JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.8 No.1 January 2026: 37-47
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v8i1.14302

Classification of Severe Weather Conditions in Nigeria: An Integrated Weather Database with Machine Learning Approach

B.O Olasunkanmi [1]*, J. D Adekunle [2], S. O Oyelakin [3], A. M Obaude [4], C.O. Afolabi [5], M. I Oyeniran [6], G. E Ideh [7], E. J Ayanlowo [8], C. K Ogu [9], C. O Robert [10], H. S Sule [11], T Anifowoshe[12]

[1,4]Department of Computer Science,
Achievers University, Owo,
Ondo state, Nigeria 341104
Blessingolawale.bo@gmail.com; obaude-pg0011@achievers.edu.ng

[2,6,7]Department of Mathematics, College of Physical Science,
[5]Department of Microbiology, College of Biological Science,
[11]Department of Statistics,
Federal University of Agriculture, Abeokuta,
Ogun State, Nigeria 110111
adekunlejd.17@student.funaab.edu.ng; oyeniranmi.17@student.funaab.edu.ng; idehge.19@student.funaab.edu.ng;
obaude-pg0011@achievers.edu.ng;
harunasulesani@gmail.com

[3]Department of Mass Communication,
Bayero University, Kano,
Kano State, Nigeria 700006
soyelakin@punchng.com

[8]Department of Geography, Environment and Population,
University of Adelaide,
Adelaide SA 5005, Australia
eniola.ayanlowo@adelaide.edu.au

[9]Medipolis GmbH, Otto-Schott-StraSSe,
Jena, Germany 07745
kizzi1428@outlook.com

[10]Department of Management Information Systems, TOPDEL Engineering Limited,
[12]Fisher of Men Technology Academy,
Lagos, Nigeria 106104
rokechukwu123@gmail.com; anifowoshejoseph3@gmail.com

* Correspondence: Blessingolawale.bo@gmail.com

Abstract – *Severe weather events pose significant risks to human safety, infrastructure, and economic activities, particularly in developing regions such as Nigeria, where reliable weather data management and analytical systems remain limited. This study presents an integrated weather data management database and a machine learning–based framework for classifying severe weather conditions using meteorological data from Nigeria. Secondary weather data was obtained from the OpenWeather platform covering the period from February 21st to 27th, 2024. A structured database was designed to store and manage the weather variables, followed by data preprocessing and exploratory statistical analysis. Supervised machine learning models were trained to classify weather conditions into severity categories based on predefined thresholds. Model performance was evaluated using training and testing datasets. Among the evaluated models, the random forest and neural network achieved the highest classification accuracy, while logistic regression showed comparatively lower but stable performance. Although high accuracy values were observed, these results may be influenced by rule-based severity labeling and potential class imbalance. This study demonstrates the feasibility of integrating weather data management systems with* *machine learning techniques for automated severe weather classification in Nigeria. Future research should incorporate expert-validated severity labels, longer temporal datasets, and external validation to improve generalizability and reduce overfitting risks.*

Keywords: *Weather Severity; Machine Learning; Classification problem*

I. INTRODUCTION

Data, also referred to as information, has significantly improved the understanding of complex phenomena that were previously difficult to analyze. Favaretto et al. (2020) note that data are generated from diverse sources, including humans and machines. Waring (2021) further categorizes data into primary and secondary sources. The increasing availability and use of data have become essential for monitoring global patterns and supporting informed decision-making across multiple sectors (Jain et al., 2020). The importance of data in modern society is emphasized in multiple studies. Lei and Ming (2023) and Osakwe et al. (2023) stress that data enables organizations to gain insights, enhance performance, and support strategic planning.

Data plays a central role in weather monitoring and analysis. Weather data are collected continuously from multiple sources, including satellites, weather stations, and remote sensors, generating large and complex datasets. Effective management of these datasets is essential for understanding atmospheric behavior and supporting timely decision-making. Consequently, the construction of resilient databases for weather data management is vital for meteorological research, disaster preparedness, and societal protection (Muppala, 2025). Weather data management supports not only meteorology but also sectors such as agriculture, transportation, energy, and urban planning. For instance, farmers depend on accurate weather information to optimize planting schedules, manage irrigation, and reduce crop losses caused by adverse conditions (Malhi et al., 2021).

Despite the availability of weather data, severe weather events continue to pose substantial threats to human life and economic activities. In Nigeria, where the economy is heavily dependent on agriculture, weather conditions have a significant influence on productivity and livelihoods (Ogbuabor & Egwuckwu, 2017). Extreme events such as floods and droughts frequently damage infrastructure, disrupt business activities, and reduce agricultural and forestry outputs. Climate change further intensifies these impacts by increasing erosion and wind-related damage, leading to declines in forest products such as timber and cane, reduced income, and rising construction material costs.

Managing and analyzing the growing volume of weather data requires robust data management systems. Relational databases such as MySQL and PostgreSQL, as well as NoSQL databases including MongoDB and Cassandra, are commonly used to store diverse weather data types, including numerical measurements, satellite imagery, and textual records (Khan et al., 2023). These systems integrate data from multiple sources into unified platforms that support efficient retrieval and analysis. Additionally, large-scale weather datasets are often stored in data warehouses or data lakes, while big data frameworks such as Apache Hadoop and Spark enable distributed processing and advanced analytics (Nambiar & Mundra, 2022).

Beyond data storage, extracting actionable insights from weather data relies on machine learning techniques. Integrating machine learning with database systems enhances data-driven decision-making, improves efficiency, and supports more accurate classification and prediction tasks (Chinta, 2025). Previous studies have demonstrated the importance of weather conditions across sectors such as agriculture, aviation, health, air quality, and transportation (Lee et al., 2014). Machine learning algorithms analyze historical weather data to identify patterns and correlations in variables such as temperature, precipitation, wind speed, and atmospheric pressure that may not be readily observable through traditional methods (Chen et al., 2023). Techniques including regression, classification, clustering, and neural networks are widely applied for short- and long-term weather analysis and forecasting (Adekunle et al., 2024; Chen et al., 2023).

However, despite the increasing application of machine learning in weather studies, there remains a gap in the development of integrated systems that combine structured weather data management with machine learning–based classification of severe weather events, particularly in the Nigerian context. Extant studies focus either on forecasting accuracy or on data storage solutions, without providing an end-to-end framework that supports data collection, management, and automated classification of severe weather conditions. Addressing this gap is significant for improving weather-related decision-making in agriculture, infrastructure planning, and disaster risk reduction in Nigeria. An integrated database and machine learning framework can support the timely identification of severe weather conditions, enhance preparedness, and reduce socio-economic losses.

Therefore, this study aims to design and develop a weather data management system and apply machine learning techniques for the classification of severe weather events by designing a weather management system using MySQL capable of handling diverse weather data and developing a binary classification model to distinguish between severe and non-severe weather conditions.

II. METHODS

To achieve the aim of this study, the methodology was structured into two main phases. The first phase involved the design and implementation of a relational database. Database functionality and integrity were verified through structured query operations that retrieved and joined data across related tables. Thereafter, a supervised machine learning model was trained to classify severe weather pattern conditions. The dataset used for this study was sourced from OpenWeather, a valuable extension to the existing Nigerian Meteorological Agency (NiMet) dataset. It included comprehensive weather information for cities across Nigeria, with a total of 22,541 observations. The dataset has been carefully structured in a CSV format with 21 attributes, providing detailed meteorological and geographical data. Attributes include: (1) country, (2) city, (3) latitude, (4) longitude, (5) temperature, (6) minimum temperature, (7) maximum temperature, (8) pressure measured in hectopascals (hPa), (9) humidity percentage, (10) sea level measured in hPa, (11) ground level measured in hPa, (12) wind speed, (13) wind degree, (14) sunrise, (15) sunset, 16) time zone, (17) cloud, (18) description of the weather condition, (19) region, (20) population and, (21) date of measurement.

2.2. Database Designs And Implementation

The database was designed to ensure minimal redundancy, scalability, data independence, and integrity. These objectives were achieved through normalization. The initial database structure was drafted in Microsoft Excel before implementation, as shown in Figure 1.

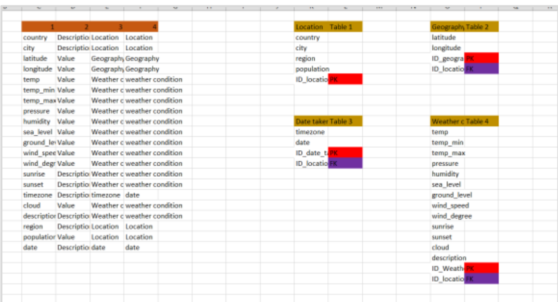


Figure 1. Ms Excel segment of the data normalization

Four tables were created based on independence and dependence by identifying entities and their corresponding descriptive and quantitative attributes. The tangible entities in the data are country, city, region, and population. The intangible entity was also identified to be date and description. For each table, there are primary keys assigned to each of the entities serving as an ID number, which makes it different from other entities. Location table with ID_location, weather condition table with ID_Weather_condition, geography table with ID_geography, and date taken with ID_date_taken. All these are uniquely defined, with no duplicates. The three tables (weather condition, geography, and data taken) all depend on the location table. This made it necessary to create a relationship between the independent and the dependent. And the relationship was created such that there is one-to-one correspondence between entities of the dependent tables to that of the independent. The creation of the database was done using a structure query language Mysql which is sufficient for creating a relational database. This process involves the implementation of the design in Figure 2[(a) - (d)].

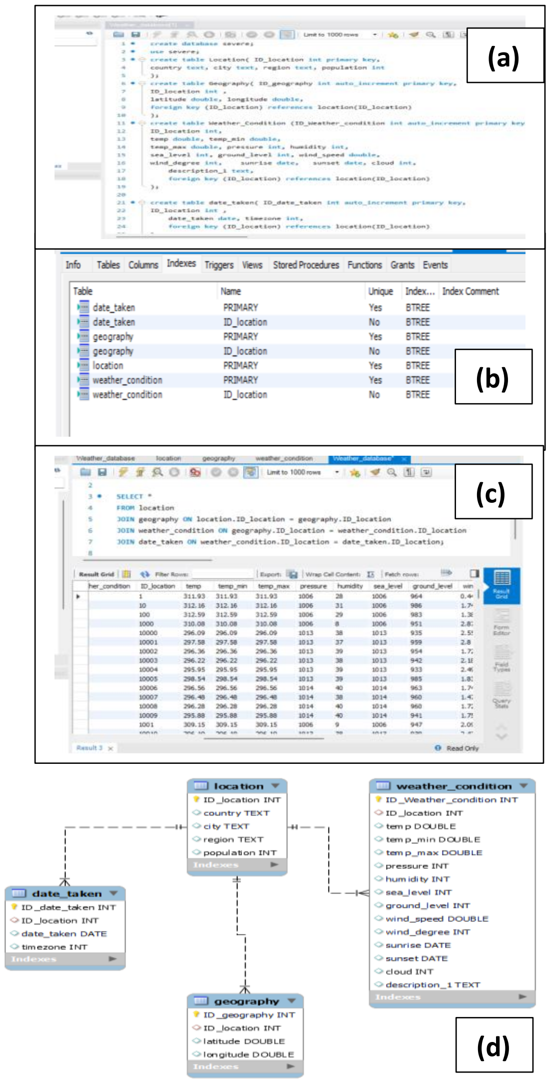


Figure 2: (a) Code snippet used to create the database; (b) Display of the tables and the primary keys; (c) Display of a pulling query of all the tables on the database;(d) Entity-relationship diagrams

2.3. Supervised Classification Algorithm

Upon successfully designing and implementing the database, the supervised machine learning (ML) was trained on it for classification (Figure 3).

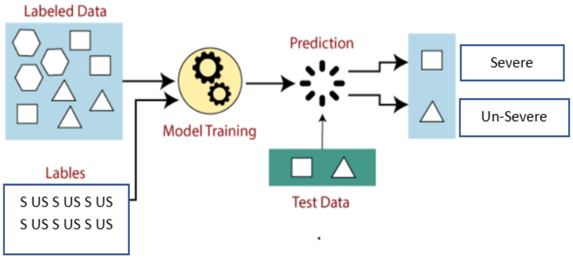


Figure 3. Supervised machine learning structure

2.4. Model Selection

A binary classifier involving three models (i.e., logistic regression, a random forest, and a binary neural network) was trained for comparison purposes using the data from the database to classify severe weather data and non-severe weather data. These models were selected based on their strength to handle variable of two classes, such as event occurring and those not occurring. The two classes are mutually exclusive and mutually exhaustive. In this case, severe and non-severe. The logistic regression uses maximum likelihood estimate to find the best least and unbiased estimate (i.e., coefficients) that minimizes the log-odd (equation 2) of predicting weather condition to be true.

$$Odds = \frac{p(Y = severe)}{P(Y = un - severe)} = \frac{P(Y=severe)}{1-p(Y=severe)} \quad (1)$$

$$\log Odds = \frac{P(Y = severe)}{1 - p(Y = severe)} = \log p(Y = severe) - \log(1 - p(Y = severe) \quad (2)$$

The logistic function, also referred to as the sigmoid function (equation 3), is employed to convert the result of the linear equation into a probability value ranging from 0 to 1.

$$S(X) = \frac{1}{1 + e^{-x}} \quad (3)$$

Logistic regression models acquire knowledge of a decision boundary that divides the two classes in the feature space. In logistic regression, the decision border is commonly depicted as a linear barrier defined as the collection of points where the probability is precisely 0.5. Any probability less than the threshold is classified as 0, and any probability greater than the threshold is classified as 1, as seen in Figure 4.

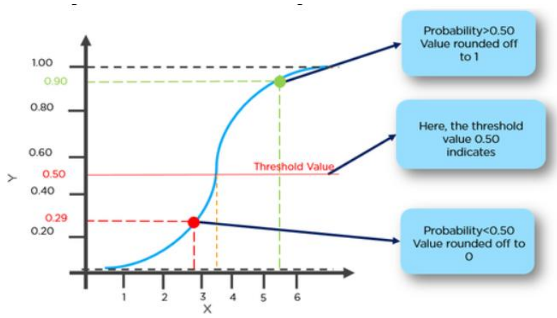


Figure 4. Decision boundary of a logistic regression

The mathematical representation of the model is the linear combination of all the independent variables with some weights (i.e., coefficients) contribution to the prediction of the probability of predicting the weather condition to be severe, which is given below:

$$p(Y = severe \mid X_i) = \frac{e^{k_0 + k_1 X_1 + \cdots + k_i X_i}}{1 + e^{k_0 + k_1 X_1 + \cdots + k_i X_i}} \quad (4)$$

2.5. Data Modification: Determining Weather Severity Classification

The collected data has no severe label. To assign this label, a crucial step involved classifying weather severity to facilitate predictive modeling. The classification was based on specific thresholds for various weather parameters, creating a binary categorization of ‘severe’ and ‘non-severe’ weather conditions. Due to the inexistence of a universally accepted event-level severity labels for Nigerian city-scale weather observations, severity in this study was operationalized using a rule-based proxy classification informed by WMO impact-based warning principles (WMO, 2015; WMO, 2018) and aligned with severe weather indicators commonly referenced in operational advisories issued by the Nigerian Meteorological Agency (NiMet, 2023a; NiMet, 2023b; NiMet, 2023c) and the U.S. National Oceanic and Atmospheric Administration and National Weather Service (NOAA, 2022; NWS, 2021).

Temperature conditions were flagged as potentially severe when values exceeded 313.15 K (40 °C), reflecting extreme heat. Similarly, observations with atmospheric pressure below 1000 hPa were treated as indicative of potentially unstable weather conditions. Extremely high or low humidity levels were also considered, with values above 90% or below 20% used as markers of abnormal atmospheric moisture conditions. Wind conditions were classified as potentially severe when wind speed exceeded 10 m/s, as high winds are often associated with disruptive weather events. In addition, textual weather descriptions containing terms such as storm, thunderstorm, or heavy rain were used as qualitative indicators of severity. Where such descriptions coincided with cloud cover greater than 80%, the observation was likewise labeled as severe. The figure [5(a) – 5(b)] below shows the geographical representation of the severed and not severed state in Nigeria.

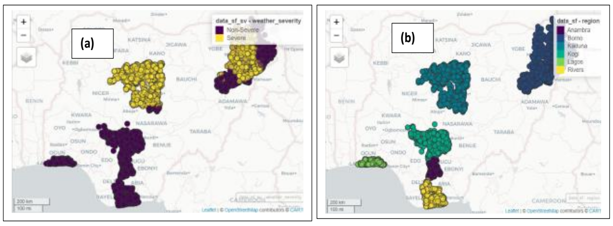


Figure 5. (a) Geographical representation of Nigerian states with severe weather conditions; (b) Geographical representation of Nigeria state identified

2.6. Data Preparation and Preprocessing

The dataset was prepared to address any form of pattern that could affect the accuracy of the result. Data normalization was done using equation 5 below. Thereafter, the dataset was partitioned into 80% for fitting and 20% for validation. Because weather severity labels were derived from the same meteorological variables used as model inputs, the learning task in this study primarily evaluates the model's ability to reproduce rule-consistent classifications rather than to infer independent causal severity signals.

$$X' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

The imbalancing issue in the fitting data was rectified using the mixed method which involves the combination of oversampling and undersampling to equalize the distribution of the target variable, as seen in the figure[6(a) – 6(b)] below. All resampling procedures to address class imbalance were applied exclusively to the training subset, while the test data were retained in their original distribution and were not exposed to oversampling.

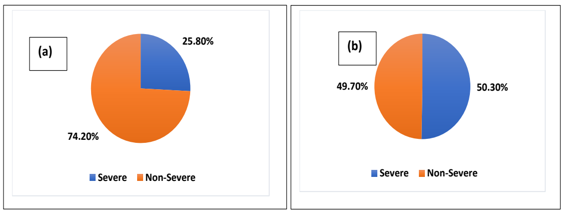


Figure 6. (a) Unbalanced state of the dataset; (b) Balanced state of the dataset

2.8. Model Evaluation Process

Model evaluation was conducted using a single hold-out test set to assess internal classification consistency by using 20% of the entire dataset to test the performance of the model. The evaluation involves the assessment of various metrics, including accuracy, precision, recall, and F1-score (Equation 6; Adekunle et al., 2024).

Equation 6 Performance metrics evaluation equation

$$\text{Accuracy} = \frac{TP_{total}}{Total\ number\ of\ instances}$$

$$\text{Sensitivity} \frac{TP}{Total\ number\ of\ instances}$$

$$\text{Specificity} = \frac{TN}{Total\ number\ of\ instances}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{TN}{TN + FN}$$

$$Prevalence = \frac{\sum y}{N}$$

$$\text{Balanced accuracy} = \frac{Sensitivity + Specificity}{2}$$

True Positives (TP): This represents the number of weather condition correctly identified as severe by the model.
True Negatives (TN): This is the number of weather condition correctly identified as not severe.
False Positives (FP): This represents the number of weather condition incorrectly identified as severe when it is not severe

.

2.7. Tools And Material

The implementation of this study was done in R programming software version 2024.12.0+467 in an R studio integrated development environment on a Windows operating system. Several packages were used for a successful implementation. Packages such as tidyverse were used for data manipulation, wrangling, and visualization while dlookr was used to explore and visualize patterns of missingness through Pareto plots. Class imbalance in the weather severity outcome was addressed using the ROSE package caTools package was employed to split the data into training and testing subsets. To handle spatial data, the sf package was utilized for working with geographic coordinates, and interactive map visualizations were generated using mapview, enabling the mapping of weather severity patterns across locations in Nigeria. For predictive modeling, logistic regression was implemented using base R's glm function, random forest model with the randomForest package, and neural network models were trained using the nnet package. Model evaluation was carried out using the caret package. All packages used are open-source and widely adopted in the scientific community for environmental and spatial data analysis, ensuring reproducibility and transparency of the results.

III. RESULTS AND DISCUSSION

The descriptive analysis of various weather condition variables, highlighting their central tendencies, dispersion, and variability (Table 2). The key metrics analyzed include the mean, standard deviation (sd), minimum, median, maximum, interquartile range (IQR), and coefficient of variation (CV). The average population is 53,158.7 with a high standard deviation of 152,217.3, indicating significant variability across different regions, and the coefficient of variation is 2.9, reflecting high relative variability in population distribution. Temperature has an average value of 303.3 K with a standard deviation of 4 K, suggesting consistent temperatures across regions. The pressure has an average of 1010.2 hPa with a standard deviation of 2.3 hPa, showing minimal variation. On the other hand, humidity presents a mean of 38% with a standard deviation of 22%, indicating moderate variability. Ground level analysis shows an average of 976.2 meters with a standard deviation of 27.6 meters, indicating relatively stable elevation levels.Table 1. Student Distribution Frequency

Table 2. *Descriptive analysis of the weather condition variables*

Variable	Mean	Std	Min	med	Max	IQR	CV
Population	53158.7	152217.3	0	10169	1777118	20981	2.9
Temp	303.3	4	291.1	303.4	314.1	5	0
Pressure	1010.2	2.3	1003	1011	1015	3	0
Humidity	38	22	7	34	85	39	0.6
Ground level	976.2	27.6	873	978	1012	51	0
Wind speed	2.9	1.5	0	2.6	8.3	2.2	0.5
Wind degree	134.3	103	0	95	360	177	0.8
Cloud	71.7	35.9	0	93	100	53	0.5

The wind speed has an average value of 2.9 m/s with a standard deviation of 1.5 m/s, indicating moderate variability. Also, wind degree has an average value of 134.3° with a standard deviation of 103°, indicating high variability in wind direction. Lastly, cloud cover presents a mean of 71.7% with a standard deviation of 35.9%, showing considerable variability. The coefficient of variation for the variables was very low. Temperature (cv = 0), pressure (cv = 0), humidity (cv = 0.6), ground level (cv = 0), wind speed (cv = 0.5), wind degree (cv = 0.8) and cloud (cv = 0.5) which highlight a very low or moderate relative variability. The weather conditions are shown in Figure 7(a – h). Similarly, the descriptive statistics for weather conditions by state and by descriptions were also provided (Supplementary 1 & Supplementary 2).

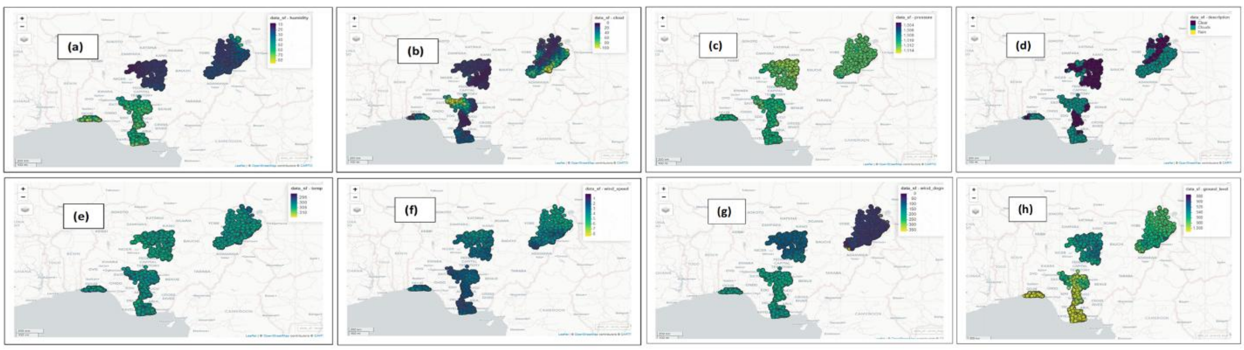


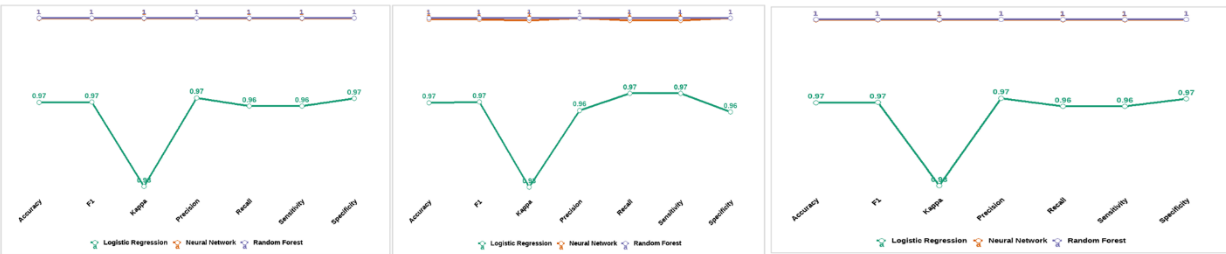
Figure 7. Weather conditions by identified states: (a) humidity; (b) cloud; (c) pressure; (d) description; (e) Temperature; (f) wind speed; (g) wind degree; (h) ground level

The relationship between the weather variables is shown in Figure 8. Population showed a weak positive correlation with temperature (0.028) and ground level (0.232), indicating slight associations where higher populations might be found in slightly warmer and elevated areas. Additionally, a moderate positive correlation was observed between population and humidity (0.249), suggesting that more populated regions tend to be more humid. Conversely, population had a weak negative correlation with pressure (-0.141) and wind speed (-0.024), implying that areas with higher populations might experience slightly lower pressure and wind speeds. Temperature exhibited a strong negative correlation with pressure (-0.640), highlighting that higher temperatures are typically associated with lower pressure. Furthermore, there was a moderate negative correlation between temperature and both humidity (-0.296) and wind degree (-0.002), suggesting that higher temperatures might correspond to lower humidity and slight variations in wind direction. Temperature also had a weak positive correlation with ground level (0.260) and wind speed (0.221), indicating that temperature might slightly increase with elevation and wind speed. Also, pressure was found to have a weak negative correlation with humidity (-0.083), ground level (-0.387), and wind speed (-0.015), suggesting that lower pressure is slightly associated with higher humidity and lower elevations.

	population	temp	pressure	humidity	ground_level	wind_speed	wind_degree	cloud
population	1.00	0.03	-0.14	0.25	0.23	-0.02	0.19	0.09
temp	0.03	1.00	-0.64	-0.30	0.26	0.22	0.00	-0.08
pressure	-0.14	-0.64	1.00	-0.08	-0.39	-0.01	-0.32	-0.08
humidity	0.25	-0.30	-0.08	1.00	0.66	-0.30	0.64	0.12
ground_level	0.23	0.26	-0.39	0.66	1.00	-0.11	0.53	0.13
wind_speed	-0.02	0.22	-0.01	-0.30	-0.11	1.00	-0.39	-0.01
wind_degree	0.19	0.00	-0.32	0.64	0.53	-0.39	1.00	0.01
cloud	0.09	-0.08	-0.08	0.12	0.13	-0.01	0.01	1.00

Figure 8. Relationships between weather condition variables

Additionally, pressure's correlation with wind degree (-0.322) and cloud cover (-0.075) was weakly negative, indicating minimal relationships with these variables. On the other hand, humidity showed a strong positive correlation with ground level (0.662) and wind degree (0.642), indicating that more humid areas tend to be at higher elevations and associated with specific wind directions. There was also a weak negative correlation between humidity and temperature (-0.296) and wind speed (-0.297), suggesting that higher humidity is generally found in cooler areas with slower winds. The ground level had a moderate positive correlation with wind degree (0.528) and a weak positive correlation with cloud cover (0.129), suggesting that higher elevations might experience certain wind directions and slightly more cloud cover. A weak negative correlation was observed between ground level and wind speed (-0.112), indicating that higher elevations might experience slower wind speeds. Wind speed showed weak negative correlations with population (-0.024), pressure (-0.015), humidity (-0.297), and ground level (-0.112), indicating minimal relationships with these parameters. The correlation with wind degree was moderate and negative (-0.388), suggesting that higher wind speeds are associated with specific wind directions. Wind degree had a moderate positive correlation with humidity (0.642) and ground level (0.528), indicating that specific wind directions are associated with higher humidity and elevations. The correlations with population (0.186), temperature (-0.002), and cloud cover (0.008) were weak, showing minimal relationships. Finally, Cloud cover exhibited weak correlations with most variables, with the strongest being a weak positive correlation with humidity (0.123) and ground level (0.129), suggesting slight associations.



(a) (b) (c)

Figure 9. Performance Comparison of Models on (a) Training dataset (b) Testing dataset (c) the complete dataset

The performance metrics of the three-models — Neural Network (NN), Random Forest (RF), and Logistic Regression (LR) revealing significant insights into their effectiveness across the training (Figure 9(a)), testing (Figure 9(b)), and full datasets (Figure 9(c)) based on key metrics including accuracy, sensitivity, specificity, precision, Kappa, and balanced accuracy [Figure 9(a) – 9(b)]. On the training dataset (Figure 9(a)), both NN and RF models showed perfect classification ability, with NN achieving an accuracy of 1(95% CI: 0.9997 to 1), Kappa of 0.9999, sensitivity of 0.9999, and specificity of 1.0. RF slightly outperformed NN with a perfect accuracy of 1.0(95% CI: 0.9998 to 1), Kappa of 1, 0.9635 sensitivity, and 100% specificity, without any misclassification. In contrast, the LR model had an accuracy of 0.9651(95% CI: 0.9625 to 0.9676), Kappa of 0.9303, sensitivity of 0.9635, and specificity of 0.9668(Supplementary

3(a)). While these values are still strong, LR misclassified 707 samples in training, highlighting a trade-off between interpretability and predictive precision (Supplementary 4(a)).

When tested on the unseen testing dataset, a similar pattern emerged (Figure 9(b)). RF maintained its perfect accuracy of 1.0(95% CI: 0.9984 to 1), while NN had a slightly lower accuracy of 0.9996 (95% CI: 0.9975 to 1), sensitivity of 0.9991, and Kappa of 0.9991, with only 1 misclassified instance. LR again trailed, achieving accuracy of 0.9663(95% CI: 0.958 to 0.973), sensitivity of 0.9700, specificity of 0.9625, and Kappa of 0.9326, misclassifying a total of 76 instances (42 false negatives and 34 false positives) (Supplementary 4(b)). Despite this, LR's consistent performance confirms its generalizability. Notably, the balanced accuracy for LR on test data was 0.9663, compared to 0.9996 for NN and 1.0 for RF, reinforcing the relative strength of the complex models in capturing patterns in unseen data (Supplementary 3(b)).

Across the entire dataset, RF continued its flawless classification, achieving 1.0(95% CI: 0.9998 to 1) accuracy, perfect Kappa (i.e., 1), and no misclassifications (Supplementary 4(c)). The NN model performed nearly as well, with an accuracy of 0.9999(95% CI: 0.9997 to 1), Kappa of 0.9998, sensitivity of 0.9998, and specificity of 1.0, making only 2 classification errors (Supplementary 4(c)). LR also maintained consistent but lower scores, recording accuracy of 0.9652(95% CI: 0.9628 to 0.9676), sensitivity of 0.9642, specificity of 0.9664, and Kappa of 0.9305, with a total of 783 misclassified instances (377 false negatives and 406 false positives) (Supplementary 4(c)). While RF and NN exhibited extremely high performance across all datasets, the lack of misclassification on both training and test sets may suggest overfitting, which calls for careful evaluation using external validation or cross-validation techniques. In contrast, LR's relatively lower yet stable performance across datasets suggests better potential for generalization and robustness, especially in real-world scenarios where perfect predictions are rare. The predicted weather conditions are visualized on the map below [Figure(a) to Figure(b)] while the model equation is expressed in equation 7 below.

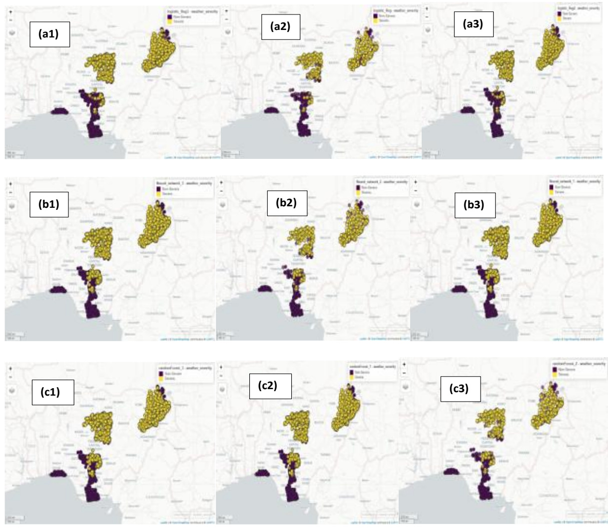


Figure 10. Prediction by (a) logistic on (a1) All the dataset(a) Training dataset (a3) Testing dataset; (b) neural network on (b1) All the dataset(b2) Testing dataset(b3) Training dataset; (c) Random Forest on (c1) All the dataset[(c2) Training dataset (c3) Testing dataset.

$$\mathrm{p(Y = severe \mid X_i)} = \frac{e^{\mathrm{equation(8)\,-\,equation(9)}}}{1 + e^{\mathrm{equation(8)-equation(9)}}} \tag{7}$$

$$4.4546 \times \text{Groun level} - 5.3948 \times \text{wind speed} \\ -6.8325 \times \text{cloud} + 14.9117 \times \\ \text{latitude} - 3.2053\ \text{longitude} \tag{8}$$

$$16.8768 - 6.8325 \times \text{Population} + 14.9117 \\ \times \text{Temperature} - 9.2941 \times \text{Pressure} - 99.9828 \\ \times \text{Humidity} \tag{9}$$

3.2. Discussion

The result of this study shows a comprehensive analysis of the weather conditions in Nigeria, presenting the complex relationships between various weather variables and their potential impacts on different regions in Nigeria, highlighting the critical influence of weather conditions on socio-economic factors. The high variability in population distribution indicates diverse population densities across regions, which may influence local climate conditions due to human activities. Temperature and pressure, showing consistent values across regions, suggest stable climatic conditions, whereas moderate variability in humidity points to differences in moisture levels, which can significantly impact agriculture and forestry. The observed relationships between weather variables reveal intricate dynamics. The weak positive correlation between population and temperature implies that more densely populated areas might experience slightly higher temperatures due to urban heat effects (Jiang et al., 2018). Conversely, the strong negative correlation between temperature and pressure indicates that higher temperatures are typically associated with lower pressure, a common meteorological phenomenon (WHO, 2018). The findings underscore the significant impact of weather conditions on agriculture and forestry in Nigeria. High variability in humidity, coupled with the strong positive correlation between humidity and ground level, suggests that elevated areas tend to be more humid, which could benefit certain crops but may also pose challenges due to increased erosion and wind damage (Malhi et al., 2021). The moderate negative correlation between temperature and humidity implies that higher temperatures could exacerbate drought conditions, adversely affecting crop yields and forest health (Cui et al., 2022). The performance metrics of the supervised binary classifier model demonstrate its efficacy in predicting severe weather conditions. The high classification accuracy observed reflects strong internal consistency between the rule-based severity definition and the supervised learning models, rather than independent event-level prediction performance. Sensitivity and specificity metrics show the model's ability to accurately identify both severe and non-severe weather conditions, crucial for practical applications in weather prediction and disaster management. Comparing the model's performance on novel datasets with established methods reveals its potential for real-world applications. The high predictive accuracy on both test and train datasets indicates the model's generalizability and reliability, essential for deploying in operational weather forecasting systems. Also, the database created demonstrates a positive possibility of storing weather data. The development and design of an efficient weather management system using MySQL are crucial for handling diverse weather data. Standardizing the database structure to ensure maximum efficiency involves implementing techniques like normalization and indexing.

3.3. Limitations

This study only captures 6 states out of 36 states in Nigeria, which indicates that the results of this study should be attributed to these states alone. However, the model could be adopted to other states within the country with caution. Nigeria is characterized by three distinct climate zones: a tropical monsoon climate in the south, a tropical savannah climate for most of the central regions, and a Sahelian hot and semi-arid climate in the north of the country. The southern regions experience strong rainfall events during the rainy season; the central regions are governed by a well-defined single rainy season and dry season. The dry season is influenced by the Harmattan wind from the Sahara. Coastal areas experience a short drier season with most rain. In the north, rain only falls from June to September while the rest of the year is hot and dry. Northern areas have a high degree of annual variation in their rainfall regime, which results in flooding and droughts. Highest temperatures occur during the dry season and vary little from the coast to inland areas. Similar to rainfall, the relative humidity in Nigeria decreases from the south to the north, with an annual mean of 88% around Lagos. The most significant temperature difference in Nigeria is between the coastal areas and its interior as well as between the plateau and the lowlands which is significantly different from other countries, and it literally impairs the generalization of this project. Also, limited access to comprehensive and accurate historical weather data may constrain the depth and accuracy of the analysis and predictive models developed in this project. Also, the severity was derived based on rule-based heuristics derived from the same meteorological variables used for model training different from using tools such as Doppler radar, weather satellites, and various ground-based instruments like thermometers, barometers, and anemometers which may inflate classification performance. As a result, reported accuracy reflects internal consistency rather than independent predictive validity. In addition, a single train–test split was employed without explicit temporal or spatial separation, limiting generalizability across locations and time periods.

.

IV. CONCLUSION

This study demonstrates the promising potential of adopting machine learning algorithms for weather prediction. The comprehensive analysis of weather variables reveals critical insights into their impacts on different regions and socio-economic factors. The high accuracy and reliability of the models in predicting severe weather conditions underscore its practical applicability in weather forecasting and disaster management. Furthermore, the development of an efficient weather data management system using MySQL highlights the importance of robust data storage solutions in supporting advanced analytical models. Implementing such systems can significantly mitigate the adverse impacts of severe weather conditions, enhance agricultural productivity, and improve disaster preparedness and response.

Future research could focus on incorporating more diverse weather variables and exploring advanced machine learning algorithms to further refine predictive accuracy. Additionally, integrating real-time weather data and developing user-friendly interfaces for stakeholders can enhance the practical utility of weather management systems, contributing to sustainable development and resilience against climate change in Nigeria

AUTHOR'S CONTRIBUTION

Conceptualization was done by B.O. O. & J. D. A. Data curation was done by B.O.O. & J. D. A., Formal analysis was done by J. D. A., Methodology B.O. O., J. D. A. & A. M. O. Validation J. D. A., A. M. O. & M. I. O. Visualization J. D. A. Writing: B.O. O., J. D. A., S. O. O., A. M. O., C.O. A., M. I. O., G. E. I., E. J. A., C. K. O., C. O. R., H. S. S. and T. A.

CONFLICT OF INTEREST

We declare that there is no conflict of interest related to this research.

FUNDING

This research received no funding

AVAILABILITY DATA AND MATERIALS

The data used in this study can be accessed via the link *josephdamilare01/Building-a-database-for-managing-weather-data-and-Algorithm-for-tracking-severe-weather-patterns.*

REFERENCES

Adekunle, J., Oyeniran, M., Ayanlowo, E. J., Dada, A. M., & Robert, C. O. (2024). Machine learning application in aquaculture: Predicting fish aggressive behavior and growth in diverse aquatic environment. TIJER – International Research Journal, 11(10).

Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine Learning Methods in Weather and Climate Applications: A Survey. Applied Sciences, 13(21), 12019. https://doi.org/10.3390/app132112019

Chinta, S. (2025). Integrating Machine Learning Algorithms in Big Data Analytics: A Framework for Enhancing Predictive Insights. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5046555

Cui, Y., Ouyang, S., Zhao, Y., Tie, L., Shao, C., & Duan, H. (2022). Plant responses to high temperature and drought: A bibliometrics analysis. Frontiers in Plant Science, 13, 1052660. https://doi.org/10.3389/fpls.2022.1052660

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PloS One, 15(2), e0228987. https://doi.org/10.1371/journal.pone.0228987

Hachimi, C. E., Belaqziz, S., Khabba, S., Sebbar, B., Dhiba, D., & Chehbouni, A. (2023). Smart Weather Data Management Based on Artificial Intelligence and Big Data Analytics for Precision Agriculture. Agriculture (Basel), 13(1), 95. https://doi.org/10.3390/agriculture13010095

Jain, A., Patel, H., Nagalapatti, L., Gupta, N., Mehta, S., Guttula, S., ... & Munigala, V. (2020). Overview and Importance of Data Quality for Machine Learning Tasks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3561–3562. https://doi.org/10.1145/3394486.3406477

Jiang, W., Huang, Y., & Sha, A. (2018). A review of eco-friendly functional road materials. Construction & Building Materials, 191, 1082–1092. https://doi.org/10.1016/j.conbuildmat.2018.10.082

Khan, W., Kumar, T., Zhang, C., Raj, K., Roy, A. M., & Luo, B. (2023). SQL and NoSQL Database Software Architecture Performance Analysis and Assessments—A Systematic Literature Review. Big Data and Cognitive Computing, 7(2), 97. https://doi.org/10.3390/bdcc7020097

Lee, J. J., Gino, F., & Staats, B. R. (2014). Rainmakers: Why Bad Weather Means Good Productivity. Journal of Applied Psychology, 99(3), 504–513. https://doi.org/10.1037/a0035559

Lei, M. Q., & Ming, C. W. (2023). The Empowerment of Digital Marketing among SMEs in Chengdu, Sichuan Province, China: The Influence of social media towards Purchase Decision. Journal of Digitainability, Realism & Mastery (DREAM), 2(09), 32–42. https://doi.org/10.56982/dream.v2i09.155

Malhi, G., Kaur, M., & Kaushik, P. (2021). Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. Sustainability, 13(3), 1318. https://doi.org/10.3390/su13031318

Muppala, M. (2025). Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5367993

Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. Big Data and

Cognitive Computing, 6(4), 132. https://doi.org/10.3390/bdcc6040132
Nigerian Meteorological Agency (NiMet). (2023a). Seasonal climate prediction (SCP) for Nigeria. Abuja, Nigeria: Nigerian Meteorological Agency.
Nigerian Meteorological Agency (NiMet). (2023b). Severe weather and heat advisory bulletin. Abuja, Nigeria: Nigerian Meteorological Agency.
Nigerian Meteorological Agency (NiMet). (2023c). Early warning advisory reports. Abuja, Nigeria: Nigerian Meteorological Agency.
National Oceanic and Atmospheric Administration (NOAA). (2022). Storm events database documentation. National Centers for Environmental Information, NOAA.
National Weather Service (NWS). (2021). Heat index and wind advisory criteria. National Oceanic and Atmospheric Administration.
Ogbuabor, J. E., & Egwuchukwu, E. I. (2017). The Impact of Climate Change on the Nigerian Economy. International Journal of Energy Economics and Policy, 7(2), 217.
Osakwe, J., Mutelo, S., & Obijiofor, N. (2023). Integrating Customer Relationship Management and Business Intelligence to Enhance Customer Satisfaction and Organisational Performance. A Literature Review. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4664792
Waring, S. M. (2021). Integrating primary and secondary sources into teaching: The SOURCES framework for authentic investigation. Teachers College Press.
World Health Organization. (2018). Low indoor temperatures and insulation. In WHO housing and health guidelines. World Health Organization. https://www.ncbi.nlm.nih.gov/books/NBK535294/
World Meteorological Organisation. (2015). WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services. World Meteorological Organisation. https://docs.edtechhub.org/lib/EYD24S8L
World Meteorological Organization (WMO). (2018). Multi-hazard early warning systems: A checklist. World Meteorological Organization.