International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

Machine Learning Algorithms for Analysing Weather Patterns: A Case Study of Western Region of Kenya

Maureen Nechesa Murambi[1], Daniel Khaoya Muyobo[2], Richard Rono[3]
Department of Computer science
Kibabii university, Bungoma, Kenya[1]
Department of IT & Computer science
Kibabii University, Bungoma, Kenya[2,3]

Abstract — Traditional meteorological models often face challenges in processing large volumes of real-time data and capturing complex nonlinear atmospheric relationships. Recent advances in Machine Learning (ML) have provided powerful tools for analysing weather patterns and improving forecasting accuracy. The paper discusses relevant literature on machine learning algorithms suitable for weather pattern analysis, identifies research gaps and proposes future research directions involving deep learning and hybrid forecasting systems. This paper presents an integrated Internet of Things (IoT) and Machine Learning (ML) model for analysing weather patterns in Bungoma County, Kenya. Historical weather data (2006–2025) from the Nzoia Sugar Factory Weather Station and simulated real-time IoT sensor observations were analysed using Random Forest (RF) and K-Nearest Neighbours (KNN). Data preprocessing included outlier detection using the IQR method, polynomial interpolation for missing values, Min-Max normalization, and feature engineering. The model was trained and evaluated with an Infinite Random Search hyperparameter optimiser (578 configurations, 3-hour window). Performance was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).. The overall average R^2 across all predicted weather targets was 0.495, with relative humidity at 15:00 achieving $R^2 = 0.836$ and maximum temperature achieving $R^2 = 0.629$. Comparative evaluation showed that RF consistently outperformed KNN in predictive accuracy, demonstrating the suitability of ensemble learning for nonlinear meteorological datasets. The integration of IoT enabled continuous monitoring and improved decision support for agriculture and disaster preparedness. These findings contribute to the growing body of knowledge on ML applications in meteorology and provide a foundation for developing localized weather forecasting systems in regions with similar climatic conditions.

Keywords— Weather forecasting, Machine Learning, IoT, Weather Pattern Analysis.

I. INTRODUCTION

Technology is the interface through which humanity shapes its environment and its future. It amplifies our abilities, addresses our needs, and is a critical driver of societal evolution. Technologies like Machine Learning is increasingly being used in our day to day life. Machine learning is Artificial Intelligence that can automatically adapt with minimal human interference. Machine learning algorithms can be broadly classified into four major categories based on the manner in which they learn from data and make predictions. These categories include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Each category employs a distinct learning mechanism and is suitable for different types of data analysis and problem-solving applications.

Machine learning applications have demonstrated significant advances across diverse domains eg cancer prediction[1],water resource management[2], precision agriculture[3] etc.Weather forecasting has become increasingly important due to climate variability, extreme weather events, and the growing need for accurate environmental information. Agriculture-dependent regions such as Western Kenya are particularly vulnerable to unpredictable weather conditions that significantly affect food production and livelihoods[4]. Traditional numerical weather prediction models rely on complex mathematical equations and extensive computational resources, making localized forecasting challenging.

Machine learning offers an alternative approach by learning patterns directly from historical and real-time weather observations[5]. By identifying complex relationships among atmospheric variables, Machine Learning algorithms can generate accurate forecasts while requiring fewer computational resources than conventional meteorological models.

Background to the Study

Globally, machine learning has enabled significant advances in weather pattern analysis. In North America , K-means and Convolutional Neural Network (CNN) algorithms were used to analyse weather patterns by classifying winter and summer days [6]. In China, a comprehensive review of ML techniques in weather forecasting demonstrated the capability of Machine Learning to complement traditional numerical weather prediction systems [5]. In Africa, various Machine Learning approaches, including deep learning and Support Vector Machines, have been applied to predict precipitation and manage water resources[7, 8] .

In Rwanda, Long Short Term Memory (LSTM) networks have been used to predict extreme rainfall events [9], though inadequate data collection infrastructure remains a continental challenge. In Kenya, the western region relies heavily on rainfall for agriculture and domestic water supply, yet experiences erratic rainfall due to climate shifts, affecting food security and public health [10, 11]. From the synthesis of extant studies, there is a significant need to analyse relevant literature on machine learning algorithms suitable for weather pattern analysis, identifies research gaps, and proposes future research directions involving weather pattern analysis in Western Region of Kenya.

Objectives of the Study
General Objective
To examine machine learning algorithms used for analysing weather patterns and their effectiveness in improving localized weather forecasting.

Specific Objectives
- To review machine learning algorithms commonly applied in weather forecasting.
- To evaluate the performance of machine learning algorithms in weather pattern analysis.
- To identify research gaps and future directions in machine learning-based weather forecasting.

II. CLIMATE CHARACTERISTICS OF WESTERN REGION OF KENYA

Western Kenya, comprising the counties of Kakamega, Bungoma, Busia, and Vihiga, experiences an equatorial climate modified by altitude and proximity to Lake Victoria. The region receives bimodal rainfall, with the long rains occurring from March to May and the short rains from October to December [12]. Mean annual rainfall ranges from 1,200 mm in rain-shadow areas to over 2,000 mm on the slopes of Mount Elgon and near the lake.

Temperature patterns exhibit diurnal variation exceeding seasonal variation, with daily maxima typically ranging from 25°C to 30°C and minima from 12°C to 16°C. Relative humidity remains high throughout the year due to proximity to Lake Victoria, often exceeding 80% during morning hours before declining to 40–50% in the afternoon[13].

These climatic characteristics create specific requirements for weather analysis systems: the ability to capture sharp spatial gradients introduced by topography and lake effects, the capacity to handle high-frequency temporal variability in humidity and precipitation, and the robustness to operate with incomplete or imperfect input data.

III. LITERATURE REVIEW

A systematic review of the literature reveals that Machine Learning algorithms have been extensively applied to weather prediction and analysis tasks across diverse geographical and climatic contexts. The following review is organised geographically to identify patterns, achievements, and limitations relevant to the present study.

1.Weather Pattern Analysis in North America and Europe
In North America, [6] used K-means and CNN algorithms to analyse weather patterns by classifying days into winter and summer regimes. An ensemble-based Random Forest algorithm was developed to analyse atmospheric pressure patterns using data from 150 weather stations across North America, identifying 14 distinct pressure pattern signatures and achieving 89% accuracy in predicting pressure system movements [14]. However, the model showed limited feature importance exploration across different geographical regions and underperformed in coastal areas with complex terrain.

Deep learning models including GraphCast, Pangu-Weather, and FourCastNet were evaluated for their ability to forecast extreme weather events in Bangladesh, South Asia, and North America [15]. In Northern Europe, a Deep Neural Network (DNN) architecture was developed to process multi-dimensional weather data from meteorological stations, outperforming traditional statistical methods by 23%, achieving 87% accuracy in three-day precipitation forecasts, and identifying regional microclimate patterns previously undetected [15]. However, its high computational requirements restricted real-time implementation.

2.Weather Pattern Analysis in Asia
In Asia, six Global Climate Models (GCMs) were used to project spatial and temporal precipitation changes across Afghanistan, demonstrating ML's capacity to accurately analyse complex weather events [16]. However, IoT integration was absent in this study. Deep learning models including CNN and LSTM networks have been developed to simultaneously

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

predict multiple weather parameters with higher accuracy than traditional methods [17]; however, this research was survey-based and did not yield a deployable model.

3.Weather Pattern Analysis in Africa

In Africa, ML has been applied to predicting rainfall patterns [7], though integration with IoT remains limited. An ensemble of ML models was used to provide reliable information on heavy precipitation occurrence in the West African Sahel[8]. A hybrid CNN-LSTM architecture was developed to analyse weather patterns across West African countries using data from 35 weather stations combined with satellite imagery, achieving 83% accuracy in predicting seasonal rainfall patterns[8]; however, performance varied significantly across ecological zones.

In Nigeria, a hybrid ensemble model combining Gradient Boosting with Artificial Neural Networks demonstrated a 22% improvement over traditional forecasting methods, achieving 81% accuracy in predicting extreme rainfall events, though significant data gaps affected model training [18, 19]. In Southern Africa, a reinforcement learning framework achieved 79% accuracy in predicting flash flood events with 24-hour notice, though limited operational deployment was noted due to infrastructure constraints [20].

In Rwanda, LSTM networks outperformed CNN and GRU models in predicting extreme rainfall events, achieving accuracy metrics of 99.7%–99.8% [9]. In Tanzania's Wami River sub-catchment, five ML algorithms were compared for drought condition prediction, with LSTM achieving the highest R^2 and Nash-Sutcliffe Efficiency (NSE) of 0.99 [21].
In Kenya's central highlands, a crop advisory system using Random Forest achieved 98.8% accuracy in agricultural recommendations[22] . In Nyando, Western Kenya, LSTM, XGBoost, Random Forest, and Support Vector Regression (SVR) algorithms were compared for rainfall prediction, demonstrating ML's feasibility in the region but noting that IoT-based real-time data could substantially enhance prediction accuracy[23]. Recent research in Kenya has explored ML in climate modelling, but lacks real-time IoT integration [24, 25].

IV. RESEARCH GAPS

The systematic review of the literature reveals the following significant research gaps that informed the present study .Most studies focus on single machine learning models rather than integrated ensemble approaches[22]. Studies such as [6] analysed weather patterns by classifying them into winter and summer regimes - a categorisation inapplicable to Western Kenya's equatorial climate, which is characterised by two rainy seasons (March–May and October–December) rather than temperate seasons. Existing ML weather models in Kenya and the broader sub-Saharan Africa region were developed primarily using secondary historical data from weather stations or internet datasets, without integration of real-time IoT sensor data[22, 23].Many studies have employed either ML algorithms on historical data or IoT systems for data collection, but few have integrated both components into end-to-end systems capable of real-time data acquisition, processing, and prediction [5].

V. MODEL DEVELOPMENT

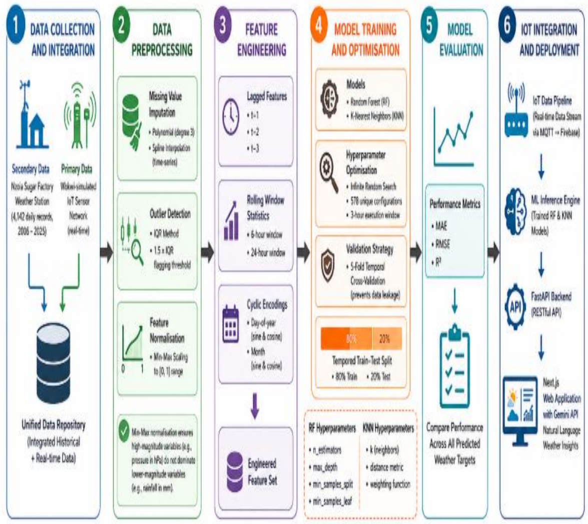

Fig 1: Model development Procedure

The study adopted a systematic machine learning model development process that began with the integration of secondary and primary weather data. Historical meteorological records collected from the Nzoia Sugar Factory Weather Station (4,142 daily records ,2006–2025) were combined with real-time observations generated through a Wokwi-simulated IoT sensor network to create a unified dataset. To ensure data quality, the dataset underwent extensive preprocessing, including missing value imputation using polynomial and spline interpolation, outlier detection through the 1.5× Interquartile Range (IQR) method, and Min–Max normalization to scale all variables uniformly and improve model performance.

Following preprocessing, feature engineering techniques were applied to capture the temporal characteristics of weather data. Lagged variables (t−1, t−2, and t−3), rolling window statistics (6-hour and 24-hour), and cyclic encodings representing the day of the year and month were generated to enhance the predictive capability of the machine learning models. These engineered features enabled the algorithms to learn seasonal and short-term weather patterns more effectively.

The Random Forest (RF) and K-Nearest Neighbours (KNN) models were subsequently trained using an 80:20 temporal train-test split. Hyperparameter optimization was performed through Infinite Random Search across 578 configurations, while five-fold temporal cross-validation minimized data leakage and improved model reliability. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). Finally, the optimized prediction model was integrated into an IoT-enabled system through a FastAPI backend and deployed as a Next.js web application capable of generating real-time weather predictions and natural language weather insights using the Gemini API.

VI. EVALUATION OF MACHINE LEARNING ALGORITHMS

A structured comparative evaluation of Random Forest (RF) and K-Nearest Neighbours (KNN) algorithms was conducted on the preprocessed dataset to determine the most appropriate ML approach for multi-target weather pattern prediction in Bungoma County.

Both algorithms were trained using 5-fold temporal cross-validation - a modification of standard k-fold cross-validation that respects the temporal ordering of the time series and prevents future data from contaminating the training of past observations, thereby avoiding the data leakage that would arise from random shuffling of a time-series dataset.

Hyperparameter optimisation was performed using Random Search, evaluating 578 unique configurations within a three-hour hardware execution window, a strategy selected over Grid Search for its computational efficiency in exploring high-dimensional hyperparameter spaces.

The RF ensemble model was trained by averaging predictions from multiple decision trees, with optimised hyperparameters including n_estimators = 150, max_depth = 12, and min_samples_split = 4. The ensemble averaging mechanism substantially reduces prediction variance and overfitting relative to a single decision tree, particularly for variables with irregular distributions such as rainfall.

The KNN model was configured with $k = 7$ nearest neighbours, a Minkowski distance metric with $p = 2$ (equivalent to Euclidean distance), identifying historically similar weather patterns in the feature space to generate predictions. Lagged features ($t-1$, $t-2$, $t-3$ time steps), rolling window averages (6-hour and 24-hour windows), and cyclic month-of-year encodings were incorporated as additional engineered features for both models to capture the temporal autocorrelation structure inherent in weather time series.

VII. DISCUSSION OF ALGORITHM EVALUATION FINDINGS

The Random Forest ensemble achieved superior predictive accuracy over KNN across all seven target variables, with an overall average R^2 of 0.495 compared to 0.371 for KNN - a 33.4% improvement in overall explained variance. These findings are consistent with [25] ,who demonstrated that RF consistently delivers competitive performance for multi-target weather prediction in the Kenyan context, and with the broader ensemble learning literature as theorised by [26], who established that ensemble averaging substantially reduces variance without increasing bias.

Limitations of the Study

The accuracy of the machine learning models could have been affected by emerging climate change factors that were not represented in the 20-year historical dataset used for training. To mitigate this limitation, rigorous feature engineering techniques were applied, and model parameters were optimised using Infinite Random Search across 578 configurations.

The historical weather dataset obtained from the Nzoia Sugar Factory Weather Station contained missing values in several variables. This challenge was addressed through the application of polynomial and spline interpolation techniques.

Recommendations for Further Study

Future studies should evaluate advanced deep learning models such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Transformer-based architectures for weather prediction in Western Kenya and compare their performance with Random Forest. Future research should also incorporate atmospheric pressure, wind direction, satellite imagery, vegetation indices, and topographical data to improve prediction accuracy, particularly for rainfall forecasting and extreme weather events.

REFERENCES

1. B. Zhang, H. Shi, and H. Wang, "Machine learning and AI in cancer prognosis, prediction, and treatment selection: A critical approach," Journal of Multidisciplinary Healthcare, vol. 16, pp. 1779–1791, 2023.
2. F. Ghobadi and D. Kang, "Application of machine learning in water resources management: A systematic literature review," Water, vol. 15, no. 4, Art. no. 620, 2023.
3. A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," IEEE Access, vol. 9, pp. 4843–4873, 2021.
4. B. K. Kogo, L. Kumar, R. Koech, and K. Hasan, "Climatic and non-climatic risks in rainfed crop production systems: Insights from maize farmers of western Kenya," Climate and Development, vol. 13, no. 10, pp. 869–878, 2021.
5. H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine learning methods for weather forecasting: A survey," Atmosphere, vol. 16, no. 1, Art. no. 82, 2025.
6. A. Chattopadhyay, P. Hassanzadeh, and S. Pasha, "Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data," Scientific Reports, vol. 10, no. 1, Art. no. 1317, 2020.
7. A. Oyerinde et al., "Rainfall prediction using machine learning techniques in Africa," Applied Water Science, vol. 9, no. 2, 2019.
8. I. A. Saley and S. Salack, "Present and future of heavy rain events in the Sahel and West Africa," Atmosphere, vol. 14, no. 6, Art. no. 965, 2023.
9. J. Kagabo, G. R. K. ..., J. Kazora, C. N. Shangwe, and F. Habiyakare, "Application of machine learning algorithms in predicting extreme rainfall events in Rwanda," Atmosphere, vol. 15, no. 6, 2024.
10. IGAD, Climate Change and Human Health in the IGAD Region. Djibouti: Intergovernmental Authority on Development (IGAD), 2019.
11. Water Resources Authority (WRA), National Water Master Plan. Nairobi, Kenya: Government of Kenya, 2021.
12. P. I. Palmer, C. M. Wainwright, B. Dong, R. I. Maidment, K. G. Wheeler, N. Gedney, J. E. Hickman, N. Madani, S. S. Folwell, and G. Abdo, "Drivers and impacts of Eastern African rainfall variability," Nature Reviews Earth & Environment, vol. 4, no. 4, pp. 254–270, 2023.
13. S. E. Nicholson, "Climate and climatic variability of rainfall over eastern Africa," Reviews of Geophysics, vol. 55, no. 3, pp. 590–635, 2017.
14. L. Lang, T. Li, S. An, and X. Tang, "An improved random forest algorithm and its application to wind pressure prediction," International Journal of Intelligent Systems, vol. 36, no. 8, pp. 4016–4032, 2021.
15. O. C. Pasche, J. Wider, Z. Zhang, J. Zscheischler, and S. Engelke, "Validating deep learning weather forecast models on recent high-impact extreme events," Artificial Intelligence for the Earth Systems, vol. 4, no. 1, Art. no. e240033, 2025.
16. P. Ebiendele, K. Dairaku, and P. Adigun, "Deep learning downscaling of West Africa summer monsoon precipitation flow: Exploring added value," Climate Dynamics, vol. 63, Art. no. 410, 2025.
17. K. Xie, L. Huang, W. Zhang, Q. Qin, and L. Lyu, "A CNN-based multi-task framework for weather recognition with multi-scale weather cues," Expert Systems with Applications, vol. 198, Art. no. 116689, 2022.
18. S.-Q. Dotse, "Deep learning-based long short-term memory recurrent neural networks for monthly rainfall forecasting in Ghana, West Africa," Theoretical and Applied Climatology, vol. 155, no. 4, 2024.
19. O. S. Ojo and S. T. Ogunjo, "Machine learning models for prediction of rainfall over Nigeria," Scientific African, vol. 16, Art. no. e01246, 2022.
20. N. Byaruhanga, D. Kibirige, S. Gokool, and G. Mkhonta, "Evolution of flood prediction and forecasting models for flood early warning systems: A scoping review," Water, vol. 16, no. 13, Art. no. 1763, 2024.
21. C. Lalika, A. U. H. Mujahid, M. James, and M. C. Lalika, "Machine learning algorithms for the prediction of drought conditions in the Wami River sub-catchment, Tanzania," Journal of Hydrology: Regional Studies, vol. 53, Art. no. 101794, 2024.
22. P. A. Lokolonyei, "Data-driven intelligent crop advisory system using machine learning approaches for Kenya's Central Highlands," Ph.D. dissertation, 2024.
23. A. Lawal, S. Y. Yerima, D. O. Olago, P. O. Amingo, C. W. Kariuki, W. Wang'ombe, L. Olaka, L. Obiero, and S. O. Wandiga, "Evaluating machine learning models for rainfall prediction: A case study of Nyando in Kenya," in Proc. IEEE, 2023, pp. 264–271.
24. D. Okoth, "AI forecasting equips East Africa for extreme weather," SciDev.Net, 2024.
25. S. Eddamiri, F. Z. Bassine, O. Hakam, P. L. K. Kambiet, and A. Chehbouni, "Applications of artificial intelligence for forecasting and managing extremes," in Climate Change and Rainfall Extremes in Africa. Amsterdam, The Netherlands: Elsevier, 2025, pp. 91–113.
26. J. Wang and G. Li, "Multidimensional meteorological data analysis based on machine learning," International Journal of Computer Applications in Technology, vol. 71, no. 3, pp. 244–250, 2023.