**UNIVERSITY OF NAIROBI**

**DROUGHT FORECASTING IN KENYA USING MACHINE LEARNING**

**BY**

**OMENY MERCY ACHIENG**

**I10/1862/2021**

**A research project proposal submitted in partial fulfilment for the award of degree of Bachelor of Science in Meteorology**

**DECLARATION**

This proposal is my original work and has not been presented for a degree in any other university under any program for examination

Signature ………………………………… Date …………………………………………

OMENY MERCY ACHIENG

I10/1862/2021

Department of Earth and Climate Sciences, Meteorology Unit

University of Nairobi

[mercyaomeny@students.uonbi.ac.ke](mailto:mercyaomeny@students.uonbi.ac.ke)

This proposal work has been submitted for examination with our proposal as university supervisors.

Signature ………………………………… Date …………………………………………

DR W. GITAU

Department of Earth and Climate Sciences, Meteorology Unit

University of Nairobi

Signature ………………………………… Date …………………………………………

PROF G. OUMA

Department of Earth and Climate Sciences, Meteorology Unit

University of Nairobi

**ACKNOWLEDGEMENT**

I would like to thank my supervisor, Dr. Gitau, for his value addition into this research work through his guidance and support in the research. His immense insight and encouragement have borne fruit in the shaping of my work. I also do appreciate my family and friends for the continued motivation and sustained support which kept me engaged and inspired. Last but not least, I also want to thank Dr. Gudoshava for the spark of interest that he gave me in the area of modelling and forecasting, which has driven me into the journey so far through academics. Your contribution is what has made this research possible, and I feel truly grateful.

**ABSTRACT**

Drought is one of the most severe natural hazards impacting many regions but is more pronounced in arid and semi-arid regions like Kenya. It adversely affects agricultural productivity, water resources and overall socio-economic stability. The increasing frequency of extreme climate conditions due to climate change has resulted in significant economic and political implications, making climate change a universal concern in terms of risk management. Climate change acts as risk multiplier increasing the likelihood of vulnerabilities and intensifying the impacts of extreme weather conditions.

This research aims to develop a machine learning-based forecasting model to assess and predict the impacts of drought using these data sets; raingauge data, maximum and minimum temperature and sea surface temperature (SST). Maximum and minimum temperature data provides an insight into level of evaporation and transpiration while raingauge data provides essential insights into precipitation patterns. Understanding the drivers of drought is crucial for developing effective mitigation strategies and improving resilience to drought impacts.

This study will employ machine learning algorithms to analyse the relationship between these variables and drought occurrences. By utilizing historical data spanning multiple years, the research will establish predictive models capable of forecasting drought conditions and assessing their potential impacts on agriculture and water resources.

Data sources include meteorological stations for raingauge data and maximum and minimum temperatures and oceanography data for SST. The analysis will incorporate advanced statistical methods and machine learning techniques to enhance predictive accuracy and reliability. Expected outcomes include a comprehensive understanding of drought dynamics in Kenya and a set of predictive models that can assist in mitigating the adverse impacts of drought on food security and water availability.

**List of Acronyms**

ASALs Arid and Semi-Arid Lands

CMIP5

CMIP6

ENSO El Niño Southern Oscillation

IOD Indian Ocean Dipole

ITCZ Inter Tropical Convergence Zone

NDVI Normalized Difference Vegetation Index

PSDI Palmers Standardised Drought Index

SPEI Standard Precipitation and Evaporation Index

SPI Standardized Precipitation Index

SST Sea Surface Temperature

CHAPTER ONE

**INTRODUCTION**

**BACKGROUND INFORMATION**

Drought is a prevalent, complex natural hazard with ever increasing impacts on water security, crop productivity and economic stability over larger portions of the world. Drought is a multifaceted phenomenon that can be classified into four main types based on its impacts and cause. There in Meteorological drought which is a prolonged period of below average precipitation often measured relative to historical averages for a specific region, Agricultural drought, hydrological drought and socio-economic drought. This research will primarily focus on meteorological drought. It is selected because it serves as the foundation for identifying drought onset and intensity through rainfall and temperature patterns. As a result of climate change the precipitation patterns have changed while temperature variability is enhancing; hence the frequency, intensity and duration of drought events have increased especially within the vulnerable areas of the world.

Several drought events have been experienced in Kenya ranging from mild ones to extreme severe. On of the worst droughts in Kenyan history occurred in in 1984 to1985. It resulted in widespread crop failure, livestock losses and humanitarian crisis. Another severe drought occurred from 2009 to 2011. It was part of the severe East African drought that affected the horn of Africa with Kenya being one of the hardest hit countries. A combination of delayed and erratic rains led to the failure of both long and short rains. Both drought occurrences severely affected food security and water availability with millions of people being affected and productivity dropped by up to 40% is some areas.

Previous studies on drought have explored various aspects, focusing on its onset, severity, duration, frequency, and impacts, as well as the underlying drivers. Accurate drought prediction faces enormous challenges despite advanced approaches being considered. Data quality and inconsistencies or incompleteness in the datasets, as well as limitations in spatial resolution, temporal resolution, stand in view of this prediction. Complexity is such that there would be a lot of processes and events which are unforeseeable, such as the El Niño–La Niña. Notably, climate change already acts to complicate matters when it comes to foretell the patterns of the drought. Drought modeling also faces serious challenges to fully represent the physical processes of soil moisture deficits, and most of them even fail in predicting regional droughts because of regional variability. In addition, socioeconomic factors, such as land use changes and slow responses to early warnings, and model limitations involving overfitting and difficulties in model integration further worsen the prediction accuracy. These challenges have given a clear indication of the need for better data collection, integration of models, and further understanding of environmental and human factors.

On the global scale, meteorological drought forecasting has been enhanced through the use of indices such as SPI and SPEI, coupled with climate models like CMIP5 and CMIP6, to project future drought frequency and intensity based on climate change scenarios. In Africa, studies have emphasized increased drought severity in areas such as the Sahel and East Africa, caused by reduced precipitation and increased evapotranspiration. Some Kenya-specific studies highlight the increase in episodes of drought, especially within arid and semi-arid areas, based on the SPI and using satellite-derived datasets such as NDVI and soil moisture. Integrated approaches, including those using machine learning, allow for improvement in forecast accuracy by integrating climatic and hydrological data.

In the sections that follow, we go into details about the objectives of our research, the methodology used and expected outputs. We set into context how our proposed approach enhances and augments the existing knowledge in drought forecasting. This will lead to the development of a more robust and adaptable tool for predicting drought impacts, thereby contributing o better preparedness for and resilience against droughts due to climate change.

**AREA OF STUDY**

The topography of Kenya is diverse and include a mixture of highlands, valleys, plateaus and coastal plains. The country has extensive highland areas. The great rift valley is a prominent feature running from the red sea passing through Kenya down to Mozambique. The region also features significant bodies of water like Lake Victoria, Indian ocean and more. This combination of highlands, valleys, coastal areas and water bodies make Kenya one of the most varied on the continent. (Cabot Venton et al,2012)

**Climatology of Kenya**

The climate of Kenya is dominated by large rainfall and temperature variabilities, given the country's complex geography. Essentially, the topography and ITCZ determine the rainfall patterns; this leads to a bimodal distribution of the phenomenon over most areas. The highest annual rainfall amount over the country is recorded along the coastal and highland areas while the north and east have semiarid to arid conditions with rains falling less frequently (Ogallo, 1988). The pattern in Kenya's temperature follows altitude; the coasts are generally hot, with temperatures ranging between 25-30°C, while the highlands, in which Nairobi is included, are cool with average temperatures ranging from 15-20°C due to their higher elevation. These fluctuations in temperature and rainfall have serious implications for agriculture and the management of water resources throughout the country.

**STATEMENT OF THE PROBLEM**

Current methods of predicting drought are usually quite limited in their precision and timing. This provides communities and decision makers with limited information to prepare and mitigate these impacts effectively. Traditional drought forecasting methods are usually based on precedents of weather patterns or single indicator analyses; neither captures the variety of interacting environmental factors involved in the onset and severity of drought.

The study attempts to fill that gap with the development of a machine learning based model for drought forecasting combines with rainfall data, maximum and minimum temperature data and sea surface temperature. Using such diverse variables, the model aims at improving accuracy and lead time in the prediction of drought within Kenyan boarders. Only this will allow for more reliable forecasting of drought impacts, proving vital information for farmers, policymakers and disaster management agencies to take appropriate precautions and resource allocations to socioeconomic damages from droughts

**RESEARCH QUESTIONS**

1. What is the historical relationship between maximum and minimum temperature, rainfall and Sea surface temperature patterns across Kenya and how do these inform drought forecasting?
2. How can machine learning models be developed and optimized to predict drought conditions across Kenya’s diverse agro-ecological zones?
3. What is the performance and accuracy of the developed drought forecasting model when evaluated across different spatial and temporal scales in Kenya

**OBJECTIVE OF THE STUDY**

**Overall Objective of The Study**

To develop and validate a machine learning-based drought forecasting framework for Kenya that integrates rainfall data, maximum and minimum rainfall data and sea surface temperature to predict drought conditions across different agro-ecological zones with improved accuracy and lead time.

**Specific Objectives of The Study**

1. To analyse the historic correlation between maximum and minimum temperature, rainfall and SST across Kenya, using time series data to identify patterns and trends.
2. Develop a machine learning model to predict drought conditions based on climate data across Kenya’s diverse agro-ecological zones.
3. Asses the performance and accuracy of the developed drought forecasting models by comparing model predictions with observed drought events.

**JUSTIFICATION OF THE STUDY**

Drought poses a significant threat In Kenya with significant repercussions on agriculture, water resources and socio-economic stability. The increasing frequency of extreme climate conditions due climate change has resulted in significant economic and political implications making climate change a universal concern in terms of risk management. This research aims to explore indicators, triggers, thresholds and develop a machine learning model that contributes towards timely, accurate and actionable forecasts in order to mitigate the effects.

**SIGNIFICANCE OF THE STUDY**

The significance of the study lies in it’s potential to advance drought forecasting capabilities, integrate multi-source data and generate insights tailored to Kenya’s agro-ecological context-ultimately supporting climate change adaptation and mitigation efforts.

**CHAPTER TWO**

**LITERATURE REVIEW**

Drought forecasting is a very important area of study, especially with the increases in frequency and severity of droughts, particularly in regions such as Kenya, which bases the majority of its economy on agriculture and livestock. Most of the traditional methods employed in forecasting rely purely on historical data and simple statistical models that are ill-suited to capture some of the complexities of drought. Recent development in machine learning begets sophisticated means of enhancing drought prediction accuracy. Included in these are the Random Forest algorithms. Random Forest is a machine learning algorithm Bierman 2001 developed that had more often than not been applied to many different predictive tasks with considerable success in environmental monitoring. It is an ensemble method that constructs multiple decision trees during training and outputs the mode of their predictions, hence helping to reduce overfitting and improving accuracy. Random Forest has a foresight advantage in drought forecasting since it can handle big datasets involving a large number of variables, hence suitable for integrating many environmental indicators such as rainfall and temperature as well as streamflow. Some papers illustrated the performance of Random Forest in drought forecast and compared its outcomes with the performance of traditional statistical modelling techniques. For instance, Ahmed et al. (2017) presents the performance of Random Forest models that outperformed regression models when predicting droughts based on climate and remote sensing data over arid regions. Random Forest was also applied to drought forecasting over China by Wang et al. (2019), who demonstrated that it can accept multiple variables from, among others, the SST and account for nonlinear relationships, typical in environmental data. Streamflow, raingauge data, and SST were critical variables for drought monitoring in Kenya. Raingauge data represents accurate local measurements of precipitation as a basic indicator of drought declaration. Meanwhile, SST, particularly of the Indian Ocean, is known to be associated with rainfall variability over East Africa through phenomena such as the Indian Ocean Dipole. Combining such variables within a random forest model allows for more comprehensive drought forecasting. Few studies have employed Random Forest for drought forecasting over Kenya, but the few report its high potential. Ndung'u et al. (2020), for example, employ machine learning, including Random Forest, in the forecasting of droughts in the semi-arid regions of Kenya. They note that the addition of climate variation variables enhances the accuracy of their forecast and hence enables appropriate resource management while experiencing drought conditions. The success of such models underlines the need for further research to customize those methods against specific climatic conditions in Kenya. Random Forest proves to be a promising approach towards improving the drought forecast in Kenya by integrating multiple data sources like streamflow, raingauge data, and SST. These variables, coupled with the non-sensitivity of the algorithm to complex and nonlinear relationships, will enable the model to make more realistic and timely predictions to support decision-making in drought preparedness. In fact, studies have shown that machine learning models, especially Random Forest, ensure high outperformance in conventional methods of forecasting, which ultimately is beneficial for a region like Kenya prone to frequent occurrences of this condition.

**DROUGHT IN KENYA**

Kenya encompasses 47 counties. Some of these counties frequently experience droughts that lead to severe food insecurity, water scarcity and economic losses. The region is characterized by highly variable climate with some areas receiving less than 200 mm of rainfall annually, while other receive up to 1200mm. The regions climate is influenced by several factors including the intertropical Convergence Zone (ITCZ), the El Nino Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) These climatic drivers contribute to the regions high susceptibility to drought, particularly in the arid and semi-arid lands (ASALs) that dominate much of the landscape. The region’s dependence on rainfed agriculture accelerates the impacts of droughts. This dependency combined with high population growth, poor infrastructure and limited access to clean and usable water, has led to frequent and severe drought that have devastating impacts on food security and livelihoods. The need for effective drought forecasting and early warning systems in the region is therefore critical to mitigating and enhancing resilience to climate variability,

**INDICATORS OF DROUGHT**

Drought forecasting requires the integration of multiple indicators that provide information on different thresholds of drought. There indicators can be classified into meteorological, agricultural, hydrological and socio-economic categories. Each indicator plays a critical role in understanding the onset intensity and impact of drought.

Meteorological indicators are the first to signal the potential onset of a drought. They are crucial in for assessing the initial stages of drought and setting early warning triggers. The common meteorological indicators are Precipitation deficit and Temperature anomalies. There are also set drought indexes and Precipitation indexes.

Agricultural indicators are essential for assessing the impact of drought on crop health and yield. These indicators help in setting thresholds that can trigger drought alerts for the agricultural sector Examples of agricultural indicators are Soil moisture level, Normalized Difference Vegetation Index (NDVI) among others.

Hydrological indicators provide insight into the availability of water resources, including river flows, reservoir levels and ground water Hydrological indicators include: Stream flow levels, Reservoir Storage level and ground water storage.

Socio-economic indicators asses the boarder impacts of drought on communities including food prices, Water Usage and Migration patterns. They are essential for understanding the multifaceted impacts of drought on human communities.

**DROUGHT TRIGGERS IN KENYA**

Similar to most other East African countries, Kenya is one of the driest countries in the world. The development of efficient forecasting models in this area therefore requires a proper understanding of the various causes of drought. The different causes of drought conditions in Kenya are intricately linked at both the local and global scale. Some of the key triggers of drought in Kenya include the following:

1. Climatic Variation and El Niño-Southern Oscillation (ENSO)

Particularly, the ENSO has a principal contribution to rainfall conditions and the incidence of drought in Kenya. El Nino events attract above-normal rainfall in East Africa while La Nina events attribute to drought-like conditions. Nicholson & Kim, 1997. In addition, the Indian Ocean Dipole controls, to a large extent, the processes influencing rainfall in Kenya. In general, positive phases are associated with above-average rainfall while negative phases are associated with reduced rainfall. Saji et al., 1999.

2. Rainfall Patterns are Extremely Irregular

The country experiences two rainy seasons in a year: from March to May, known as the "long rains", and from October to December, nicknamed the "short rains". Drought is often declared when either one or both of these rainy seasons fail. Climate change alters this reliability of seasonal and intensity of rain, hence increasing the frequency of a drought event and flood events.

3. Land Use Changes and Deforestation

Human activities, especially deforestation and change in land use, predisposes Kenya to drought vulnerability. For a fact, the removal of natural vegetation reduces the capacity for soils to retain moisture and increases surface runoff, thus enhancing the effects of reduced rainfall. This has been evident through changes in local climate patterns associated with deforestation occurring in such areas as the Mau Forest complex.

4. Global Climate Change

Long-term climate change is altering temperature and precipitation patterns across East Africa. Increased temperatures raise the rate of evaporation, diminishing the soil moisture and surface water availability. Based on generalized global temperature rises, climate models predict that the frequency and intensity of drought will increase across much of East Africa, including Kenya.

5. Topography and Regional Climate Zones

Topography in Kenya varies from coastal lowlands through high mountain ranges, giving quite a varied climate with different susceptibilities to drought. Accounting for about 80% of the land area, ASALs are naturally very low and erratic in rainfall, thus making them particularly prone to drought conditions.

6. Ocean-Atmosphere Interaction

Sea-surface temperatures of both the Indian and Atlantic Oceans have a great influence on rainfall patterns in East Africa. The SST anomalies are responsible for the changes in atmospheric flow patterns, impacting moisture transport and subsequent precipitation over Kenya (Behera et al., 2005).

7. Atmospheric Circulation Patterns

Some of the influence on rainfall distribution in Kenya includes movement of ITCZ and strength of the East African Low-Level Jet; their anomalies lead to drought conditions.

These triggers are very important in the development of accurate drought forecasting models for Kenya. By incorporating such data as NDVI for vegetation cover, rain gauge data for local precipitation patterns, and sea surface temperature for ENSO and IOD effects, our machine learning approach desires to capture complex interactions that yield either drought or favourable conditions in Kenya. This shall yield a deeper understanding that can be used in making more correct predictions and better-informed decisions concerning drought management.