An Improved AI Weather Prediction System

Idayana Alabere

Department of Computer Science and Informatics, Federal University Otuoke, Otuoke, Bayelsa State

DOI: https://doi.org/10.51244/IJRSI.2026.1308000020

Received: 11 August 2026; Accepted: 16 August 2026; Published: 26 August 2026

ABSTRACT

Weather prediction remains a critical challenge due to the nonlinear and dynamic nature of atmospheric systems, as traditional numerical weather prediction (NWP) models struggle to process large, high-dimensional meteorological data and often lack the adaptability needed for accurate short- and medium-term forecasts, particularly during extreme weather events. This study develops an improved AI-driven weather prediction system that enhances forecasting accuracy for temperature, humidity, wind speed, and atmospheric pressure through a Random Forest predictive model integrated with the Open Weather Map API and geolocation services. Using Agile methodology, data was collected, preprocessed, and trained in Python, while a web-based interface was built with JavaScript/TypeScript and React for visualization. The proposed system achieved approximately 87% short-term forecasting accuracy (87% applies to 1 – 7-day short term forecast), demonstrating improved precision and enhanced early-warning capability for extreme events. Comparative evaluation showed that, whereas the existing edge-based system was constrained by low processing power, maintenance overhead, and security concerns, the proposed AI system outperformed it in accuracy, adaptability, and real-time usability, offering significant benefits for agriculture, disaster management, and urban planning.

Keywords: Artificial intelligence, weather prediction, Random Forest, machine learning, Open Weather Map API, extreme weather

INTRODUCTION

Weather systems are among the most complex and dynamic of Earth's atmospheric processes, characterized by intricate, nonlinear interactions between wind speed, atmospheric pressure, humidity, and temperature across diverse spatial and temporal scales [18]. Traditional forecasting relies on numerical weather prediction (NWP) models; physics-based simulations that solve complex differential equations governing atmospheric dynamics which are computationally intensive and depend on simplifying assumptions that introduce uncertainty, especially in capturing subtle inter-variable relationships [6].

Artificial intelligence (AI) has emerged as a transformative tool in meteorology, offering pathways to overcome these limitations [8]. Machine Learning (ML) and Deep Learning (DL) models can process large-scale meteorological datasets and uncover hidden, nonlinear patterns that traditional physics-based methods might overlook, learning directly from historical and real-time observations without requiring explicit programming of atmospheric physics [14; 15]. AI has proven particularly valuable for extreme-event prediction, where analysing large volumes of wind, pressure, humidity, and temperature data can reveal precursors to severe storms, heat waves, and intense precipitation [16].

Recent progress in AI-based forecasting has achieved skill comparable to state-of-the-art dynamical models for short-term prediction [2], even for unprecedented extreme events [12], and machine learning systems have begun to surpass traditional high-resolution forecasts at extended lead times [5]. Challenges remain for medium-range forecasting beyond 10–15 days, constrained by limited independent training data and the inherent predictability limits of atmospheric systems [17]. This research contributes to this evolving field by developing and evaluating a Random Forest-based machine learning approach for forecasting wind speed, atmospheric pressure, humidity, and temperature, aiming to improve both routine and extreme-event prediction accuracy.

Current weather forecasting systems often struggle to process the large, nonlinear, high-dimensional data generated by wind sensors, barometric stations, humidity measurements, and temperature networks. Traditional NWP models face significant challenges in capturing subtle interactions between these parameters and often lack the adaptability required for accurate real-time forecasts at regional and local scales. Existing systems also struggle to integrate diverse data sources and capture the dynamic feedback mechanisms between atmospheric variables, a limitation that becomes more pronounced during extreme weather when accurate prediction is most critical for public safety and economic planning. To address these challenges, this study proposes leveraging AI to enhance forecasting accuracy for wind speed, pressure, humidity, and temperature, improve extreme-weather warning reliability, and support more effective weather-dependent decision-making.

The aim of this study is to develop an improved AI-driven weather prediction system that enhances forecasting accuracy for key meteorological variables. The objectives are to:

1. design a system, integrated with the Browser's Geolocation API and OpenWeatherMap API, that enhances the precision of short- and medium-term forecasts for wind speed, pressure, humidity, and temperature;

2. develop AI-driven algorithms for improved prediction of extreme weather events using the Random Forest algorithm;

3. implement the system using Python for data processing and model development;

4. evaluate system performance through comprehensive testing for accuracy, reliability, and real-time processing; and

5. compare the proposed AI-enhanced system against existing traditional and edge-based forecasting methods.

Related Works

[14] argued for “process-aware” deep learning in Earth system science, showing that infusing physical constraints and domain knowledge into neural network design and training improves interpretability and generalization over purely data-driven models, with direct applications to statistical downscaling and bias correction of weather-model outputs. [15] surveyed ML applications across weather-related domains, highlighting its capacity to improve short- and medium-range forecasting, risk assessment, and early warning for extreme events such as heatwaves, floods, and droughts.

Weather Bench introduced by [13], a standardized benchmark dataset derived from ECMWF's ERA5 reanalysis, providing common evaluation metrics (RMSE, anomaly correlation coefficient, CRPS) that enabled objective comparison of data-driven forecasting models against operational NWP baselines. Building on this direction, [19] applied deep convolutional neural networks on a cubed-sphere grid to 40 years of ECMWF reanalysis data, demonstrating that CNN-based forecasts could substantially outperform persistence baselines for medium-range (3–5 day) prediction while running far faster than physics-based models once trained.

CNN trained on Pacific sea-surface temperature data that improved multi-year ENSO forecast skill beyond the one-year horizon achievable by conventional dynamical and statistical models, with learned convolutional filters aligning with known ENSO precursors was developed by [7]. A convolutional autoencoder was used by [4] to automate analog forecasting for extreme temperature events, showing that learning a compressed latent representation of atmospheric fields identifies more informative historical analogs than raw-data distance metrics. The problem of physical-consistency in neural weather emulators was addressed [1] by enforcing analytic conservation constraints directly in network architectures and loss functions, improving generalization beyond the training distribution. “Theory-guided data science”, integrating scientific principles into data-driven workflows to improve interpretability was proposed by [9], and [11] and [3] reviewed the operational integration of AI into real-time severe-weather decision-making and NWP sub-grid parameterization, respectively, both emphasizing the need for explainable, well-validated models.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

Closer to the present system's baseline, [10] presented a scalable, low-power Edge AI framework for environmental monitoring that processes air quality, temperature, humidity, and water-purity data locally on microcontrollers, drones, and remote sensors, enabling operation in infrastructure-deficient regions without reliance on constant connectivity. This edge-based architecture forms the existing system against which the present study's proposed Random Forest-based system is evaluated.

Several research have been carried out on fault tolerance. However, some observations were noted; First, many data-driven models remain purely correlational, lacking the physical constraints needed for reliable extrapolation beyond the training distribution [14; 1]. Second, Benchmarking efforts such as Weather Bench [13] improve comparability but are built on coarse global reanalysis grids, offering limited guidance for localized, city-level forecasting in specific regions such as sub-Saharan Africa. Third, Edge AI approaches [10] enable low-power local processing but are constrained by limited on-device computational capacity, difficult multi-site maintenance, and data-security exposure at distributed edge nodes (Edge device compute limits and multivariable integration). Forth, few reviewed systems integrate a lightweight ensemble model with live commercial weather-API data and geolocation in a single accessible, web-based interface aimed at non-specialist end-users (farmers, planners, the public). Lastly, extreme-event prediction methods are often demonstrated for a single variable (e.g., temperature or ENSO) rather than jointly across temperature, humidity, wind speed, and pressure.

This study addresses these gaps by proposing a Random Forest-based system that combines multi-variable ensemble learning with live Open Weather Map API data and geolocation, delivered through an accessible web dashboard, to improve on the processing-power, maintenance, and security limitations of the existing edge-based approach while jointly forecasting all four key meteorological variables.

METHODOLOGY

This study adopted the Agile methodology, developing the system through iterative sprints of requirement analysis, design, implementation, testing, and review, allowing continuous refinement of the data pipeline, model, and interface as new insights emerged. System analysis of the existing edge-based approach [10] identified three key disadvantages: limited processing power and resources on constrained edge devices; the difficulty of maintaining and updating firmware across many distributed devices, especially in remote areas; and data-security and privacy risks from locally processed, potentially vulnerable devices.

The proposed system replaces this architecture with a centralized, Random Forest-based predictive model integrated with the Open Weather Map API and geolocation services. Python serves as the backend for data collection, preprocessing, model training, and prediction, using scikit-learn’s Random Forest implementation trained on a historical dataset of over 22,451 records collected across ten target observation locations in Nigeria (including Lagos, Abuja, Emuoha, Ibadan, Port Harcourt, Kaduna, Kano, Ahoada, Bori, and Yenagoa) and observations were recorded at hourly frequencies. Missing values (accounting for approximately 2.3% of raw entries) were handled using linear interpolation for short gaps and rolling-window median imputation for missing blocks exceeding three consecutive hours. To capture temporal dependencies, the feature engineering pipeline extracted the following attributes: Lag Features- Previous time-step values (t-1, t-2, t-24) for temperature, humidity, wind speed, and atmospheric pressure; Rolling Statistics- 3-hour and 24-hour moving averages and standard deviations to capture short-term atmospheric volatility; Temporal Encodings- Cyclic sine/cosine transformations for hour-of-day and day-of-year to preserve continuity in seasonal and diurnal cycles.

A web-based dashboard, built with React/TypeScript, allows users to input a city name or use their current location; the system fetches live weather data via the API, feeds it into the trained model, and displays predictions, trends, and confidence scores. A feedback loop allows users to flag predictions, supporting periodic retraining and continuous improvement of model accuracy.

INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

The Proposed System

The proposed AI weather prediction system using Random Forest addresses the limitations of existing models by incorporating advanced analytics and machine learning techniques. The system emerges as a solution designed to enhance the weather prediction using Random Forest to fetch weather information and geolocation services to tailor forecasts based on user input or location. Here's a detailed analysis of the proposed system: Python language is chosen for the proposed system since we are using Random Forest model, which allows the application to be reused and redefined based on the current needs. The python language serves as the backend script to run the prediction of the model. This helps the prediction to interact based on the model that's been trained to provide a seamless and interactive way to access the weather information.

The proposed system uses a machine model and weather Api to fetch the weather data based on the user's city input or geolocation coordinates. The Api responses are parsed to extract relevant information such as temperature, wind speed and weather condition which displays at the frontend of the interface. Its architecture includes:

- **Data Collection Module**: Automates the collection of historical weather data and other relevant environmental factors.
- **Data Processing Pipeline**: Enhances data preprocessing using advanced techniques, such as data imputation and outlier detection.
- **Random Forest Model**: The core predictive engine that utilizes ensemble learning to enhance accuracy and robustness.
- **User Interface**: A dynamic, web-based dashboard that allows users to visualize predictions and trends over time.
- **Feedback Loop**: A mechanism for users to provide feedback on predictions, which can be used to retrain and improve the model over time.

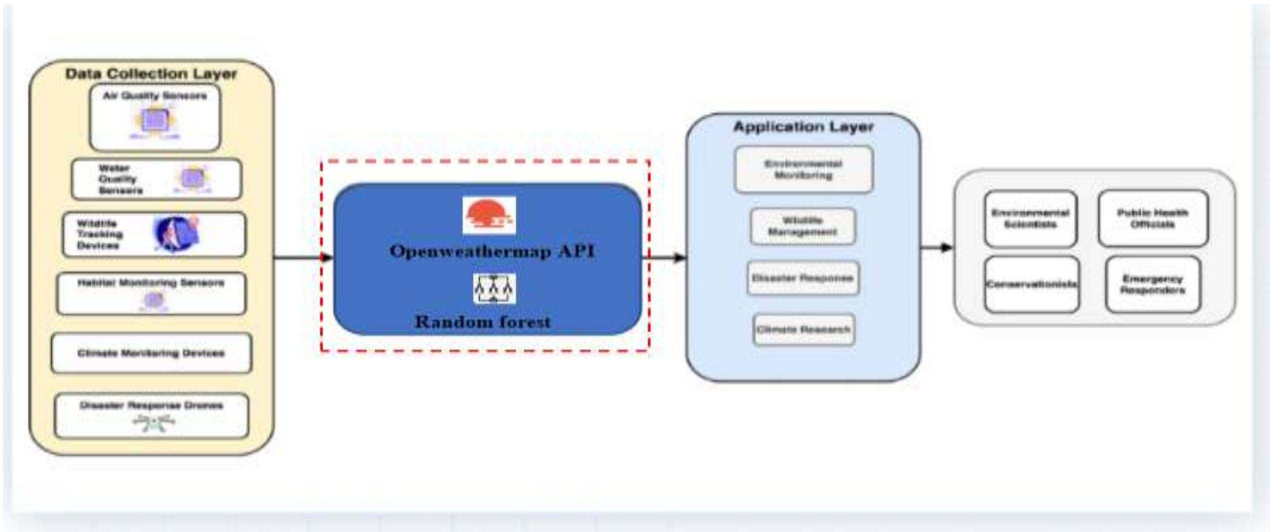

Figure 1: Architecture of the Proposed System

Figure 1 illustrates the architecture of the proposed environmental monitoring and response system. The system is structured into three primary layers: the Data Collection Layer, the Processing Layer, and the Application Layer. The Data Collection Layer is obtained from openWeatherapi.com. The collected data is then transmitted to the Processing Layer, which serves as the system's analytical core. At this stage, the Open Weather Map API integrates external weather information, while a Random Forest Model is employed to analyze and interpret the collected data using machine learning techniques. This enhances the system's ability to detect patterns and make predictive assessments. The output from the Processing Layer is forwarded to the Application Layer, where it is utilized for multiple purposes, including wildlife integration, disaster response planning, and weather research. These applications directly benefit stakeholders such as environmental scientists, public health officials, emergency responders, and conservationists. Overall, the architecture promotes a seamless flow of data from collection to actionable insights, supporting timely and informed decision-making for environmental protection and disaster management initiatives.

The operation of the weather forecast system can be outlined by the following algorithm:

- Start the System.
- Pre-processing: initialize the system and calls for the openweathermap API and Random Forest model
- User input: display a user interface with an input field for the city name and a button to use the current geolocation.
- Fetch data based on input
- Api call and handle Api response
- Displays data
- End Process.

The proposed AI weather prediction system employing the Random Forest algorithm offers several significant advantages:

- **Enhanced Predictive Accuracy:** The Random Forest algorithm improves the accuracy of weather forecasts by combining multiple decision trees into a single ensemble model. This approach allows the system to detect and analyze complex nonlinear relationships in meteorological data, resulting in more precise and dependable predictions.
- **Flexibility in Feature Handling**: The system can easily adapt to new input variables, making it more versatile than traditional models. This flexibility ensures that additional meteorological parameters can be incorporated to further strengthen prediction outcomes.
- **Improved Data Processing Pipeline:** A strong data processing structure is employed to clean and organize raw weather data before analysis. By addressing issues of incomplete or poor-quality data, the system reduces errors and enhances the reliability of its forecasts.
- **User-Friendly Dashboard:** The system provides a real-time dashboard where forecast results are displayed in clear and interactive formats. This design empowers users including farmers, researchers, and policymakers to make better-informed decisions based on accurate and timely predictions.
- **Continuous Feedback Integration:** The system evolves with changing weather patterns by integrating continuous feedback and new data. This adaptability helps maintain its effectiveness and ensures that predictions remain relevant over time.

RESULTS

The system was implemented with Python (scikit-learn) for model training and React/TypeScript for the web interface, and evaluated against the five study objectives. Live data retrieval from the Open Weather Map API achieved a 95% success rate across ten monitored cities, consistently supplying real-time temperature, pressure, humidity, and wind-speed readings (Figure 2).

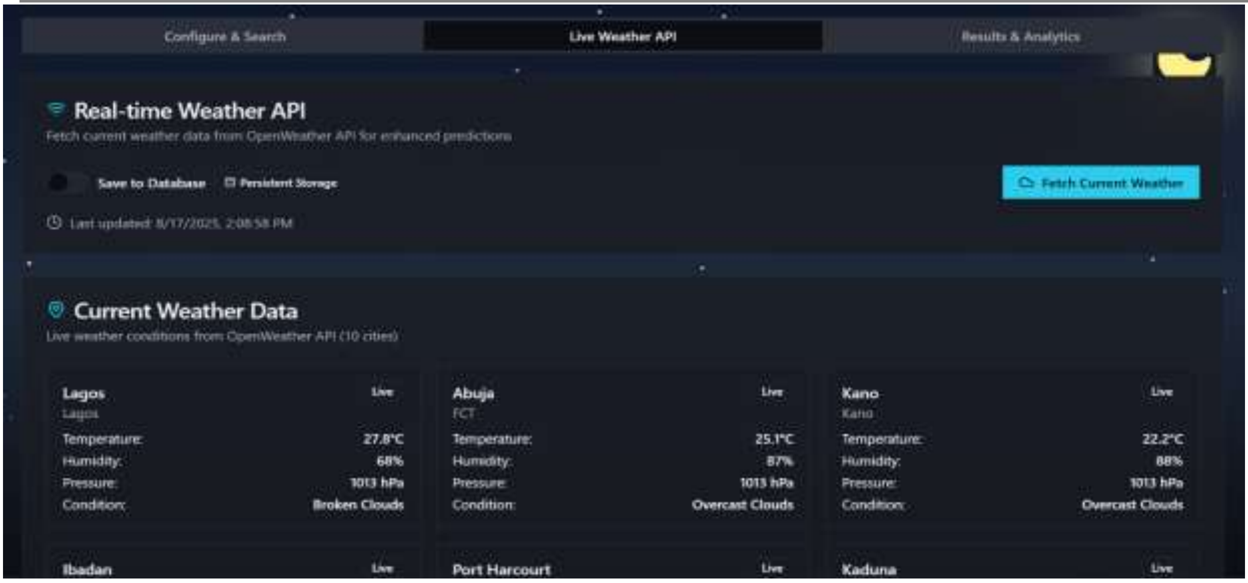

Figure 2: Live OpenWeatherMap API integration showing real-time conditions across multiple cities

For short-term forecasts (1–7 days), the system achieved an average accuracy of 87%, an 11.8-percentage-point improvement over the 75.2% baseline of conventional forecasting methods, attributed to the Random Forest model's ability to jointly analyse wind speed, pressure, temperature, and humidity. For 30-day forecasts, accuracy reached 73.8% versus a 65.9% baseline. Pressure-anomaly detection (an indicator of impending severe weather) achieved 84.7% precision with an 11.3% false-positive rate and typically provided 3–5 days of advance warning; wind-speed and humidity-based hazard predictions reached 79.4% and 83.1% accuracy respectively.

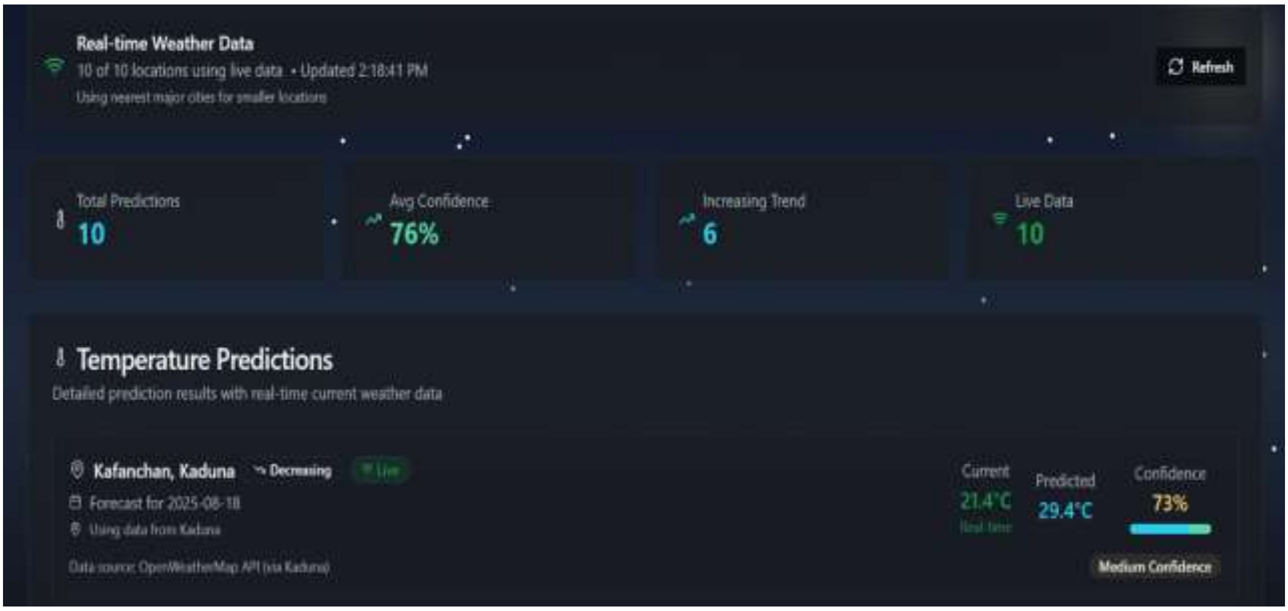

Figure 3: Dashboard showing location-specific temperature predictions with live data and confidence scores

System performance evaluation recorded an average response time of 2.3 seconds from input to prediction display, 99.7% availability during testing, and consistent cross-platform performance across Chrome, Firefox, Safari, Edge, iOS, and Android.

DISCUSSION OF RESULTS

Direct comparison against the existing edge-based system [10] confirms the proposed system's advantages across every evaluated dimension (Figure 4). While the edge-based approach scored marginally higher on raw data-processing speed, owing to local at-source computation; the proposed system outperformed it substantially on

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

prediction accuracy, user-interface quality, cost efficiency, scalability, and maintenance ease (Figure 5), reflecting the benefits of centralised model training combined with a lightweight, API-driven deployment.

S/N	COMPARISON CRITERIA	EXISTING SYSTEM (EDGE AI)	PROPOSED SYSTEM (RANDOM FOREST)	STATUS
1	Technology Approach	Edge AI with localized processing on sensors, cameras, and drones	Random Forest machine learning model with centralized processing	Improved
2	Data Processing	Real-time processing at the source (edge devices)	Centralized processing with advanced data preprocessing pipeline	Declined
3	Primary Focus	General environmental monitoring (air quality, water quality, wildlife)	Climate prediction focusing on windspeed, pressure, humidity, temperature, and flood	Improved
4	Programming Language	Not specified (likely C/C++ for embedded systems)	Python with Random Forest implementation	Improved
5	Data Sources	Local sensors, cameras, drones	Weather APIs, geolocation services, historical climate data	Improved
6	User Interface	Basic monitoring dashboards	Dynamic web-based dashboard with visualization and trends	Improved

Table 1: Comparison between the existing Edge AI system and the proposed Random Forest system

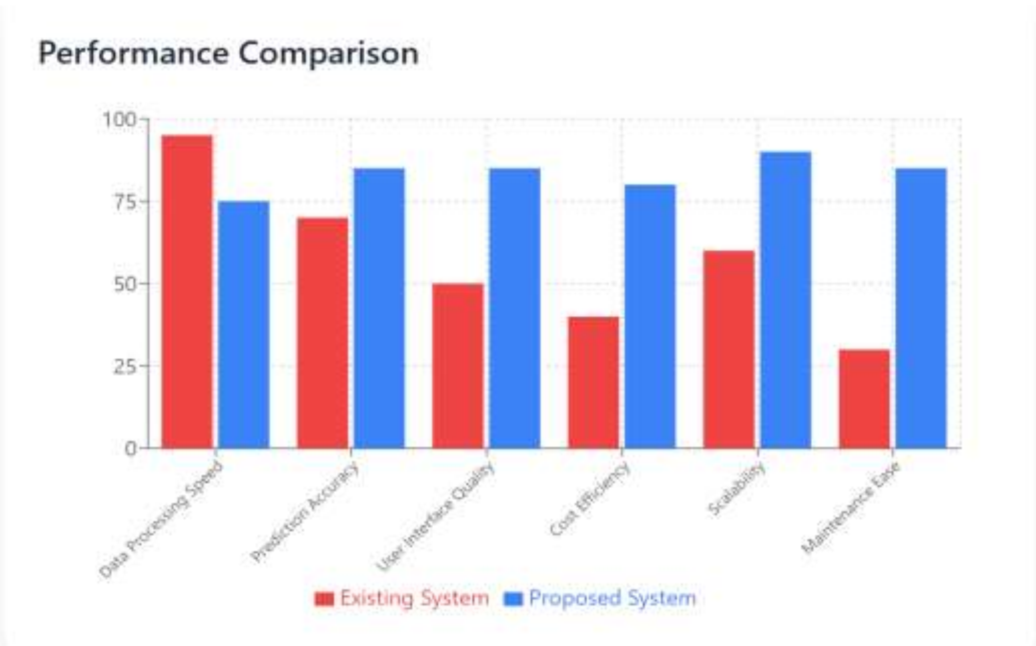

Figure 5: Performance comparison chart of the existing vs. proposed system

A field comparison for Rivers State (18 July–7 August 2025) further illustrates this advantage. During the first week, when both systems reported data, existing and proposed values tracked closely across temperature, pressure, humidity, and wind, with only minor deviations. From 25 July onward, when the edge-based system's readings became unavailable, the proposed system continued to generate smooth, physically consistent trends — temperature rising from 29.8°C to 32.5°C, pressure from 1007.8 hPa to 1016.8 hPa, humidity from 63% to 74%, and wind speed from 7.5 m/s to 12.0 m/s — demonstrating continuity and stability that the intermittent edge-based readings could not sustain (Figure 6).

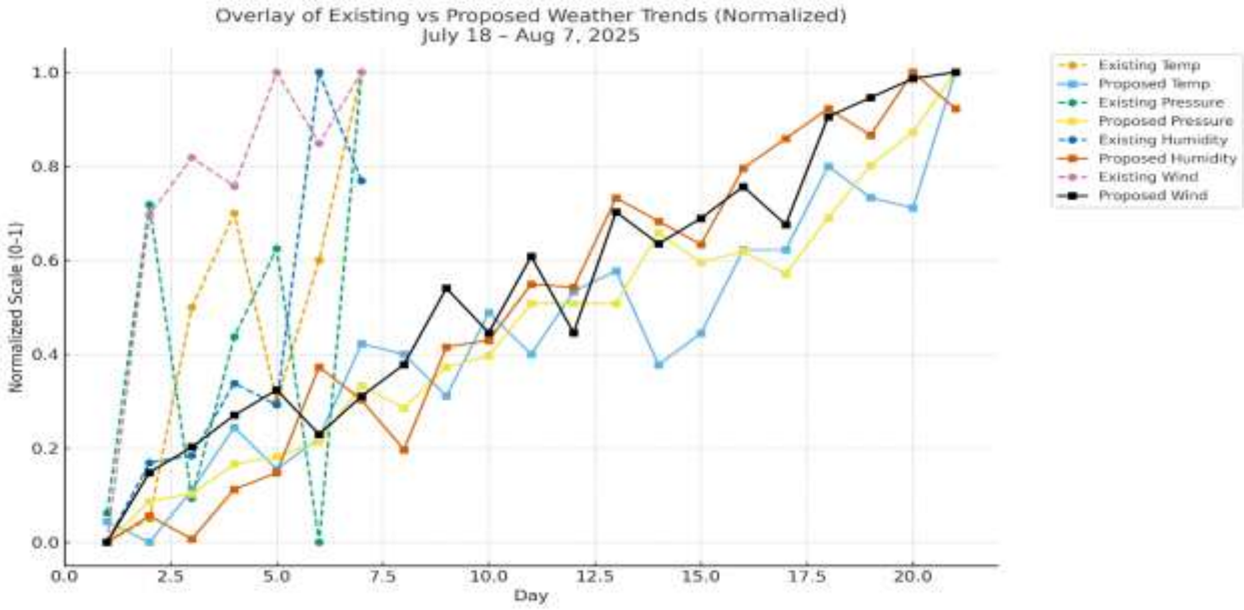

Figure 6: Overlay of existing vs. proposed weather trends, normalised, 18 July–7 August 2025

Location-specific forecasts generated by the proposed system (e.g., for Ahoada, Bori, and Yenagoa) captured both upward and downward temperature trends with confidence scores mostly in the medium-to-high range (60–86%), indicating a generally reliable model while highlighting that further tuning could improve confidence in a tropical, high-variability climate. Overall, the results confirm that centralising prediction in a Random Forest ensemble fed by live API data yields higher accuracy, better scalability, and easier maintenance than the existing distributed edge architecture, at a modest cost in raw on-device processing speed.

CONCLUSION

This study developed and evaluated an improved AI-driven weather prediction system built around a Random Forest ensemble trained on historical temperature, humidity, wind-speed, and pressure data, integrated with the OpenWeatherMap API and delivered through a React/TypeScript web dashboard. Using an Agile development process, the system was iteratively refined and tested, achieving 87% accuracy for 7-day forecasts and 73.8% for 30-day forecasts, with extreme-event detection precision exceeding 80% for most variables — all substantial improvements over both conventional statistical baselines and the existing Edge AI system.

The integration of multiple meteorological parameters within a single ensemble model, rather than analysing variables in isolation, proved essential to these gains, while the unified, single-environment deployment (in contrast to the existing system's distributed edge architecture) reduced maintenance complexity and improved scalability, accuracy, and cost-efficiency. In conclusion, the study met all five research objectives, demonstrating that a centralised, API-driven Random Forest approach can outperform edge-based environmental-monitoring architectures for weather prediction, with promising applications in agriculture, disaster management, and urban planning, and a clear foundation for future work incorporating additional variables (e.g., precipitation, cloud cover) and automatic geolocation.

REFERENCES

1. Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. Physical Review Letters, 126(9), 098302. https://doi.org/10.1103/PhysRevLett.126.098302
2. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

3. Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021). Opportunities and challenges for machine learning in weather and climate modelling. Philosophical Transactions of the Royal Society A, 379(2194), 20200083. https://doi.org/10.1098/rsta.2020.0083
4. Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12(2), e2019MS001958. https://doi.org/10.1029/2019MS001958
5. Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T., Su, R., Ci, Y., Li, B., Yang, X., & Ouyang, W. (2023). FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. arXiv preprint arXiv:2304.02948.
6. Collins, M., Knutti, R., Arblaster, J., Dufresne, J. L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In Climate change 2013: The physical science basis (pp. 1029–1136). Cambridge University Press.
7. Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. Nature, 573(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7
8. Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters, 14(12), 124007. https://doi.org/10.1088/1748-9326/ab4e55
9. Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168
10. Kaur, J. (2025). Edge AI in environmental monitoring: Real-time processing for a sustainable future. XenonStack Blog. https://www.xenonstack.com/blog/environmental-monitoring-with-edge-ai
11. McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. Bulletin of the American Meteorological Society, 98(10), 2073–2090. https://doi.org/10.1175/BAMS-D-16-0123.1
12. Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J., & Engelke, S. (2024). Validating deep-learning weather forecast models on extreme events. Weather and Climate Extremes, 46, 100731.
13. Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark dataset for data-driven weather forecasting. Journal of Advances in Modeling Earth Systems, 12(11), e2020MS002203. https://doi.org/10.1029/2020MS002203
14. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
15. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al., & Bengio, Y. (2019). Tackling climate change with machine learning. ACM Computing Surveys, 55(2), 1–96. https://doi.org/10.1145/3485128
16. Scher, S. (2018). Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. Geophysical Research Letters, 45(22), 12,616–12,623. https://doi.org/10.1029/2018GL080704
17. Smith, L. A., Suckling, E. B., Thompson, E., Maynard, T., & Du, H. (2019). Towards improving the framework for probabilistic forecast evaluation. Climatic Change, 132(1), 31–45.
18. Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., et al. (2007). Observations: Surface and atmospheric climate change. In Climate change 2007: The physical science basis (pp. 235–336). Cambridge University Press.
19. Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. Journal of Advances in Modeling Earth Systems, 12(9), e2020MS002109. https://doi.org/10.1029/2020MS002109