Geospatial Artificial Intelligence (GeoAI) Insights
ISSN: XXXX Vol. 01, Issue: 01, 44-72, 2026
DOI: XXXX

Gis and Data Science in Enhanced Forecasting and Risk Assessment for Disaster Management in Africa

Owan Ayuk Daisy Ajijia[1,*], Roger Blaise Limaleba[1], Abba Nkasse Alain[1,2], Jean Chills Amba[2], Georges Kogge Kome[1], Yassine Mouniane[3], Merlin Bodol Momha[2], Mohamed Achab[4], Bikie Gérald Anicet[1,2,4]

[1] Department of Land Surveying, National Advanced School of Public Works, Yaoundé, Cameroon.
[2] Laboratory of Architecture, Urban Planning and Eco-construction, Advanced School of Urban Planning and Tourism Sciences, P.O. Box 47, Abong-Mbang, Cameroon.
[3] Natural Resources and Sustainable Development Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco.
[4] Geo-Biodiversity and Natural Resources Laboratory, Mohammed V University, Rabat, Morocco.

** Corresponding author: yassine.mouniane@uit.ac.ma*

Abstract: Climate change presents one of the most pressing challenges of the 21st century, with profound implications for ecosystems, human health, and socioeconomic stability, particularly in Africa. This study aims to integrate Geographic Information Systems (GIS), data science, and data modeling techniques to improve climate change forecasting and risk assessment in Africa. Through a comprehensive analysis of climate variables, disaster event data, and socioeconomic factors, this research develops predictive models that enhance the ability to forecast climate hazards such as floods, droughts, and heatwaves. Utilizing data from meteorological stations and global climate databases, this study implements machine learning models, time series analysis, and spatial analysis to forecast future climate scenarios. The methodology used in this study includes data cleaning, correlation analysis, model calibration and validation to ensure accuracy in predicting the onset of climatic events. The findings emphasize the role of GIS in visualizing climate impacts across Africa's diverse geographic regions, while the data models provide valuable insights for developing evidence-based policies aimed at disaster preparedness and mitigation. By integrating GIS and advanced data analysis techniques, this work contributes to a deeper understanding of Africa's climate dynamics and provides tools to support more effective decision-making in the face of growing climate threats. This interdisciplinary approach underscores the importance of leveraging technological innovations to build a resilient and sustainable future in Africa.

Keywords: Climate change, Modeling, Forecasting, Disaster, Geographic Information Systems.

Received: 4 Sept. 2025; Revised: 10 Sept. 2026; Accepted: 12 Sept. 2026; Published: 3 Nov. 2026

© 2026 The Author(s). Published by Science Transactions. License: CC BY 4.0

1 Introduction

Climate change represents one of the most pressing challenges of the 21st century, posing significant threats to ecosystems, human health, and economic stability worldwide [1]. It has a huge potential to undermine years of innovation and progress in global health, social living and civilization at large [2]. Countries in East Africa are prone to climate variability and extreme climate events such as droughts and floods [3], [4] with high variability in seasonal rainfall and daily rainfall extremes [5], [6]. In March 2024, a heat wave in East Africa led to emergency measures, including a two-week school closure in South Sudan, with temperatures projected to reach 41 to 45 degrees Celsius (106 to 113 degrees Fahrenheit). The said heat wave followed February's unusually high temperatures, particularly in southern Africa—Botswana, Namibia, Mozambique, South Africa, Zambia, and Zimbabwe—which reported temperatures that were 4 to 5 °C above average [7]. For the Sahel region of West Africa, there is little consistency amongst climate models with regard to the direction of future rainfall changes, and most models are not capable of representing the interannual rainfall variability observed over this region [8]. Current global climate models (GCMs) are known to exhibit significant systematic errors in representing African climate [9]. These errors include large biases in representing sea-surface temperatures along the coast of west Africa and inadequate representations of the African Monsoon [10]. By leveraging robust data analysis techniques, we can uncover hidden correlations, assess risks and develop evidence-based policies [11]. Scientific research provides the foundational knowledge necessary to comprehend the mechanisms driving climate change and its impacts. Through interdisciplinary studies, we investigate the interactions between atmospheric, oceanic, and terrestrial systems, with a minimal study on the influence of human activities on these processes [12]. This scientific understanding is crucial for developing accurate models, assessing the effectiveness of interventions, and eventually guiding policy decisions [13]. Following the aforesaid, this study sought to answer the question of how to mitigate the impact of climate induced events in Africa by improving preparedness/response strategy through the use of Data Science. The main objective of this study was to explore the synergistic use of data analysis, science, and GIS to address climate change and predict future events to facilitate risk management. Specifically, the objectives were;

1. To study the existing relationship between climatic variables such as temperature and climate-related events such as floods.
2. To develop a model for predicting future values of climatic variables such as temperature, pressure, wind speed, and cloud cover, among others, for different African countries.
3. To develop a model for predicting the occurrence of climatic events such as droughts, heat waves, and floods in different countries.

2 Literature Review

Geographic Information Systems (GIS) and data science are increasingly important for addressing growing climate-related disaster challenges on the continent of Africa [14]. The

literature demonstrates the value of spatial data combined with predictive analytics in improving disaster preparedness and risk management. The study in [15] state that surface temperatures in Africa are likely to rise faster than any other part of the world, making the continent particularly vulnerable to climate extremes, such as floods, droughts, or heatwaves. The adverse effects of these climate extremes tend to be worsened by lack of infrastructure, rapid urbanization, and limited access to early warning systems, so further 4 data-driven solutions are urgently needed.

Many studies have shown the promise of statistical and machine learning models for predicting climatic variables and extreme weather events. One simple example is [16], who utilized simple linear regression to predict climate changes. This study showed that linear regression was able to model temperature trends, but due to the complexity of climate systems where non-linear interactions are present and heavily affect the data, the model illustrated the limits of linear regression. More complex models like Random Forest and XGBoost have become more popular due to their complexity and improved accuracy with the amount of non-linear data in typically large datasets with multiple dimensions. Recent studies by [17], and [18] demonstrated the success of using Random Forest and XGBoost models, respectively, the study [19] also illustrated the application of these models for time series forecasting to succeed in other real-time forecasting tasks, including the pandemic and climatic events.

GIS instruments have supported analyses about the identification of hazardous areas and simulation of risk spread associated with hazards in the field of disaster risk assessment [20]. In a forensic analysis of East African countries, demonstrated how GIS based analyses could identify patterns of vulnerability and therefore identify the areas of intervention [21]. The use of socio-economic characteristics, land use and land cover, and topographic characteristics within a GIS allows for a multi-dimensional context [22]. These methods complement the increasing amount of research taking place on multi-criteria spatial analysis as part of risk/multirisk management approaches.

Time series analysis to examine climate variables was broadly discussed and examined. For example, the Box-Jenkins process remains the basis for modelling temporal components particularly ARIMA and SARIMA models [23]. More recently, the study by [24] noted the importance of stationarity, autocorrelation diagnostics, and model building. Climate researchers have also been employing the decomposition of time-series data into trend, seasonal, and residuals, applied to better understand long-term variability and short term shocked anomalies. In addition, classification models have been used to predict the occurrence of discrete disaster events, based on climatic predictors. Although logistic regression has important limitations on complexity, it has been useful in classifying binary events [25]. Nevertheless, Support Vector Machines (SVMs) will struggle with real world disaster datasets and comparatively will fall short pre-preprocessing and feature engineering despite their theoretical robustness [26]. In addition, ensemble learning methods, especially Random Forest and XGBoost will outperform other classifiers, because they are capable of modeling complex and non-linear relationships while handling imbalanced datasets [27].

3 Research Methodology

3.1 Presentation of the Study Area

This study was conducted across 12 African countries, spanning the biophysical diversity of the continent. Africa is a diverse continent with a rich tapestry of geography, cultures, and history. It is the second-largest continent in the world, both in size and population. Africa is made up of 54 countries and is home to over a billion people and was expected to reach 2.5 billion people by 2005 who speak a vast array of languages and practice numerous traditions and customs [28].

Africa's geographical extent is vast, affecting its climate, ecology, and daylight hours across the continent. The northernmost point in Africa is Ras ben Sakka in Tunisia, located at approximately 37°21' North. This places it in the temperate zone, experiencing a Mediterranean climate. The southernmost point is Cape Agulhas in South Africa, situated at around 34°50' South. This also experiences a Mediterranean climate, albeit much cooler and windier. The equator passes through the middle of Africa, crossing countries like Gabon, Republic of the Congo, Democratic Republic of the Congo, Uganda, Kenya, and Somalia. This equatorial crossing contributes to the tropical climates found in these areas, characterized by consistent temperatures year-round and abundant rainfall. The westernmost point is Pointe des Almadies in Senegal, at about 17°33' West. This part of Africa experiences a tropical savannah climate with a pronounced rainy season and a dry season. The easternmost point is Ras Hafun in Somalia, at 51°27' East. The climate here is arid to semiarid, influenced by its proximity to the Indian Ocean but shielded from heavy rainfall by highlands. These latitudinal and longitudinal ranges give Africa its diverse climates and ecosystems, from lush rainforests near the equator to harsh deserts in the north and south, along with varied day lengths and seasonal patterns across the continent. The Köppen climate classification, one of the most widely used climate classification systems [29], identifies thirteen (13) different climate types in Africa, largely reflected in the combination of different precipitation patterns and vegetation types across different geographical zones of Africa. Based on climatic diversity and the geographical repartition of each African country, sixteen countries were selected for this study. Each of these countries uniquely represent at least one climate type as well as emerge from each zone of Africa i.e Northern, Southern, Western, Central and Eastern Africa. The selected countries involved in this study all fit into the 13 climate types as shown in Table 1.

3.2 Data

In this study, the need for climate-related data, demographic data and disaster-related data was assessed and collected for a period of 23 years (2000 - 2023), with a monthly interval for each country. Climatic data included Surface Temperature, Earth Skin Temperature, Temperature at 2 meters, Specific Humidity and Relative Humidity at 2 meters, Wind Speed at 2 meters, Wind Direction at 2 meters, Wind Speed at 10 meters, Cloud Amount, Corrected Precipitation and Root Zone Soil Wetness, obtained from the NASA POWER Project's Data Access Viewer (DAV) which provides solar radiation and meteorological datasets from NASA research for supporting renewable energy, building energy efficiency and agricultural needs.

Table 1. Countries involved in our study and their associated Climate Type as per the Koppen climate classification.

Climate Type	**Countries**
Tropical Rainforest (Af)	Cameroon, Congo and Nigeria.
Tropical Monsoon (Am)	Mozambique, Sierra Leone
Tropical Savanna (Aw/As)	Kenya, Burkina Faso, Tanzania, South Sudan and Ethiopia
Hot Desert (BWh)	Algeria, Namibia
Hot Semi-Arid (BSh)	Mauritania and Chad
Subtropical Oceanic Highland (Cwb)	Lesotho
Hot Mediterranean (Csa)	South Africa
Temperate Mediterranean (Csb)	Parts of Morrocco and Algeria, western South Africa
Temperate Oceanic Climate (Cfb)	Coastal South Africa
Cold Desert (BWk)	Parts of Namibia
Cold Semi-Arid (BSk)	Morocco and Algeria.
Humid Subtropical Climate (Cwa)	Mozambique
Hot Oceanic Climate (Cfa)	Coastal South Africa

Additionally, disaster related data was gathered which included Origin of Climatic Event, Magnitude/Severity of Impact, Magnitude Unit, Presence of River Basin, Start Month of Event, Start Day of Event, End Year of Event, End Month of Event, End Day of Event, Total Deaths due to Event, Number of People Injured due to Event, Number of People Affected due to Event, Number of People Homeless due to Event and Total Affected due to Event. This dataset was sourced from the EM-DAT database (https://www.emdat.be), an emergency events database which records natural and technological disasters, created in 1988 as a joint initiative between the World Health Organization and the Centre for Research on the Epidemiology of Disasters (CRED). Furthermore, topographic and demographic data comprising of each country's Total Population, Percentage of rural population, Carbon dioxide emissions, Percentage of Urban Population, Air Pollution, Percentage of Land Use and Socio-Economic Vulnerability was gathered from the World bank open database. (https://data.worldbank.org) After sourcing of these data, restructuring and merging into .csv files from different sources was done to ensure better manipulation, analysis and inferencing and to produce a structured table with all the data. This resulted in a data frame of 36 columns(variables) and 9936 rows (data points) for each country in our study, giving an overall count of 129168 datapoints. The variables used in this study are shown in Table 2.

3.3 Data cleaning

Data cleaning involved elimination of duplicates, handling missing data, correcting inconsistent data, handling outliers, handling incorrect data types, removing irrelevant data, renam-

ing columns for clarity, standardizing data, filtering out unnecessary observations, dealing with data inconsistencies, and addressing syntax errors. Given the voluminous nature of our data (which could be considered as big data) and its variety, a parallel big data cleaning system was used as well, known as the Cleanix, designed and proposed by [19] which tackles four types of data quality issues namely abnormal value detection, incomplete data filling, deduplication and conflict resolution [30].

Table 2. Presentation of variables used in this study.

Variable	**Variable Definition**
PS	Surface Pressure (kPa)
TS	Earth Skin Temperature (°C)
T2M	Temperature at 2 Meters (°C)
QV2M	Specific Humidity at 2 Meters (g/kg)
RH2M	Relative Humidity at 2 Meters (%)
WS2M	Wind Speed at 2 Meters (m/s)
WD10M	Wind Direction at 10 Meters (m/s)
WS10M	Wind Speed at 10 Meters (m/s)
CLOUD_AMT	CERES SYN1deg Cloud Amount (%)
PRECTOTCORR	Precipitation Corrected (mm/day)
PRECTOTCORR_SUM	Precipitation Corrected Sum (mm)
GWETROOT	Root Zone Soil Wetness (1)
Average precipitation in depth	Average precipitation in depth (mm per year)
CO2 emissions	Carbon dioxide Emissions (kt)
Forest area	Forest area (% of land area)
Country	Country Affected
Origin	Origin of Climatic Event
Magnitude	Magnitude/Severity of Impact
Magnitude Scale	Magnitude Unit
River Basin	Presence of River Basin
Start Month	Start Month of Event
Start Day	Start Day of Event
End Year	End Year of Event
End Month	End Month of Event
End Day	End Day of Event
Total Deaths	Total Deaths due to Event
No. Injured	Number of People Injured due to Event
No. Affected	Number of People Affected due to Event
No. Homeless	Number of People Homeless due to Event
Total Affected	Total Affected due to Event
total_population	Country's Total Population
perc_rural_pop	Percentage of Rural Population
perc_urban_pop	Percentage of Urban Population
air_pollution	Air Pollution
land_use	Percentage of Land Use
socio_eco_vul	Socio-Economic Vulnerability

3.4 Descriptive and exploratory data analysis

Descriptive statistics are used to summarize data in an organized manner by describing the relationship between variables in a sample or population [31]. Different statistical attributes (such as median, mode, mean etc.) for the variables were computed and interpreted. Variability in the data was computed using standard deviation, variance, and range to understand its spread. Various graphs were generated to further understand the distribution of these variables over space and time in order to identify patterns and trends in the data, such as increasing temperatures or changes in precipitation distribution across regions. An impact analysis is conducted as well by visualizing which countries are most affected, in terms of people affected and infrastructural damage. The level of impact can give insight to the level of severity of the event.

3.5 Correlation Analysis

Correlation is meant for exploring the degree of relationship between two variables in consideration. A correlation coefficient is the measure to quantify such degree of relationship of the variables [32]. This analysis is done in order to deeply understand the effect of each variable on another. Due to the nature of the data's distribution, the Spearman's rank correlation and the Pearson's product moment correlation were used. Spearman's rank correlation is a non-parametric measure of relation between variables, using ranks to calculate the correlation. Pearson's correlation describes how well relationship between variables can be described in [33].

3.6 Time-Series Analysis/Trend Analysis

After the correlation analysis, a script is written to analyze and forecast climate variables over time. Results of the time series analysis permit us to break down time-series data into trend, seasonal, and residual components to better understand cyclical patterns.

3.7 Statistical testing

Statistical testing plays a crucial role in time series prediction to ensure that the model used is robust, unbiased, and accurately represents the underlying data patterns. The methods used for our statistical testing in time series prediction include;

3.7.1 Stationarity testing.

Time series data must often be stationary (i.e., its statistical properties like mean and variance remain constant over time) for many predictive models like ARIMA. We test for stationarity in three ways.

i. Using Augmented Dickey-Fuller (ADF) Test

It is a test commonly used in econometrics and finance to test the stationarity of time series data. It helps determine if differencing or other transformations are needed to stabilize the time series and make them stationary [34].

ii. Visual Assessment

By plotting the time series, we can visually assess whether the data is stationary or not.

Seasonal decomposition is a statistical technique used to break down a time series into three distinct components: Trend(T), Season(S) and Residual(R). Seasonal decomposition is useful for understanding the underlying patterns in a time series, especially when forecasting future values or analyzing fluctuations in data over time. There are two common types of decomposition.

a. Additive Model decomposition. This assumes that the data is a linear combination of the three components.

$$y(t) = T(t) + S(t) + R(t) = T(t) + S(t) + R(t)y(t) = T(t) + S(t) + R(t) \quad (1)$$

We use this model when the seasonal variation and the trend remain constant over time. When the dataset is additive, we apply a normal differencing algorithm to make the data stationary.

b. Multiplicative Model decomposition. This assumes that the data is a multiplicative combination of the components.

$$y(t) = T(t) \times S(t) \times R(t) = T(t) \times S(t) \times R(t)y(t) = T(t)S(t)R(t) \quad (2)$$

In this case when the series is multiplicative, we apply a log transformation to make the series stationary.

3.7.2 Autocorrelation and Partial Autocorrelation Function application.

To determine a proper model for a given time series data, it is necessary to carry out the ACF and PACF analysis. These statistical measures reflect how the observations in a time series are related to each other. For modeling and forecasting purpose it is often useful to plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) against consecutive time lags [35]. These plots help in determining the order of AR and MA terms. Another measure, known as the Partial Autocorrelation Function (PACF) is used to measure the correlation between an observation k period ago and the current observation, after controlling for observations at intermediate lags (i.e. at lags ¡ k, where k stands for a defined period). At lag 1, PACF(1) is same as ACF(1) [35]. The Ljung-Box Test is also used to statistically test whether residuals from the models are white noise (no correlations in the residual) and the Jarque-Bera test is used to inform on whether the residuals are normally distributed.

3.8 Risk Analysis

Additionally, a risk analysis is carried out to identify regions with significant climate anomalies at high risk of being affected by climatic events. This risk analysis is done using ArcGIS pro following a careful study of topographic variables, climatic and event history. Following the identification of these different spots, this parameter is integrated during model building.

3.9 Predictive Modeling

For events prediction, similar machine learning models are applied, assessed and finetuned in order to ensure the most appropriate model is being captured and selected. The different models include;

a) Linear Regression Models

Regression involves estimation of the value of a variable from that of one or more other variables, with the relevant coefficients derived from empirical observations [36]. The dependent variable is usually denoted as the y-variable and the independent variables as the x-variables – i.e $x_1, x_2, x_3 \ldots x_n$. A regression equation identifies the relationship between the variables of the form:

$$y = a + b_1 x_1 + b_2 x_2 \ldots b_n x_n \pm se_y \tag{3}$$

where:

a is the constant term – the value of y when the values of all of the x variables are set at 0;

$b_1 \ldots b_n$ are the regression or slope coefficients, indicating the change in the value of y for each unit of change in the relevant x value;

se_y is the standard error, which is the measure of the variability of the estimated coefficients.

These models are commonly used for predicting temperature, precipitation, and other climatic variables over time. They are effective for modeling linear relationships in the data. The expected challenge here is climatic data often have non-linear and complex patterns, so linear models may not always capture these accurately.

b) Decision Trees and Random Forests

Decision Trees and Random Forests are effective in handling large datasets and capturing complex interactions between variables. The training algorithm for random forests applies the general technique of bootstrapping aggregating or bagging, to tree learners. Given a training set $X = x_1, \ldots, x_n$ with responses $Y = y_1, \ldots, y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples:

For $b = 1, \ldots, B$:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b
2. Train a classification or regression tree f_b on X_b, Y_b

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$f'' = \frac{1}{B} \sum_{b=1}^{B} (f_b)(x') \tag{4}$$

Random Forests are useful in predicting extreme events (such as rainfall, floods and droughts) based on multiple climatic variables [17]. Random Forests can be used to model non-linear relationships between climate factors like humidity, wind speed, and temperature to predict the likelihood of extreme weather events.

c) Support Vector Machines (SVM)

SVMs are supervised learning models that work by finding the optimal hyperplane (Decision boundary (linear or nonlinear) separating classes) that maximizes the margin between classes in high-dimensional space [37]. Mathematically, it aims at obtaining the hyperplane ($w^T x + b = 0$) that maximizes the margin ($2/\|w\|$). This leads to the following optimization problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{5}$$

Subject to the constraint:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i \tag{6}$$

where:
w = weight vector b = bias term $y_i \in \{-1, +1\}$ = class labels

SVMs are often used for classification problems, such as rainfall prediction, weather forecasting or predicting whether a particular climatic event (like a heatwave or storm) will occur [38]. SVMs work well with smaller datasets but can capture complex boundaries between classes. An important consideration though is, SVMs can be computationally expensive and may not scale well for very large climatic datasets.

d) Extreme Gradient Boosting Model (XGBoost)

Boosting refers to a class of learning algorithms that fit models by combining base models with basic functions to improve the fit of the final model. The gradient boosting framework supports various objective functions such as regression, classification and ranking. It optimizes a regularized loss function [39]:

$$L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{7}$$

where:
$l(y_i, \hat{y}_i)$ = loss function (e.g., MSE, log loss)
$\Omega(f_k)$ = regularization term for the k-th tree:

$$\Omega(f) = \gamma T + 0.5\lambda w^2 \tag{8}$$

T = number of leaves w = leaf weights γ = complexity control (min. loss reduction for a split) λ = L2 regularization

XGBoost provides state of the art results when it comes to climate prediction [40]. They are highly accurate, can handle a mix of continuous and categorical features, and provide feature importance, offering insights into which climatic factors are driving the predictions.

These models are tested on the dataset, which is split into testing and training data, precisely 70% for training and 30% for testing.

3.10 Model Selection and Validation

Time series models can be complex, and validating the performance is critical to ensure the reliability of predictions. To carry out the model selection and validation, different models with different parameters and orders (the values p, d, q, d, P, Q, D, S) are trained and the following metrics implemented in order to assess their accuracy. The model parameters are defined as follows:

p – the order of the autoregressive (AR) model and represents the number of lagged observations used to predict the current value

q – the order of the moving average(MA) model and represents the number of lagged error terms used to predict the current value.

d – the degree of differencing. It indicates the number of times the data has been differenced to achieve stationarity

S – the number of periods in each season (the seasonal period)

P – the order of seasonal AR model

Q – the order of the seasonal MA model

D – the degree of seasonal differencing.

- **AIC (Akaike Information Criterion)**

  AIC estimates the relative quality of statistical models for a given dataset. It balances the fit of the model against its complexity [41]. It is given in the Eq. (9).

$$\text{AIC} = 2k - 2\ln(L) \tag{9}$$

- **BIC (Bayesian Information Criterion)**

  BIC is similar to AIC but introduces a stronger penalty for the number of parameters, particularly as the size of the dataset increases [41]. Its formula is given below

$$\text{BIC} = k\ln(n) - 2\ln(L) \tag{10}$$

Where:

- k is the number of model parameters.
- n is the number of data points (observations).
- L is the likelihood function, which measures how well the model fits the data.

Our objective here is to minimize the BIC value and AIC value. The lower the BIC, the better the model in terms of balancing fit and simplicity. Likewise, the lower the AIC, the better the model fits the data with fewer parameters. The BIC introduces a heavier penalty for models with more parameters than AIC, especially as the number of data points increases. This makes BIC more conservative in favor of simpler models and better for larger datasets. AIC penalizes models that have more parameters. It encourages the use of simpler models while avoiding overfitting, though it typically tends to favor more complex models than BIC.

3.11 Model Diagnostics and Performance metrics

In order to diagnose the model generated, we focus on the residuals to the training data. The residual is the difference between the model's one-step-ahead predictions and the real values of the time series. We quantify our residuals through several evaluation metrics such as the accuracy, precision, recall and f1 score.

4 Results and Discussion

4.1 Descriptive and Exploratory Data Analysis Results

The different counts of events are plotted to give an overview of the most dominant events across our different countries of study. This result is presented in the histogram (Figure 1) displayed subsequently.

As per our dataset and bar chart, the most occurring event across the countries involved in our study is flood, with over 400 cases of flood recorded from 2000 to 2024. From the bubble plot shown in Figure 2, we realize that even though the most reoccurring events are floods, drought seem to create an extremely significant impact. Additionally, we observe that every country that experienced a storm, additionally experienced a flood. This could imply some sort of correlation (not causation though) as well as give insight to how much rainfall intensity (which is a climatic trait) affects flooding, which will be verified as we progress with our results and analysis.

Also seeing the distribution of events and magnitude of impact across countries give an insight into which countries are most vulnerable and so could be at a higher risk of these events occurring again. It is also important to observe that the most affected countries are Ethiopia, Nigeria, Kenya, South Africa, Burkina Faso, Nigeria, and South Sudan, which represent West and East Africa which could lead us to start considering the fact countries at the equator are more sensitive to climatic changes. This observation also concords with UNDRR's forensic analysis for countries in the Horn of Africa such as Somalia, Kenya and parts of Ethiopia which suffered severe droughts and floods, affecting more than 30 million people [42].

4.2 Correlation Analysis Results

The correlation matrix (Figure 3)shows that the year variable demonstrates very weak correlations (close to zero) with all other climate variables, indicating that the year itself does

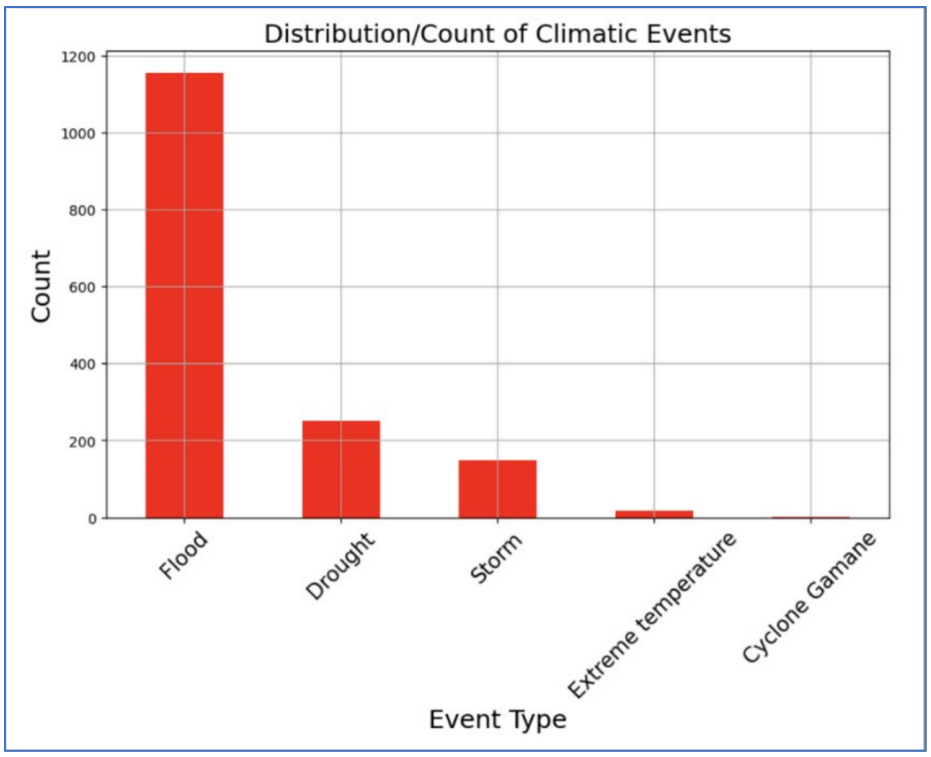


Fig. 1. Histogram showing count of events.

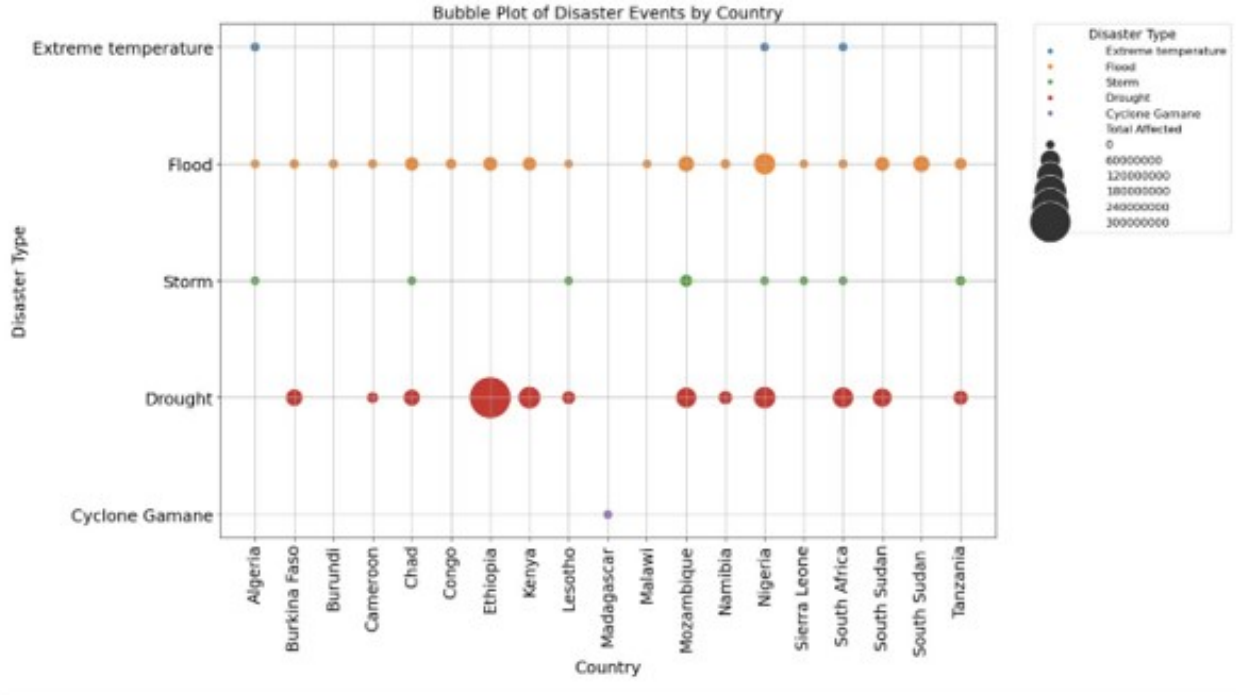


Fig. 2. Bubbleplot showing impact proportion of events per country.

not strongly correlate with any of the observed variables. This suggests no significant long-term upward or downward trend for these specific variables when looked at in isolation by year. Many of the climate variables (especially temperature, humidity, wind speed, and precipitation) are strongly positively correlated with each other. This reflects the interconnected nature of climate processes, where changes in one variable often influence others. Precipitation (PRECTOTCORR) has a significant impact on several other variables, such as soil moisture (GWETROOT), humidity, and cloud cover, all of which are expected in regions experiencing precipitation-driven climatic variations. The high correlation between temperature and wind speeds suggests that temperature fluctuations are often accompanied by windier conditions, which might be related to the movement of weather fronts or storm systems. This correlation matrix (Figure 3) provides insights into how various climate factors are interrelated and enable us to make predictions or understand weather patterns in a region.

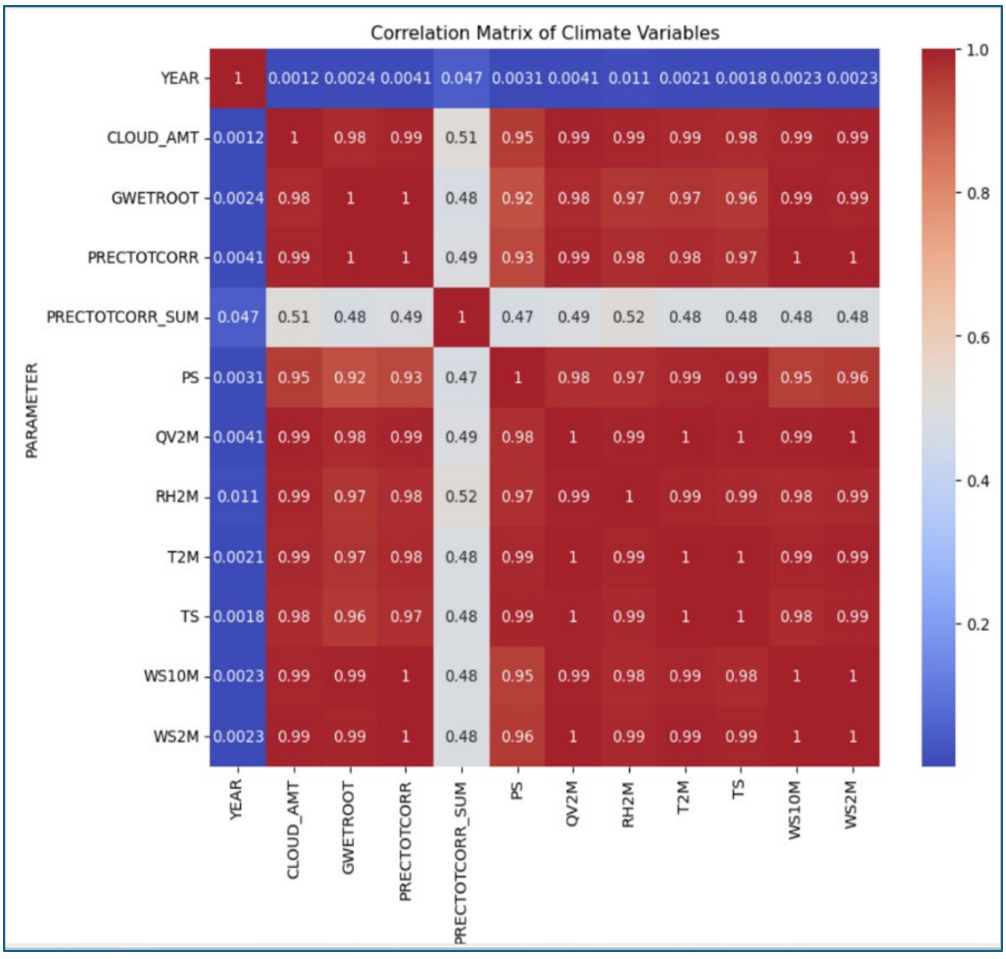


Fig. 3. Correlation Matrix showing degree of correlation between variables.

4.3 Time Series/Trend Analysis Results

Following the results of the correlation analysis are the results of the trend/temporal analysis. This enabled the visualization of the activity of our variables over time and especially give insight into what to expect in the coming years. Primarily, we begin by an overall analysis of our climatic factors to see how they behave over time, before visualizing their specific activity per country.

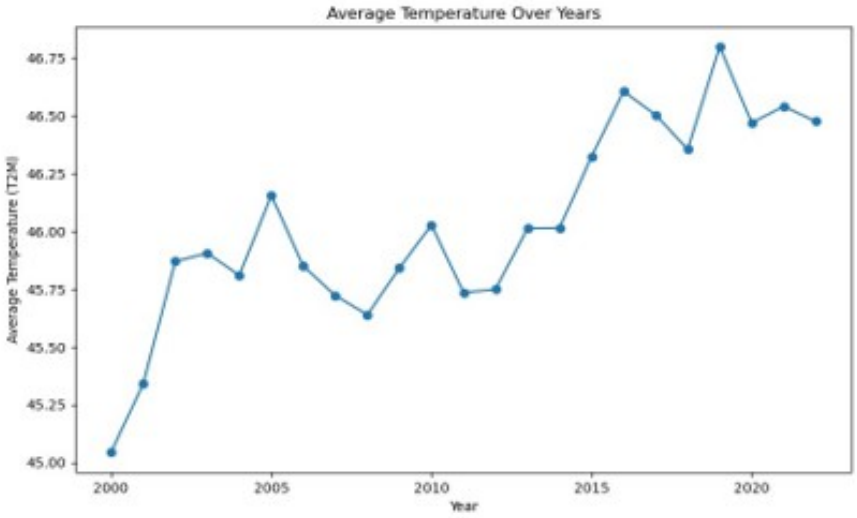


Fig. 4. Average Temperature in Africa over the years.

The line chart (Figure 4) illustrates the average temperature (T2M) over a span of years (from 2000 to 2024). An overall trend showing a gradual increase in average temperature over time is observed. The temperature starts at around 45.0°C in the year 2000 and increases to approximately 46.5°C by 2022. This suggests that the has been an overall warming trend over the last two decades. This observed increase is in line with climate change and other studies [43]. The temperature shows significant fluctuations year by year, where some years experience notable increases and others show dips. Between 2000 and 2005, there is a sharp increase in average temperature, rising by almost 1°C in five years. Around 2005, there's a notable peak in the temperature (around 46°C), followed by a significant decline until about 2010.

From 2010 onwards, there is a relatively consistent upward trend with minor dips, reflecting a general increase in temperature, punctuated by occasional decreases (e.g., in 2016 and 2020). The period from 2010 to 2022 shows a steady rise with a noticeable peak around 2016, and another around 2019. However, the last few years (post-2020) show a slight cooling or plateau, but the temperature remains well above earlier years (around 46.3°C). The chart (Figure 4) reflects a typical warming pattern observed in many regions globally due to climate change and suggests that we could be experiencing global warming impacts. The fluctuations seen year-to-year may be related to climatic variability—influenced by factors like atmospheric conditions, ocean currents (e.g., El Niño/La Niña cycles), and localized

weather phenomena [44]. We then proceed to visualize the climatic trends for each country closely.

Figure 5 represents a series of line plots showing temperature trends over time for various African countries from the year 2000 to around 2024. In terms of overall temperature stability, many countries show stable long-term trends without significant warming or cooling.

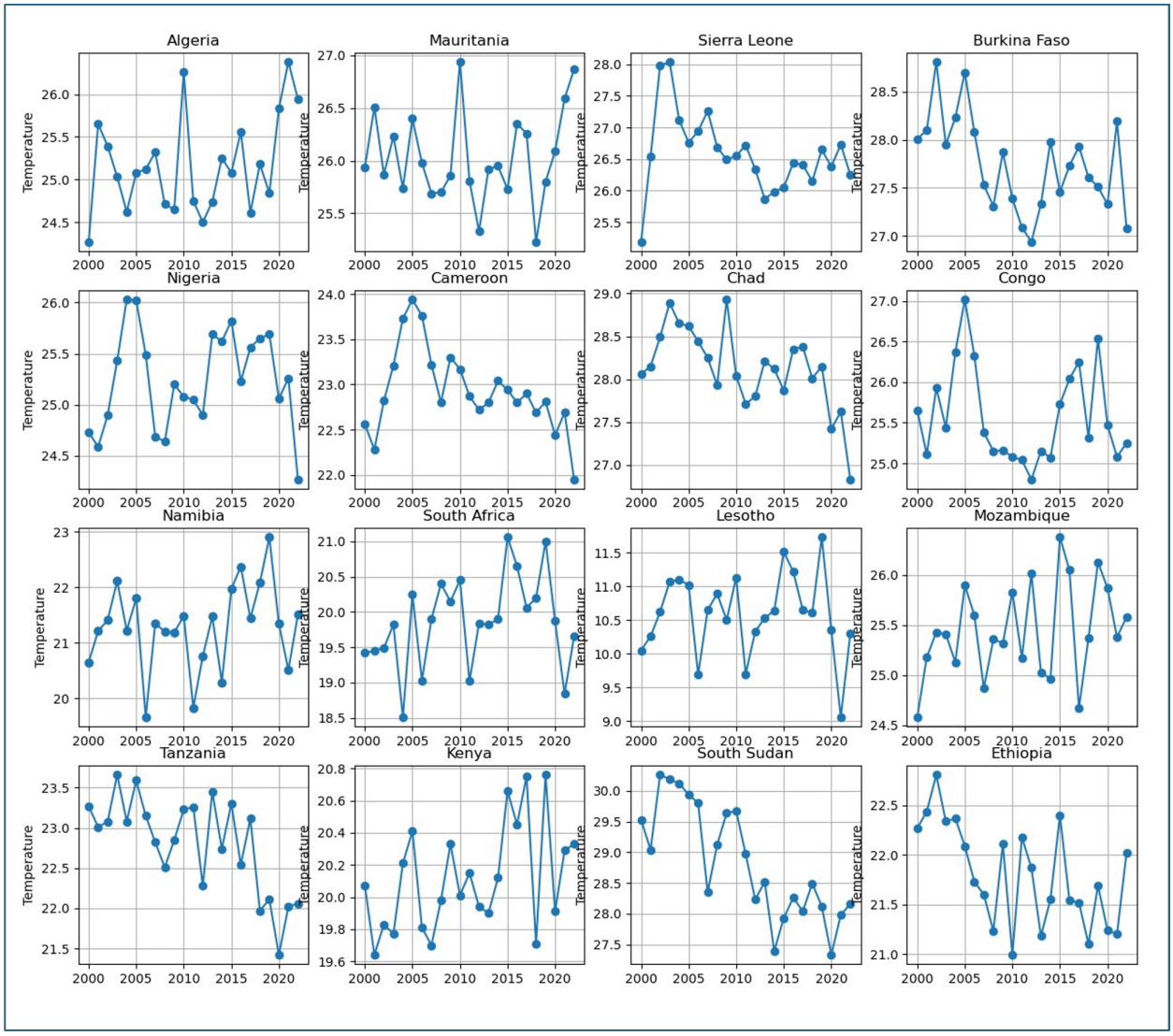


Fig. 5. Temperature distribution across sixteen African countries from 2000 to 2024.

The temperatures oscillate within a relatively narrow band, which suggests stable climates in some regions. However, there is some noticeable variability where some countries like Sierra Leone, Chad, and South Sudan exhibit high variability in their temperatures, with significant peaks and troughs, which may indicate more dynamic weather patterns or the influence of climate events (e.g., El Niño, La Niña) [44], [45]. Cooler Countries like Lesotho have the coolest average temperatures, consistently below 12°C. The stable yet low temperatures may reflect the country's higher altitude and unique climatic conditions. Warmer

Countries like South Sudan, Chad, and Burkina Faso show higher overall temperatures compared to other countries, reflecting their location in warmer regions with drier climates.

4.4 Statistical Testing Results

Results of the Augmented-Dickey fuller (ADF) test is depicted in Table 3. Upon application of this test and evaluation of this Table 3, the following is observed.

Table 3. Statistical measures at different lags for each country.

Country	ADF Statistic	p-value	Lags Used	Observations
Algeria	-3.286	0.016	14	261
Burkina Faso	-2.686	0.076	13	262
Cameroon	-1.491	0.538	13	262
Chad	-1.451	0.558	14	261
Congo	-2.772	0.062	14	261
Ethiopia	-3.130	0.024	13	262
Kenya	-3.835	0.003	13	262
Lesotho	-4.177	0.001	15	260
Mauritania	-3.927	0.002	14	261
Mozambique	-4.490	0.000	15	260
Namibia	-3.869	0.002	14	261
Nigeria	-4.054	0.001	16	259
Sierra Leone	-2.613	0.090	13	262
South Africa	-3.839	0.003	15	260
South Sudan	-1.849	0.356	13	262
Tanzania	-2.508	0.114	13	262

Countries with Strong Evidence of Stationarity

- **Lesotho:** ADF Statistic (-4.177) and p-value (0.0007) show significant stationarity at all levels (1%, 5%, 10%).
- **Mozambique:** ADF Statistic (-4.490) and p-value (0.0002) suggest very strong stationarity, as the test statistic is well below the critical values.

Countries with Moderate Evidence of Stationarity

- **Kenya:** ADF Statistic (-3.835) and p-value (0.0026) indicate stationarity at the 1% and 5% levels but are closer to the 10% boundary.
- **South Africa:** Similar to Kenya, its ADF Statistic (-3.839) and p-value (0.0025) imply moderate evidence for stationarity.

Countries with Weak or No Evidence of Stationarity

- **South Sudan:** ADF Statistic (-1.849) and p-value (0.356) do not provide sufficient evidence for stationarity.
- **Cameroon:** ADF Statistic (-1.491) and p-value (0.538) show the series is non-stationary.

Countries Near the Marginal Significance Levels

- **Congo:** ADF Statistic (-2.772) and p-value (0.062) indicate borderline stationarity, significant only at the 10% level.
- **Burkina Faso:** Similar results with an ADF Statistic (-2.686) and p-value (0.076), close to the 10% significance threshold.

Patterns Across Countries

- **Strong Evidence for Stationarity:** Countries like Mozambique, Lesotho, and Mauritania have highly negative ADF Statistics and low p-values, making them strongly stationary.
- **Borderline Evidence:** Countries like Congo and Burkina Faso are borderline cases where stationarity might depend on context or adjusted significance levels.
- **Non-stationary Series:** Countries like South Sudan, Cameroon, and Chad have higher p-values, suggesting their time series data are likely non-stationary.

Next, we proceed by performing ACF and PACF analysis to identify the lag relationships within the time series for each country, the proper time series model to use (AR, MA, ARMA, ARIMA, SARIMA, SARIMAX), the period and the proper seasonal and non-seasonal orders. A sample of the results of our ACF and PACF plots are presented in Figure 6 .
From the plots, we deduce the non-seasonal orders range from 1 to 3, each country displays a periodicity of 12 with DRC exceptionally having a periodicity of 24 and a seasonal order of 1 for our 13 countries. Conclusively, we end our statistical testing by calculating the AIC and BIC values, shown in Table 4.

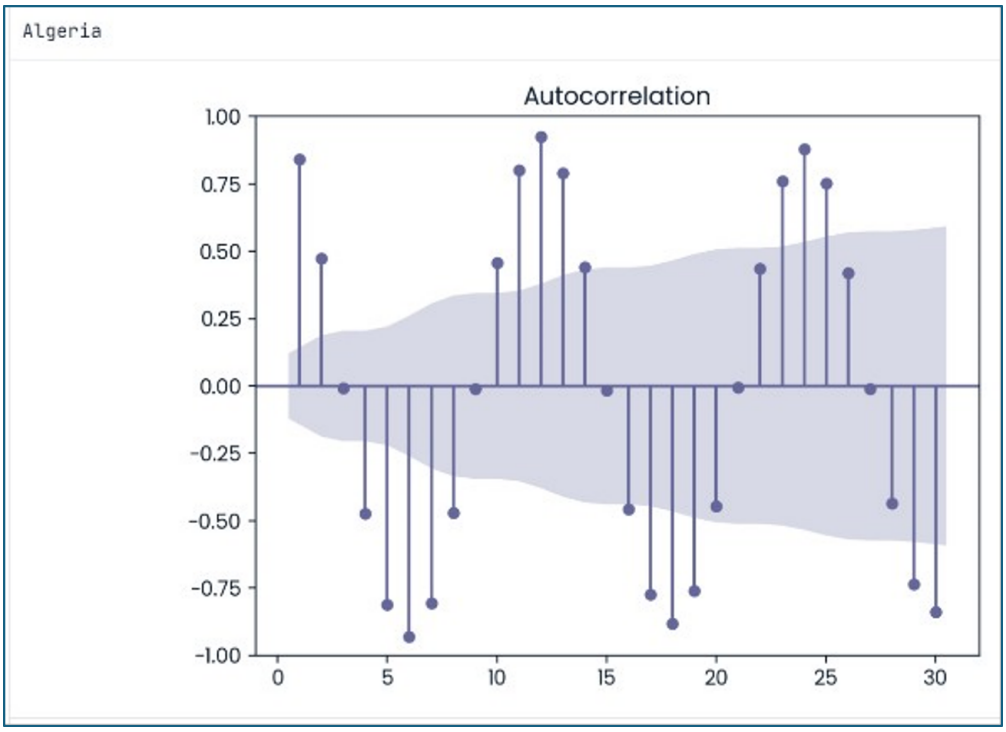


Fig. 6. ACF Time series plot for Algeria.

Table 4. AIC and BIC values resulting from predicting wind speed at 2 meters.

Country	Parameter	AIC	BIC
Algeria	WS2M	373.069	390.930
Burkina Faso	WS2M	-38.421	-16.988
Cameroon	WS2M	-296.056	-278.195
Chad	WS2M	323.070	340.931
Congo	WS2M	-658.034	-643.746
Ethiopia	WS2M	-11.224	10.209
Kenya	WS2M	207.517	225.378
Lesotho	WS2M	55.919	70.208
Mauritania	WS2M	216.286	234.147
Mozambique	WS2M	34.622	45.339
Namibia	WS2M	15.791	26.507
Nigeria	WS2M	254.293	275.726
Sierra Leone	WS2M	-218.571	-200.710
South Africa	WS2M	2.000	5.572
South Sudan	WS2M	45.763	63.624
Tanzania	WS2M	71.676	89.537

4.5 Predictive Modeling results - Climate Variable Prediction

The AIC and BIC values obtained guided the choice of our model parameters. For the AIC and BIC values presented in Table 4 for Wind Speed at 2 meters, the SARIMA model parameters were deduced for the same climatic property as shown in Table 5.

Table 5. SARIMA Model Parameters for predicting wind speed at 2 meters.

Country	Parameter	p	q	P	Q	S	D	d
Algeria	WS2M	0	1	1	2	12	1	1
Burkina Faso	WS2M	0	2	2	1	12	1	1
Cameroon	WS2M	1	1	1	1	12	1	1
Chad	WS2M	0	1	1	2	12	1	1
Congo	WS2M	1	1	0	1	12	1	1
Ethiopia	WS2M	2	1	0	2	12	1	1
Kenya	WS2M	2	1	0	1	12	1	1
Lesotho	WS2M	0	2	0	1	12	1	1
Mauritania	WS2M	2	1	0	1	12	1	1
Mozambique	WS2M	2	1	0	1	12	1	1
Namibia	WS2M	0	1	0	1	12	1	1
Nigeria	WS2M	1	1	2	1	12	1	1
Sierra Leone	WS2M	1	1	0	2	12	1	1
South Africa	WS2M	0	0	0	0	12	1	1
South Sudan	WS2M	1	2	0	1	12	1	1
Tanzania	WS2M	1	2	0	1	12	1	1

4.6 Predictive Modeling results - Event Prediction

After finalising our climate variable prediction, we move to the events prediction given these events are predicted from climatic variables. Our problem type is a classification problem. Therefore, we proceed to train several adapted classification models on our data and the compare them to see which gives the most accurate result. The models adapted such as Logistic Regression, SVM, Random Forest, and XGBoost produced the outcomes with high accuracies mentioned in Table 6, 7, 8, and 9, respectively.

4.6.1 Model: Logistic Regression

Table 6. Classification report for Logistic Regression Model.

Class	Precision	Recall	F1-Score	Support
Storm	0.60	0.69	0.64	210
Flood	0.83	0.40	0.54	235
Drought	0.50	0.54	0.52	202
Extreme Temperature	0.71	1.00	0.83	193
Accuracy			0.64	840
macro avg	0.66	0.66	0.63	840
w. avg	0.67	0.64	0.63	840

4.6.2 Model: Support Vector Machine

Table 7. Classification report for SVM Model.

Class	Precision	Recall	F1-Score	Support
Storm	0.63	0.73	0.68	193
Flood	0.44	0.62	0.51	202
Drought	0.55	0.36	0.43	210
Extreme Temperature	0.52	0.43	0.47	235
Accuracy			0.53	840
macro avg	0.54	0.53	0.52	840
w. avg	0.53	0.53	0.52	840

4.6.3 Model: Random Forest

Table 8. Classification report for Random Forest Model.

Class	Precision	Recall	F1-Score	Support
Storm	0.99	1.00	0.99	210
Flood	1.00	0.92	0.96	235
Drought	0.93	1.00	0.96	202
Extreme Temperature	1.00	1.00	1.00	193
Accuracy			0.98	840
macro avg	0.98	0.98	0.98	840
w. avg	0.98	0.98	0.98	840

4.6.4 Model: Extreme Gradient Boosting

Table 9. Classification report for XGBoost Model.

Class	Precision	Recall	F1-Score	Support
Storm	1.00	1.00	1.00	193
Flood	0.92	1.00	0.96	202
Drought	0.98	1.00	0.99	210
Extreme Temperature	1.00	0.91	0.95	235
Accuracy			0.97	840
macro avg	0.97	0.97	0.97	840
w. avg	0.97	0.97	0.97	840

Where w. avg represents weighted average and macro avg represents macro average.

4.7 Model Comparison, Selection and Validation

Model selection is done based on the evaluation of the outcomes og model metrics mentioned in Tables 6, 7, 8, and 9, and as follows.

Logistic Regression: Precision, Recall, F1-Score. The model performs reasonably well for most classes, but struggles with the flood class (which is a minority class), where it has a recall of only 40%. The extreme temperature class performs the best with an F1-score of 0.83. This observation is in line with another studies [46], which demonstrates this model's ability to predict simple relationships such as temperature increase but no insight on complex patterns. Overall, the model is okay but not very strong. This suggests that Logistic Regression is a simpler model and may not capture complex relationships in the data and so is not the best option given the performance, especially with imbalanced classes.

Support Vector Machine: Precision, Recall, F1-Score. The SVM model struggles with the dataset, as the overall accuracy is lower (52.6%). It particularly struggles with Drought and extreme temperature classes, with low F1-scores (0.43 and 0.47). The recall and precision are generally inconsistent across classes. This suggests that SVM does not perform well on our dataset. Similarly [47] provided an attempt at using SVM for weather prediction demonstrated its inability to capture heavy precipitation events on its own, leading to a large negative bias in estimated precipitation, until combined with Quantile Mapping to produce higher statistics. Such underestimation was also found in previous studies but explained by non-Gaussian outliers in the training data. This potentially suggests that SVM is not adequate for capturing complex events.

Random Forest: Precision, Recall, F1-Score. Random Forest performs exceptionally well across all classes, with nearly perfect precision and recall. All classes achieve F1-scores close to 1.00, except for the flood class (slightly lower recall of 92%). Random Forest is a strong model for this dataset. It handles both the complexity and class imbalance very well, making it a highly suitable model. The efficiency of Random Forest on climatic data has also been observed in other works, with an accuracy ranging from 76% to 99% for forecasting [48].

XGBoost: Precision, Recall, F1-Score. XGBoost also performs very well, with an accuracy similar to Random Forest. It achieves near-perfect scores across all metrics, though the extreme temperature class has a slightly lower recall (91%). Like Random Forest, XGBoost performs excellently and is well-suited for our dataset, which is expected as studies by Ranjan et al., [19]., enunciated this fact as well while comparable results were found by [40]. for predicting wave height during storm events. It might be a good alternative to Random Forest, especially for more finetuned or large-scale datasets.

Furthermore, ROC curves and confusion matrices (Figure 7) were used to further illustrate the diagnostic ability of the different models. Our ROC curves were plotted using the One-vs-Rest (OvR) multiclass classification strategy. Here, class 0 represents storms, class 1 represents floods, class 2 represents droughts, and class 3 represents extreme temperature.

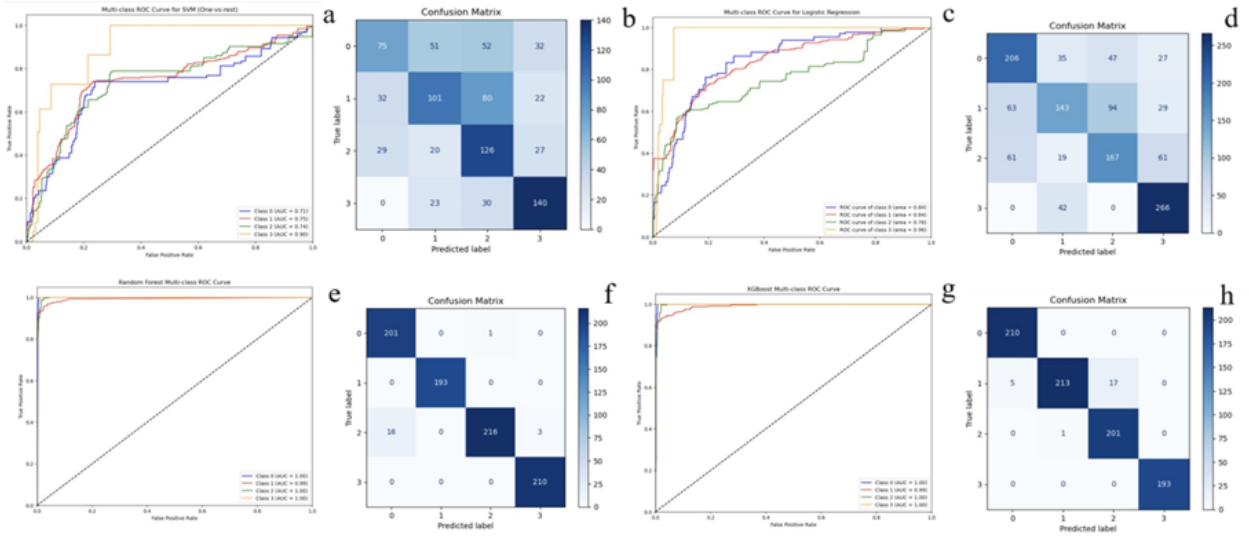


Fig. 7. a: ROC Curve for SVM; b: Confusion Matrix for SVM; c: ROC Curve for Logistic Regression; d: Confusion Matrix for Logistic Regression; e: ROC Curve for Random Forest; f: Confusion Matrix for Random Forest; g: ROC Curve for XGBoost ; h: Confusion Matrix for Confusion Matrix for XGBoost.

It is observed that SVM is best at classifying Class 3 (Extreme temperature, AUC = 0.90) and handles non-linear boundaries well (if using RBF kernel). Nonetheless, it struggles with Class 0 (Storm, AUC = 0.71) as the performance is inconsistent across classes. This could be due to the fact that the kernel choice may not be optimal for all classes. Logistic Regression has a balanced performance (AUC 0.76–0.84) and is particularly simple, interpretable, and computationally efficient but struggles with complex data as Class 2 (Drought, AUC = 0.76) has an overall weak link. On the other hand, XGBoost has a perfect classification (AUC = 1.00) for Classes 0, 2, 3 with Class 1 (flood, AUC = 0.99) being near-perfect. This indicates it handles non-linearity, missing data, and feature importance well. Random Forest is comparable to XGBoost as it achieves a perfect classification (AUC = 1.00) as well for Classes 0, 2, 3 with Class 1 (flood, AUC = 0.99) being nearperfect. It is also quite robust to outliers and noise. Conclusively, it is established that as per this work, Random Forest with an accuracy of nearly 98% and strong metrics across all classes is the most suitable model for our prediction, followed by XGBoost. This then justifies this model as the most

appropriate as per our study, for events prediction. Logistic Regression and SVM struggled with the dataset, showing lower accuracy and performance, especially with class imbalances.

5 Conclusions

This study aimed to establish a framework that enhances the precision of climate-related predictions and assessments, thus aiding in more informed decision-making processes for disaster preparedness and mitigation strategies. The methodology employed was the Box -Jenkins methodology (Box and Jenkins, 1970), applied over several models namely the AR, MA, SARIMA, ARFIMA, ARIMA, ARMA, SARIMAX, Support Vector Machine, XGBoost, Random Forest and Logistic Regression. The predictive models developed were instrumental in forecasting climatic conditions. Despite its innovative approach and significant findings, a major limitation of this study is the generalization of models. While the models developed are robust, their applicability might be limited to similar datasets and may not fit across all types of climatic events or geographic locations without adjustments. Another limitation is computational resources. Computational resources are required to process and analyze large datasets, which might not be readily available in all research or operational settings. Given the extensive nature of this study and inherent limitations, three areas are proposed for further research and improvement. Firstly, incorporating more ground truth data from local weather stations and real-time data could enhance model accuracy. Secondly, future research should focus on developing real-time forecasting models that can dynamically incorporate new data as it becomes available. Finally, the exploration of deep learning and neural networks could uncover deeper insights and improve the predictive capabilities of the models.

Author Contributions

All authors have made equal contributions to the content presented in this paper.

Conflict of Interest

The authors declare no conflict of interest.

Research Involving Human and/or Animals

Not Applicable.

Data Availability

Not applicable.

Funding

The authors received no financial support for this research.

AI Use Disclosure

The author(s) confirm that generative AI tools were not used in the preparation of this article.

References

[1] Goniewicz, K., Burkle, F. M., Khorram-Manesh, A. (2025). Transforming global public health: climate collaboration, political challenges, and systemic change. Journal of infection and public health, 18(1), 102615.10.1016/j.jiph.2024.102615

[2] Rubio, J. L., Reyes-Sánchez, L. B., Duihu, N., Costantini, E. A., Horn, R., Zlatic, M. (2024). Protecting the soil is protecting the climate WASWAC and IUSS position paper on the inter linkages of soil and climate change. Soil Security, 14, 100124.10.1016/j.soisec.2023.100124

[3] Taye, M. T., Dyer, E. (2024). Hydrologic extremes in a changing climate: a review of extremes in East Africa. Current Climate Change Reports, 10(1), 1-11. https://doi.org/10.1007/s40641-024-00193-9.

[4] Gebrechorkos, S. H., Hülsmann, S., Bernhofer, C. (2019). Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. International Journal of Climatology, 39(1), 18-30. https://doi.org/10.1002/joc.5777.

[5] Cattani, E., Merino, A., Guijarro, J. A., Levizzani, V. (2018). East Africa rainfall trends and variability 1983–2015 using three long-term satellite products. Remote Sensing, 10(6), 931.https://doi.org/10.3390/rs10060931

[6] Gebrechorkos, S. H., Hülsmann, S., Bernhofer, C. (2020). Analysis of climate variability and droughts in East Africa using high-resolution climate data products. Global and Planetary Change, 186, 103130. https://doi.org/10.1016/j.gloplacha.2020.103130

[7] Chemura, A., Nangombe, S. S., Gleixner, S., Chinyoka, S., Gornott, C. (2022). Changes in climate extremes and their effect on maize (Zea mays L.) suitability over Southern Africa. Frontiers in Climate, 4, 890210. https://doi.org/10.3389/fclim.2022.890210

[8] Engelbrecht, F., Adegoke, J., Bopape, M. J., Naidoo, M., Garland, R., Thatcher, M., ... Gatebe, C. (2015). Projections of rapidly rising surface temperatures over Africa under low mitigation. Environmental Research Letters, 10(8), 085004.https://doi.org/10.1088/1748-9326/10/8/085004

[9] Karypidou, M. C., Katragkou, E., Sobolowski, S. P. (2022). Precipitation over southern Africa: is there consensus among global climate models (GCMs), regional climate models (RCMs) and observational data?. Geoscientific Model Development, 15(8), 3387-3404. https://doi.org/10.5194/gmd-15-3387-2022

[10] Paxian, A., Sein, D., Panitz, H. J., Warscher, M., Breil, M., Engel, T., ... Paeth, H. (2016). Bias reduction in decadal predictions of West African monsoon rainfall using regional climate models. Journal of Geophysical Research: Atmospheres, 121(4), 1715-1735. https://doi.org/10.1002/2015JD024143.

[11] Kim, E. S. (2024). Can data science achieve the ideal of evidence-based decision-making in environmental regulation?. Technology in Society, 78, 102615. https://doi.org/10.1016/j.techsoc.2024.102615

[12] Meier, Walter N., Greta K. Hovelsrud, Bob E.H. Van Oort, Jeffrey R. Key, Kit M. Kovacs, Christine Michel, Christian Haas, et al. 2014. "Arctic Sea Ice in Transformation: A Review of Recent Observed Changes and Impacts on Biology and Human Activity: ARCTIC SEA ICE: REVIEW OF RECENT CHANGES." *Reviews of Geophysics* 52 (3): 185–217. https://doi.org/10.1002/2013RG000431.

[13] Brozek, J. L., Canelo-Aybar, C., Akl, E. A., Bowen, J. M., Bucher, J., Chiu, W. A., ... GRADE Working Group. (2021). GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence—an overview in the context of health decision-making. Journal of clinical epidemiology, 129, 138-150. https://doi.org/10.1016/j.jclinepi.2020.09.018

[14] Agbehadji, I. E., Schütte, S., Masinde, M., Botai, J., Mabhaudhi, T. (2023). Climate risks resilience development: A bibliometric analysis of climate-related early warning systems in Southern Africa. Climate, 12(1), 3. https://doi.org/10.3390/cli12010003

[15] Engelbrecht, F., Adegoke, J., Bopape, M. J., Naidoo, M., Garland, R., Thatcher, M., ... Gatebe, C. (2015). Projections of rapidly rising surface temperatures over Africa under low mitigation. Environmental Research Letters, 10(8), 085004.https://doi.org/10.1088/1748-9326/10/8/085004

[16] Sreehari, E., Ghantasala, P. G. (2019). Climate changes prediction using simple linear regression. Journal of Computational and Theoretical Nanoscience, 16(2), 655-658.https://doi.org/10.1166/jctn.2019.7785

[17] Raniprima, S., Cahyadi, N., Monita, V. (2024). Rainfall prediction using random forest and decision tree algorithms. Journal of Informatics and Communication Technology (JICT), 6(1), 110-119. https://doi.org/10.52661/jict.v6i1.253

[18] Chen, Z., Wu, Q., Han, S., Zhang, J., Yang, P., Liu, X. (2022). A study on geological structure prediction based on random forest method. Artificial Intelligence in Geosciences, 3, 226-236. https://doi.org/10.1016/j.aiig.2023.01.004

[19] Fang, Z., Morales, A. B., Wang, Y., Lombardo, L. (2025). Climate change has increased rainfall-induced landslide damages in central China. International Journal of Disaster Risk Reduction, 119, 105320. 10.1016/j.ijdrr.2025.105320

[20] Abdalla, R. (2016). Evaluation of spatial analysis application for urban emergency management. SpringerPlus, 5(1), 2081. https://doi.org/10.1186/s40064-016-3723-y.

[21] Eze, E., Siegmund, A. (2024). Identifying disaster risk factors and hotspots in Africa from spatiotemporal decadal analyses using INFORM data for risk reduction and sustainable development. Sustainable Development, 32(4), 4020-4041. https://doi.org/10.1002/sd.2886.

[22] Narendr, A., Vinay, S., Aithal, B. H., Das, S. (2022). Multi-dimensional parametric coastal flood risk assessment at a regional scale using GIS. Environment, Development and Sustainability, 24(7), 9569-9597. https://doi.org/10.1007/s10668-021-01839-6?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle.

[23] Sultanad, N., Hasan, M. M. (2015). Forecasting temperature in the Coastal area of Bay of Bengal-an application of Box-Jenkins seasonal ARIMA model. J Civ Environ Res, 7(8), 149-160.https://www.iiste.org/Journals/index.php/CER/article/view/24652/0

[24] Hyndman, J. (2001). The field as here and now, not there and then. Geographical Review, 91(1-2), 262-272. https://www.tandfonline.com/doi/abs/10.1111/j.1931-0846.2001.tb00480.x

[25] Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Ghimire, S., Cerro-Prada, E., Gutierrez, P. A., ... Salcedo-Sanz, S. (2022). Machine learning regression and classification methods for fog events prediction. Atmospheric Research, 272, 106157. https://doi.org/10.1016/j.atmosres.2022.106157

[26] Alabdulmohsin, I. M., Gao, X., Zhang, X. (2014, November). Adding robustness to support vector machines against adversarial reverse engineering. In Proceedings of the 23rd ACM International conference on conference on information and knowledge management (pp. 231-240). https://doi.org/10.1145/2661829.2662047.

[27] Gündoğdu, S. (2023). Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique. Multimedia Tools and Applications, 82(22), 34163-34181. https://doi.org/10.1007/s11042-023-15165-8?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle.

[28] Rotberg,R. (2013).Africa emerges: Consummate challenges, abundant opportunities. John Wiley Sons. https://www.wiley.com/en-us/shop/generalintroductory-economics/africa-emerges-consummatechallenges-abundant-opportunities-p-9780745661636.

[29] Griffiths, J. F. 1972. "Climates of Africa." World Survey of Climatology 10.

[30] Ridzuan, F., Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. Procedia Computer Science, 161, 731-738. https://doi.org/10.1016/j.procs.2019.11.177

[31] Kaur, P., Stoltzfus, J., Yellapu, V. (2018). Descriptive statistics. International Journal of Academic Medicine, 4(1), 60-63.10.4103/IJAM.IJAM_7_18

[32] Senthilnathan, S. (2019). Usefulness of correlation analysis. Available at SSRN 3416918. https://www.researchgate.net/publication/334308527_Usefulness_of_Correlation_Analysis.

[33] Rebekic, A., Lončarić, Z., Petrović, S., Marić, S. (2015). Pearsonov ili spearmanov koeficijent korelacije-Koji koristiti?. Poljoprivreda, 21(2), 47-54. https://doi.org/10.18047/poljo.21.2.8

[34] Petrica, A. C., Stancu, S., Ghițulescu, V. (2017). Stationarity–The central concept in time series analysis. International Journal of Emerging Research in Management and Technology, 6(1), 6-16. https://doi.org/10.23956/ijermt/V6N1/107

[35] Hyndman, R. J. (2001). It's time to move from what to why'. International Journal of Forecasting, 17(1), 567-570.

[36] Johnston, Ron. 2023. Linear Regression Model." In Encyclopedia of Quality of Life and Well-Being Research, edite par Filomena Maggino, 3918–3924. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_1659.

[37] Joachims, Thorsten. 2002. "Support Vector Machines." In *Learning to Classify Text Using Support Vector Machines*, par Thorsten Joachims, 35–44. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-0907-3_3.

[38] Hayaty, N., Kurniawan, H., Rathomi, M. R., Chahyadi, F., Bettiza, M. (2023). Rainfall prediction with support vector machines: a case study in Tanjungpinang City, Indonesia. In Bio web of conferences (Vol. 70, p. 01003). EDP Sciences. https://doi.org/10.1051/bioconf/20237001003.

[39] Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Engineering Journal, 12(2), 1545-1556. https://doi.org/10.1016/j.asej.2020.11.011

[40] Guo, X., Gui, X., Xiong, H., Hu, X., Li, Y., Cui, H., ... Ma, C. (2023). Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms. Journal of Hydrology, 621, 129599.https://doi.org/10.1016/j.jhydrol.2023.129599

[41] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.https://doi.org/10.1007/BF02294361

[42] Alasow, A. A., Hamed, M. M., Rady, M., Arab, M. A., Muhammad, M. K. I., Shahid, S. (2024). Spatiotemporal analysis of soil moisture drought in the Horn of Africa.

Theoretical and Applied Climatology, 155(8), 7165-7176. https://doi.org/10.1007/s00704-024-05052-z.

[43] Gebrechorkos, S. H., Hülsmann, S., Bernhofer, C. (2019). Long-term trends in rainfall and temperature using high-resolution climate datasets in East Africa. Scientific reports, 9(1), 11376.https://www.nature.com/articles/s41598-019-47933-8

[44] Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern Africa. Reviews of Geophysics, 55(3), 590-635. https://doi.org/10.1002/2016RG000544.

[45] McPhaden, M. J., Santoso, A., Cai, W. (2020). Introduction to El Niño Southern Oscillation in a changing climate. El Niño Southern Oscillation in a changing climate, 1-19. https://doi.org/10.1002/9781119548164.ch1.

[46] Imon, A. R., Roy, M. C., Bhattacharjee, S. K. (2012). Prediction of rainfall using logistic regression. Pakistan Journal of Statistics and Operation Research, 655-667. https://doi.org/10.1234/pjsor.v8i3.535

[47] Yin, G., Yoshikane, T., Yamamoto, K., Kubota, T., Yoshimura, K. (2022). A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. Journal of Hydrology, 612, 128125. https://doi.org/10.1016/j.jhydrol.2022.128125

[48] Irmanda, H. N., Ermatita, E., bin Awang, M. K., Adrezo, M. (2024). Enhancing weather prediction models through the application of random forest method and chi-square feature selection. JOIV: International Journal on Informatics Visualization, 8(3-2), 1506-1514. https://dx.doi.org/10.62527/joiv.8.3-2.2356