INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
VOLUME: 05 ISSUE: 02 | FEB – 2026 DOI: 10.55041/ISJEM05554
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

Machine Learning-Based Predictive Modeling of Weather Components

Monika Jakhar
Research Scholar
Department of Computer Science and Applications
Baba Mastnath University
Rohtak, Haryana, India

Dr. Vinod Kumar Srivastava
Professor
Department of Computer Science and Applications
Baba Mastnath University
Rohtak, Haryana, India

Abstract- The study focuses on the application of artificial intelligence techniques to forecast key weather parameters with improved accuracy and efficiency. Traditional weather prediction methods rely heavily on physical models and large-scale simulations, which, while effective, often face challenges in handling complex, nonlinear relationships among atmospheric variables. Machine learning (ML) offers a data-driven alternative by learning patterns from historical datasets and generating predictive models that can adapt to diverse climatic conditions. This research aims to design and evaluate ML models capable of predicting weather components such as temperature, humidity, thereby contributing to more reliable forecasting systems. The methodology involves collecting extensive meteorological datasets, preprocessing them to remove inconsistencies, and applying supervised learning algorithms such as random forests. Feature selection techniques are employed to identify the most influential variables, while cross-validation ensures robustness and generalizability of the models. The results demonstrate that ML-based models can capture nonlinear dependencies more effectively than conventional statistical approaches, offering higher accuracy in short-term predictions.

Keywords: Weather Forecasting, Machine Learning, Climate Change, Trends and challenges etc.

I. INTRODUCTION

Weather prediction [1] has always been a subject of immense importance due to its direct impact on human life, agriculture, industry, and disaster management. Accurate forecasting of weather components such as temperature, humidity, rainfall, and wind speed is essential for planning and preparedness across multiple sectors. Traditionally, weather prediction has relied on physical models and large-scale numerical simulations, which, although effective, often struggle to capture the highly nonlinear and dynamic interactions among atmospheric variables. These limitations have created a demand for innovative approaches that can complement conventional methods and provide more reliable short-term forecasts.

In recent years, artificial intelligence (AI) and machine learning (ML) [2] have emerged as transformative tools in meteorology. Unlike traditional models, ML approaches are data-driven, learning directly from historical datasets to identify hidden patterns and correlations. This ability to handle complex, nonlinear dependencies makes ML particularly [3] suitable for weather prediction, where multiple variables interact in unpredictable ways. By applying algorithms such as random forests [4], regression models, and neural networks, researchers can design predictive systems that adapt to diverse climatic conditions and improve forecasting accuracy.

The integration of machine learning [5] into weather prediction not only enhances scientific understanding but also carries significant societal benefits. Reliable short-term forecasts can help farmers plan crop cycles and irrigation, support disaster management agencies in mitigating the effects of floods and storms and assist energy sectors in optimizing renewable energy generation. Moreover, accurate weather information improves everyday decision-making for individuals and communities, contributing to safety, convenience, and resilience.This paper focuses on developing and evaluating machine learning-based models for predicting weather components, specifically temperature and humidity. By employing supervised learning algorithms,

feature selection techniques, and cross-validation, the study aims to build robust models capable of delivering accurate forecasts. The findings are expected to demonstrate the potential of ML in capturing nonlinear dependencies more effectively than conventional statistical approaches, thereby offering a pathway toward intelligent, adaptive, and socially impactful weather forecasting systems.

II. RELATED WORK

Sevgin (2025) [6] developed a machine learning-based temperature forecasting model aimed at supporting sustainable climate change adaptation and mitigation strategies. The study employed deep learning architectures, including LSTM and convolutional neural networks, to model temporal and spatial temperature variations. The model was validated using long-term climate datasets and demonstrated high accuracy in both short- and medium-term forecasts. The research underscored the role of data-driven forecasting in informing policy decisions and resource allocation for climate resilience.

Šuljug et al. (2024) [7] conducted a comparative study of machine learning models for predicting meteorological data in agricultural applications. The research evaluated models such as k-Nearest Neighbours, Random Forest, and Gradient Boosting for forecasting temperature, humidity, and rainfall in farming regions. The study demonstrated that ensemble methods provided the highest accuracy and robustness, particularly when trained on localized datasets. The findings supported the integration of machine learning into precision agriculture systems to optimize irrigation, planting schedules, and yield forecasting.

Wu and Xue (2024) [8] examined data-driven approaches to weather forecasting and climate modelling from a development perspective. The study emphasized the importance of integrating machine learning into climate services for developing countries, where traditional infrastructure may be lacking. By analysing case studies and model performance across different socioeconomic contexts, the authors highlighted the potential of AI to support sustainable development goals through improved climate risk management and informed decision-making.

Jankauskas et al. (2024) [9] sought to conduct a thorough analysis of five weather forecasting models sourced from the Open-Meteo historical data collection, focusing on their efficacy in predicting wind power output. With the growing emphasis on renewable energy, particularly wind power, precise weather forecasting is essential for optimizing energy production and maintaining the integrity of the power grid. This study's analysis includes various models: ICO sahedral Nonhydrostatic (ICON), the Global Environmental Multiscale Model (GEM Global), Meteo France, the Global Forecast System (GFS Global), and the Best Match approach. This strategy was validated by juxtaposing the model predictions with empirical data on wind power generation. The ICON model achieved a root mean squared error of 1.7565, surpassing Best Match, which recorded a root mean squared error of 1.7604, representing a minor yet significant enhancement. GEM Global and GSF Global exhibited significant alterations, with root mean squared errors (RMSEs) of 2.0086 and 2.0242, respectively, reflecting a decline in predictive accuracy of around 24% to 31% relative to ICON. The findings indicated substantial discrepancies in the accuracy of the different models employed, with some models demonstrating markedly superior predicted precision.

Guo et al. (2024) [10] created a rainfall forecasting platform utilizing the GNSS-assisted weather research and forecasting (WRF) model and quantitatively evaluated the impact of GNSS precipitable water vapor (PWV) on the accuracy of WRF model predictions for light rain (LR), moderate rain (MR), heavy rain (HR), and torrential rain (TR). In 2021, three strategies were developed and evaluated utilizing data from seven ground meteorological stations in Xi'an City, China. The findings indicated that using GNSS PWV markedly enhanced the forecasting precision of the WRF model across various rainfall intensities, with root mean square error (RMSE) improvement rates of 8%, 15%, 19%, and 25% for LR, MR, HR, and TR, respectively. The results validated the platform's high precision, visualization capabilities, and robustness in rainfall forecasting.

Teixeira et al. (2024) [11] utilized regression models to estimate absent historical data across three distinct time horizons, integrating long short-term memory (LSTM) to predict short- to medium-term weather conditions at Quinta de Santa Bárbara in the Douro region. A genetic algorithm (GA) was employed to optimize the hyperparameters of the LSTM. The findings indicated that the proposed improved LSTM significantly diminished the assessment measures across various time horizons. The results emphasized the significance of precise weather forecasting in informing critical choices across all industries.

Zhang et al. (2024) [12] enhanced the precision of atmospheric temperature and humidity profile retrieval and examined the impacts of cloud data (cloud-base height and cloud thickness) on these retrievals. The observational data from the ground-based multichannel

microwave radiometer (GMR) and the millimetre-wave cloud radar (MWCR) were integrated into the retrieval procedure of atmospheric temperature and relative humidity profiles. The retrieval was executed with the backpropagation neural network (BPNN). The retrieval outcomes were evaluated using the mean absolute error (MAE) and root mean square error (RMSE). The statistical results indicated that the temperature profiles were less influenced by the cloud data than the relative humidity profiles. In comparison to the retrieval profiles devoid of cloud information, the MAE and RMSE values for most height levels exhibited varying degrees of reduction following the incorporation of cloud data, with the relative humidity (RH) errors in certain altitude layers diminishing by almost 50%. The greatest decrease in RMSE and MAE values for temperature profile retrieval using cloud data was approximately 1.0 ◦C at 7.75 km, whereas the maximum reduction in RMSE and MAE values for relative humidity profiles was roughly 10%, achieved at 2 km.

Sim et al. (2024) [13] introduced a sophisticated ocean fog prediction model for the Yellow Sea region, utilizing satellite-based detection and high-performance data-driven techniques. The study utilized Himawari-8 satellite data to acquire extensive spatiotemporal ocean fog references and implemented Auto ML to integrate numerical weather prediction (NWP) outputs with sea surface temperature (SST)-related variables. The model exhibited enhanced performance relative to conventional NWP-based approaches, with impressive metrics in both quantitative probability of detection at 81.6%, false alarm ratio at 24.4%, F1 score at 75%, and proportion correct at 79.8%and qualitative assessments over lead periods ranging from 1 to 6 hours. Significant contributing factors comprised relative humidity, cumulative shortwave radiation, and atmospheric pressure, underscoring the necessity of using varied data sources. The study highlighted the potential of satellite-derived data to enhance ocean fog prediction, while also addressing the issues of overfitting and the necessity for more complete reference data.

Shiferaw et al. (2024) [14] sought to determine an optimal configuration of the Weather, Research, and Forecasting (WRF) model for Ethiopia. Thirty-five WRF simulations employing various combinations of parameterization approaches for cumulus (CU), planetary boundary layer (PBL), cloud microphysics (MP), longwave (LW), and shortwave (SW) radiation were evaluated throughout the summer season (June to August, JJA) of 2002. The WRF simulations employed a two-domain design featuring a 12 km nested domain encompassing Ethiopia. The starting and boundary forcing data for WRF were sourced from the Climate Forecast System Reanalysis (CFSR). The simulations were assessed against station and gridded observations to determine their efficacy in replicating various characteristics of JJA rainfall. An objective ranking methodology employing an aggregate score of multiple statistics was utilized to identify the optimal model configuration. All models accurately represented the regional distribution of JJA rainfall, with the pattern correlation coefficient (PCC) varying between 0.89 and 0.94.

Ahmadgourabi et al. (2024) [15] sought to develop a dependable water-demand forecasting system utilizing Long Short-Term Memory networks. The model incorporated hourly water needs from ten District Metered Areas within a Water Distribution Network in northeastern Italy, alongside weather data, addressing missing values through LSTM-based data imputation. It took into account temporal factors such as time, weekdays, holidays, and weekend routines, utilizing sine and cosine transforms to represent daily cycles. The model's robustness was ensured by conducting testing in the final week of the dataset, especially week 81, while making iterative tweaks to the LSTM's hyperparameters to enhance prediction accuracy. The tuning efforts concentrated on the learning rate, layer count, and batch size, optimized to enhance the system's performance.

A. Research Gap

Although traditional weather prediction methods based on physical models and numerical simulations have been widely used, they often struggle to capture the nonlinear and dynamic interactions among atmospheric variables, especially for short-term forecasts. These approaches require extensive computational resources and may not adapt well to localized climatic variations. While machine learning has recently been applied to weather prediction, most studies have focused on limited parameters or specific regions, leaving gaps in comprehensive modeling of multiple weather components such as temperature and humidity. Furthermore, there is insufficient exploration of how feature selection and algorithmic optimization can enhance predictive accuracy and generalizability. This gap highlights the need for robust, data-driven models that can complement conventional approaches and provide more reliable short-term forecasts.

III. OBJECTIVE OF WORK

A. Problem Statement

The central problem addressed in this study is the inaccuracy and inefficiency of traditional weather prediction methods in handling nonlinear relationships among atmospheric variables for short-term forecasts. Conventional models often fail to capture localized variations and require significant computational resources, limiting their adaptability. The problem is therefore formulated as the need to develop a machine learning-based predictive framework that can learn from historical meteorological data, identify key influencing factors, and generate accurate forecasts of weather components such as temperature and humidity. By applying supervised learning algorithms like random forests, the study seeks to overcome the limitations of traditional approaches, ensuring robustness, generalizability, and improved accuracy in short-term weather prediction.

B. Research Objective

The primary objective of this study is to design and evaluate machine learning-based models for predicting key weather components, specifically temperature and humidity. To achieve this, the study aims to preprocess and analyze meteorological datasets to ensure consistency, apply supervised learning algorithms such as random forests, and employ feature selection techniques to identify the most influential variables.

IV. MACHINE LEARNING BASED PREDICTIVE MODELING OF WEATHER COMPONENTS

The application of machine learning (ML) in weather prediction represents a significant advancement over traditional forecasting methods. Conventional approaches rely on physical models and numerical simulations, which, although effective, often struggle with the nonlinear and dynamic interactions among atmospheric variables. Machine learning, on the other hand, offers a data-driven alternative by learning directly from historical datasets and identifying hidden patterns that influence weather components such as temperature, humidity, rainfall, and wind speed. This ability to capture complex relationships makes ML particularly suitable for short-term forecasting, where accuracy and adaptability are critical.

In this study, extensive meteorological datasets were collected and preprocessed to ensure consistency and reliability. Preprocessing involved cleaning the data, handling missing values, and normalizing variables to eliminate inconsistencies that could affect model performance. Feature selection techniques were then applied to identify the most influential parameters, such as atmospheric pressure, relative humidity, and temperature trends, which play a key role in determining weather outcomes. By focusing on relevant features, the models were optimized for efficiency and accuracy, reducing computational complexity while improving predictive power.

Supervised learning algorithms formed the backbone of the predictive modeling framework. Among these, random forests were employed due to their robustness in handling nonlinear dependencies and their ability to reduce overfitting through ensemble learning. The algorithm constructs multiple decision trees and aggregates their outputs, resulting in more reliable predictions. Cross-validation techniques were used to test the generalizability of the models across different datasets, ensuring that the predictions were not limited to specific conditions but adaptable to diverse climatic scenarios. Comparative analysis with conventional statistical methods demonstrated that ML-based models consistently outperformed traditional approaches in short-term forecasting accuracy.

V. MAJOR FINDINGS

The study confirms that machine learning-based models, particularly random forests, outperform conventional statistical approaches in predicting short-term weather components such as temperature and humidity. By leveraging historical meteorological datasets and applying feature selection techniques, the models were able to capture complex nonlinear dependencies among atmospheric variables more effectively than traditional methods. This resulted in improved accuracy, robustness, and adaptability across diverse climatic conditions, demonstrating the strength of ML as a data-driven alternative to physical simulations.

Another significant finding is that feature selection and preprocessing play a critical role in enhancing model performance. By identifying the most influential variables, such as atmospheric pressure and relative humidity, the models achieved greater efficiency and reduced computational complexity. Cross-validation further ensured generalizability, confirming that the predictive framework is not limited to specific datasets but can be applied to broader scenarios. Overall, the results highlight that ML-based predictive modeling provides a reliable, flexible, and scalable solution for weather forecasting, with strong potential for integration into real-world applications such as agriculture, disaster management, and renewable energy optimization.

VI. CONCLUSION

The study demonstrates that machine learning provides a robust and efficient framework for predicting weather components such as temperature and humidity. By leveraging algorithms like random forests, the models can capture complex, nonlinear relationships among atmospheric variables that traditional statistical approaches often fail to address. The results confirm that ML-based models deliver higher accuracy in short-term forecasts, offering a reliable alternative to conventional methods. This establishes machine learning as a valuable tool in meteorology, capable of enhancing forecasting systems and supporting decision-making in sectors dependent on accurate weather information.

Looking ahead, the scope of this research can be expanded in several directions. Incorporating additional weather parameters such as rainfall, wind speed, and solar radiation would make the models more comprehensive and applicable to diverse climatic conditions. Hybrid approaches that combine physical simulation models with machine learning could further improve predictive accuracy by integrating domain knowledge with data-driven insights. The use of advanced techniques such as deep learning, ensemble modeling, and real-time data integration from IoT-based weather sensors can enhance adaptability and precision.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of researchers and institutions whose pioneering work in machine learning and weather forecasting analysis has laid the foundation for this review. It extends our sincere thanks to the developers of open-access datasets and platforms, which have enabled reproducible and data-driven research in this domain.

REFERENCES

[1] Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.

[2] Wilks, D.S.; Hamill, T.M. Comparison of ensemble-MOS methods using GFS reforecasts. Mon. Weather. Rev. 2007, 135, 2379–2390

[3] Akaike, H. A New Look at Statistical Model Identification. IEEE Trans. Automat Contr. 1974.

[4] Ejike, O.; Ndzi, D.; Shakir, M.Z. Comparative Study of Machine Learning-Based Rainfall Prediction in Tropical and Temperate Climates. Climate 2025, 13, 167. https://doi.org/ 10.3390/cli13080167

[5] Jankauskas, M.; Serackis, A.; Paulauskas, N.; Pomarnacki, R.; Hyunh, V.K. (2024) The Impact of the Weather Forecast Model on Improving AI-Based Power Generation Predictions through BiLSTM Networks. *Electronics*, *13*, 3472. https://doi.org/10.3390/ electronics13173472

[6] Sevgin, F. (2025), Machine Learning-Based Temperature Forecasting for Sustainable Climate Change Adaptation and Mitigation. *Sustainability*, *17*, 1812. https:// doi.org/10.3390/su17051812.

[7] Šuljug, J.; Spiši´c, J.; Grgi´c, K.; Žagar, D. (2024), A Comparative Study of Machine Learning Models for Predicting Meteorological Data in Agricultural Applications. *Electronics*, *13*, 3284. https://doi.org/ 10.3390/electronics13163284

[8] Wu, Y.; Xue, W. (2024), Data-Driven Weather Forecasting and Climate Modeling from the Perspective of Development. *Atmosphere*, *15*, 689. https://doi.org/10.3390/ atmos15060689.

[9] Jankauskas, M.; Serackis, A.; Paulauskas, N.; Pomarnacki, R.; Hyunh, V.K. (2024) The Impact of the Weather Forecast Model on Improving AI-Based Power Generation Predictions through BiLSTM Networks. *Electronics*, *13*, 3472. https://doi.org/10.3390/ electronics13173472.

[10] Guo, H.; Ma, Y.; Li, Z.; Zhao, Q.; Zhai, Y. (2024) The Evaluation of Rainfall Forecasting in a Global Navigation Satellite System-Assisted Numerical Weather Prediction Model. *Atmosphere 15*, 992. https://doi.org/ 10.3390/atmos15080992.

[11] Teixeira, R.; Cerveira, A.; Pires, E.J.S.; Baptista, J. (2024) Enhancing Weather Forecasting Integrating LSTM and GA. *Appl. Sci.*, *14*, 5769. https://doi.org/10.3390/ app14135769.

[12] Zhang, L.; Ma, Y.; Lei, L.; Wang, Y.; Jin, S.; Gong, W. (2024) Improving Atmospheric Temperature and Relative Humidity Profiles Retrieval Based on Ground-Based Multichannel Microwave Radiometer and

Millimeter-Wave Cloud Radar. *Atmosphere 15*, 1064. https:// doi.org/10.3390/atmos15091064.

[13] Sim, S.; Im, J.; Jung, S.; Han, D. (2024) Improving Short-Term Prediction of Ocean Fog Using Numerical Weather Forecasts and Geostationary Satellite-Derived Ocean Fog Data Based on AutoML. *Remote Sens.*, *16*, 2348. https://doi.org/10.3390/ rs16132348.

[14] Shiferaw, A.; Tadesse, T.; Rowe, C.; Oglesby, R. (2024) Weather Research and Forecasting Model (WRF) Sensitivity to Choice of Parameterization Options over Ethiopia. *Atmosphere*, *15*, 974. https://doi.org/10.3390/ atmos15080974

[15] Boloukasli ahmadgourabi, F.; Khashei Varnamkhasti, M.; Nosrati Habibi, M.; Hedaiaty Marzouny, N.; Dziedzic, R. (2024) Enhancing Water Demand Forecasting: Leveraging LSTM Networks for Accurate Predictions. *Eng. Proc.*, *69*, 120. https://doi.org/10.3390/ engproc2024069120