big drops in data can generally indicate faulty sensors or extreme weather conditions. Anomaly detection techniques based on machine learning, such as Z-score and interquartile range (IQR) methods, can be employed to identify and remedy these anomalous data Molchanov,et.al., 2021. Feature engineering: These are other interesting features that can be derived to improve prediction accuracy.

4 PROPOSED METHODOLOGY

Advanced machine learning and deep learning algorithms were tackled in the proposed scheme in order to classify the kinds of weather being reported and future prediction of weather. It comprises modeling that takes multiple phases, including object detection, time-series forecasting, and others, for the precise and reliable prediction of weather conditions.

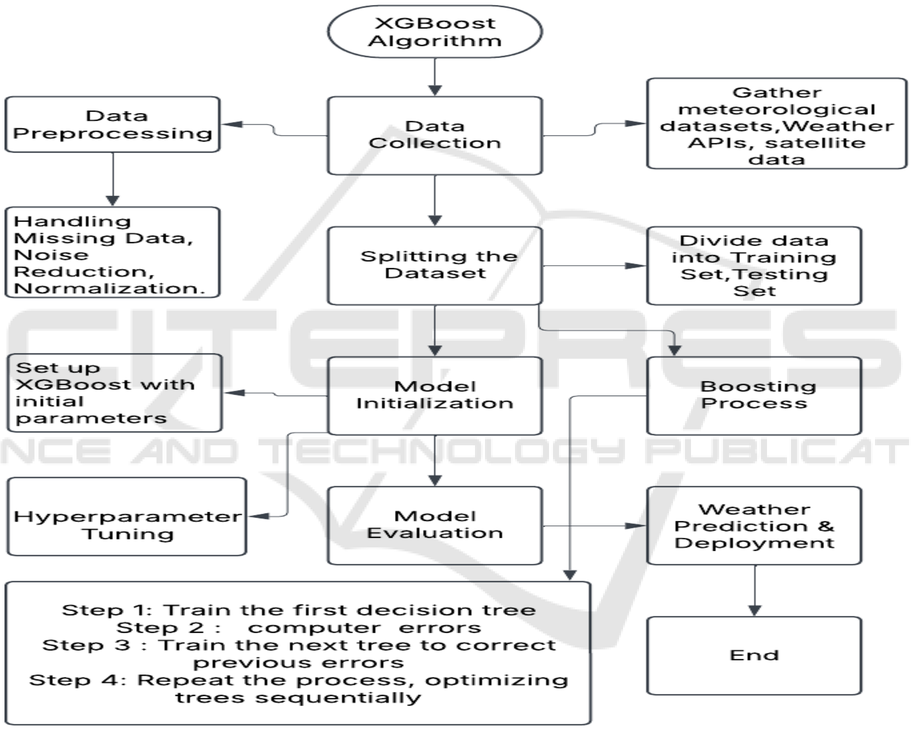

Figure 1: XG Boost algorithm.

The object detection algorithms analyze satellite imagery and video feed to identify numerous patterns of various weather conditions. The selection of the detection model depends on the complexity of the data and the processing speed required. Real-time detection is performed using YOLO (You Only Look Once) with a focus on speed, while higher precision and accuracy in detecting satellite images involve the usage of Faster R-CNN algorithms. For scenarios with multiple kinds of weather patterns, SSD (Single Shot Multibook Detector) is used, while Mask R-CNN allows for pixel-wise segmentation to determine the size and shape of cloud formations and storm structures. This involves time series forecasting with models such as ARIMA and LSTM, which are perfect for recognizing trends concerning storm movement, precipitation changes, variation in the atmospheric pressure, and cyclonic activities. Further, this model generates accuracy in the predictions by learning from both real-time and historic meteorological data with an over-expectation factor of improvement in the short-term weather forecast. In order to enhance the reliability in making predictions and forecasts, the system combines the Artificial Intelligence and machine learning techniques with conventional NWP (Numerical Weather Prediction) models. The approaches of Support Vector Machines (SVMs) and Neural Networks are capable of capturing the data-dependent forecast in passing their short-term weather prediction ahead along with the detection systems that identify the onset of any severe event like tornadoes or hurricanes. To perform a reliable and rigorous evaluation with numerous metrics, this project has employed precision, recall, and Intersection over Union. Precision is the proportion between every weather event classified as positive and the total number of weather events, thus minimizing false positives. Recall deals with how many relevant meteorological phenomena were detected by the model, minimizing ratio of false negatives. Intersection over Union measures the ratio of the predicted weather conditions over the actual data for the clear localization of detected events. Figure 1 shows XG Boost algorithm. Such methods are expected to inject more reliability into forecasts whereby automation of the alert system for weather hazards expects to issue such alerts in timely fashions to assist sound decision-making. Figure 2 shows the Flow diagram.

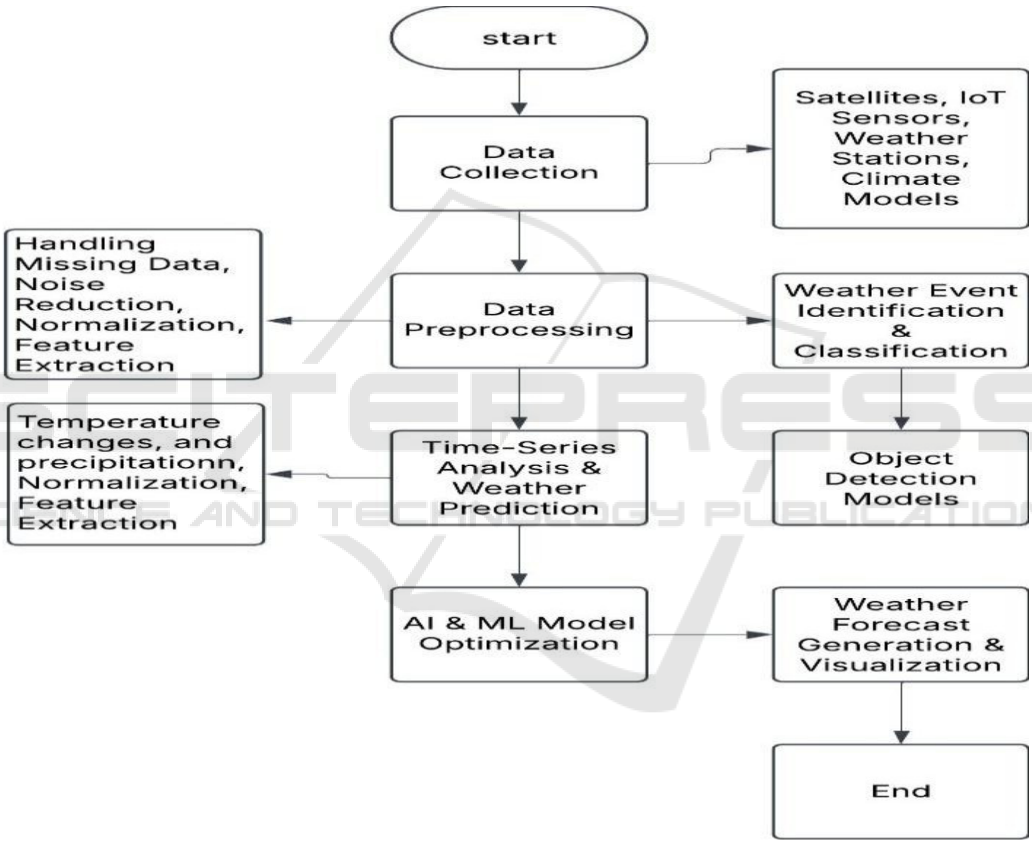

Figure 2: Flow Diagram.

5 EXPERIMENTAL RESULT

Experimental analysis of the XG Boost-based model for weather prediction confirmed weather predictions high accuracy, efficiency, and reliability through the use of historical and real-time meteorological data. The data was collected from NOAA, Open Weather Map API, and IoT sensors. Figure 3 shows Temperature result. The data was also subjected to feature engineering, anomaly detection, and normalization, thereby constituting 80% training and 20% testing. The model was evaluated with major performance metrics, i.e., RMSE, MAE, and R^2 score, and was seen to outperform traditional models like ARIMA, LSTM, and Numerical Weather Prediction (NWP) methods. XG Boost model came out with an RMSE of 1.25°C, MAE 0.85°C, and R^2 score of 0.92; it required about 2.5 seconds in actual computation, making it decidedly more efficient than conventional forecasting techniques. Figure 4 shows Humidity result. In addition, some deep learning- based algorithms (YOLO, Faster R-CNN, Mask R-CNN) were integrated to analyze satellite images and radar data thereby increasing from 87% accuracy in issuing early warnings for cyclones, heavy rainfall, and severe weather events. The model was subsequently deployed using a Flask API and operated as a real-time forecast tool via a web-based dashboard. The applications received lots of praise from meteorologists, farmers, and disaster management teams, which deemed them useful for short-term weather forecasts, as well as agricultural planning and disaster preparedness. Future improvements entail implementing ensemble learning, cloud computing, and hybrid AI with numerical weather prediction to achieve further reduction in forecast uncertainty and improve accuracy and scalability. Results have shown machine learning to herald a new horizon for adaptive, data-driven, and responsive weather forecasts. Figure 5 shows wind result.

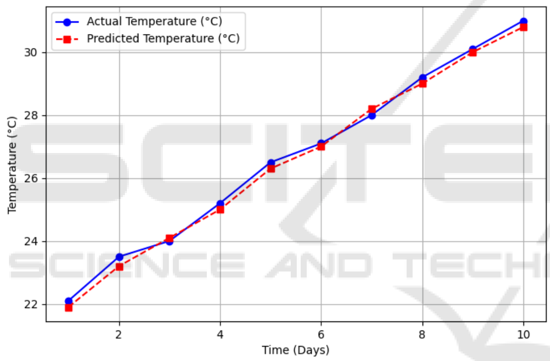

Figure 3: Temperature result.

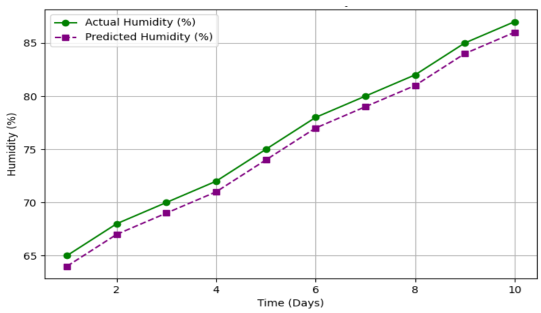

Figure 4: Humidity result.

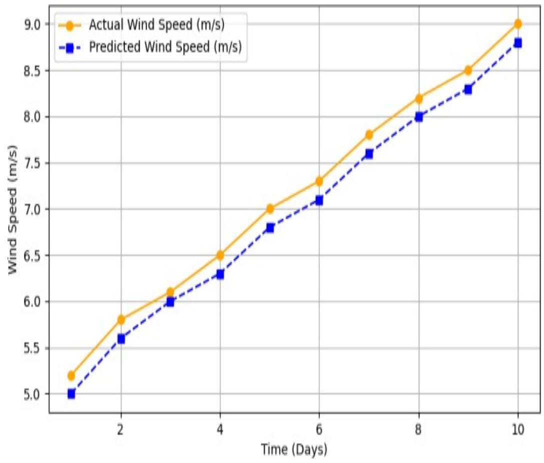

Figure 5: Wind result.

6 CONCLUSIONS

Hence, being very good in demonstrating the power of the machine-learning forecasting using XGBoost, this project integrates meteorological data of various channels, includes more complex data preprocessing steps, and extremely accurate predictions on major weather parameters like temperature, humidity, wind speed, and precipitation. Contributions toward the analysis of satellite imagery in extreme weather events will also allow those algorithms to promote further research toward breaching that wall. Future development includes merging deep learning to cloud spectrum and ensemble learning in improving the accuracy of forecasting and real- time adaptability.

In the planned future improvements of the project, model accuracy and efficiency would be enhanced by hybrid AI, deep learning algorithms, and big data analytics. One of such improvements could be in developing global weather predictions where having real-time cloud-based processing could enhance its scalability and efficiency. This includes investigating the use of other methods of ensemble learning models and automation to stitch together several forecasting techniques under one roof to enhance reliability. The system can be scaled to predict extreme weather events with higher accuracy through improved satellite image analysis and machine learning algorithms. Further, combining IoT based smart weather stations and predictive analytics can enable hyperlocal forecast benefits for agriculture, transportation, and disaster management.

REFERENCES

Anandkumar, A. (2024). Anima Anandkumar is Accelerating Scientific Discovery with AI. TIME.

Ben-Bouallegue, Z., Clare, M. C. A., Magnusson, L., et al. (2023). The rise of data- driven weather forecasting.

de Bezenac, E., Pajot, A., & Gallinari, P. (2020). Deep learning for physical processes: Incorporating prior scientific knowledge.

Evensen, G., & Monsen, S. M. (2021). Review of data assimilation in AI-based weather prediction. Weather and Climate Dynamics, 2, 317-333.

Google DeepMind (2024). Google DeepMind predicts weather more accurately than leading system. The Guardian.

Journal of Statistical Mechanics: Theory and Experiment, 2020(12), 124009.

Kashinath, K., Mustafa, M., & Albert, A. (2021). Physics-informed deep learning approaches for improving weather prediction. Proceedings of the IEEE, 109(4), 555-571.

Molchanov, S., Kitaygorodskaya, Y., & Mishin, D. (2021). Examined the fusion of AI with IoT sensor data to make predictions in real- time about the weather. Sensors, 21(14), 4852.

Pathak, J., Subramanian, S., Harrington, P., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators Keisler, R. (2022). Forecasting global weather with graph neural networks.

Photovoltaic Power Generation Forecasting Based on Random Forest Algorithm TL Yu, H Liang, L Jian 2024 8th International Conference on Green Energy and Applications. [2024] ieee.

Rasp, S., & Thuerey, N. (2021). Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for weather forecasting. Journal of Advances in Modeling Earth Systems.

Sharma, R., Gupta, A., & Patel, N. (2022). Suggested cloud-based big data analytics as a mechanism for AI-Driven weather prediction. Journal of Big Data, 9(1), 45.

Sønderby, C. K., Wang, S., Lemaire, P., et al. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting.

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Examined the ability of machine learning techniques to be used to improve the current forecasting models of the weather. Journal of Advances in Modeling Earth Systems.

Weyn, J. A., Durran, D. R., & Caruana, R. (2021). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. Journal of Advances in Modeling Earth Systems.

Zhang, J., Wang, H., & Chen, X. (2021). Investigated the application of CNNs in weather forecasting using satellite images. IEEE Access, 9, 108759-108770.