OPTIMIZING HEALTHCARE RESILIENCE THROUGH CLIMATE-INFORMED DATA ANALYTICS.

JARED MURUNDU

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE AND IT IN THE SCHOOL OF MATHEMATICS AND COMPUTING IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE OF THE COOPERATIVE UNIVERSITY OF KENYA

2025

DECLARATION

Declaration by the Candidate

This Project is my original work and has not been presented for award of a degree in any other University or for another award

………………   ……………18/11/2025………………

Signature   Date:

Student Name: JARED MURUNDU

Student Registration Number: MDATC01/6067/2022

Declaration by the supervisors

We confirm that the work reported in this Project was carried out by the candidate under our supervision and has been submitted with our approval as university supervisors

…………………………   ………18/11/2025……………………

Signature   Date:

Dr. Fidelis Mukudi, PhD

Department of Mathematical Sciences

School of Computing and Mathematics

The Co-operative University of Kenya

…………………………   ……………18/11/2025……………

Signature   Date:

Dr. Ronald Ojino, PhD

School of Science and Technology

Open University of Kenya

DEDICATION

I dedicate this to my parents Enos Khaweri and Esther Andayi, siblings and Blessy. In memory of Naftali Indeche, whose faith and counsel inspired me.

ACKNOWLEDGMENT

I am deeply grateful to my supervisors, Dr. Ronald Ojino and Dr. Fidelis Mukudi, whose mentorship, support, and guidance have been instrumental to the direction and success of this project. I also acknowledge Migori County for providing the vital data that formed the backbone of this research, and for the cooperation of local authorities during data collection. My heartfelt appreciation goes to my parents, Enos Khaweri and Esther Andayi, thank you for your unconditional love, encouragement, and constant support throughout my academic journey. To my siblings Rodgers, Roselyne, and Brian I am truly grateful for the inspiration, motivation, and unwavering belief you have always given me. Finally, I extend my gratitude to The Co-operative University of Kenya, School of Computing and Mathematics, for offering me the platform to advance my academic skills.

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

UNFCCC	United Nations Framework Convention on Climate Change
KMD	Kenya Meteorological Department
AUC ROC	Area Under the Receiver Operating Characteristic Curve
GBM	Gradient Boosting Machine
ANN	Artificial Neural Network

DEFINITION OF TERMS

Climate Data Analytics	The process of collecting, analyzing, and interpreting climate-related data, such as temperature, humidity, and rainfall, to identify patterns and trends that can influence healthcare outcomes.
Healthcare Resilience	The ability of healthcare systems, institutions, and personnel to maintain essential healthcare services during climate-induced challenges, such as disease outbreaks, natural disasters, and environmental disruptions. This study aims to enhance healthcare resilience through the integration of climate data analytics.
Ensemble Machine Learning Model	A predictive model that combines multiple algorithms to improve accuracy and reliability in forecasting outcomes. In this study, an ensemble machine-learning model will be developed to predict disease outbreaks by incorporating climate and health data.
Waterborne Diseases	Contaminated water diseases such as cholera and typhoid. The study will predict and manage the impact of waterborne diseases affected by climate variability.
Respiratory Diseases	There are serious concerns pertaining to lung and airway diseases like asthma, pneumonia and chronic obstructive pulmonary disease (COPD). In this research, we will study how the spread of respiratory diseases is influenced by the factors of climate change, such as fluctuations in air quality and temperature.

Predictive Modeling	The process of using statistical techniques and algorithms to forecast future events based on historical data. In this study, predictive modeling will be employed to anticipate disease outbreaks driven by climate variability.
Data-Driven Decision Making	A decision-making process that relies on data analysis and interpretation to guide actions and strategies. In this context, healthcare decisions will be informed by climate data and predictive models to enhance resource allocation and outbreak management.
Disease Incidence	The occurrence of new cases of a particular disease within a specific period. This study will measure the incidence of malaria and respiratory diseases in response to changing climate conditions.
Healthcare Resource Allocation	The distribution of healthcare resources, such as medical supplies, personnel, and infrastructure, to respond to disease outbreaks. This study will explore how predictive models can optimize resource allocation based on climate data forecasts.
Machine Learning	A subset of artificial intelligence (AI) that allows computer systems to learn from data and improve predictions or decisions without being explicitly programmed.

ABSTRACT

Climate variability is a major driver of malaria surges in Kenya's Lake Victoria Basin, yet routine decision-making often reacts after peaks have begun. This project develops, validates, and operationalizes an ensemble machine-learning framework that integrates climate and health-system data to provide short-lead malaria early warning for Migori County. The study population comprised public health facilities in Migori; a purposive census of ten facilities yielded 5,000 facility-week records (2015–2024). Secondary data were extracted from routine surveillance (weekly malaria cases and facility capacity indicators) and matched to weekly climate series (rainfall, temperature, humidity, wind). Data collection used standardized extraction templates; instruments for processing and analysis were reproducible Python notebooks. Pre-processing created short lags (t–1, t–2), harmonic calendar terms, and facility fixed effects; blocked time-series splits and rolling-origin cross-validation were used to avoid leakage. Three base learners Random Forest, XGBoost, and a feed-forward ANN were trained and stacked via a Ridge meta-learner; complementary Negative Binomial regression supported inference. Key results show the ensemble outperformed single models on the independent test set (R^2 = 0.75; RMSE = 2.87; MAE = 2.22), with stable ROCV performance across seasons (R^2 = 0.72–0.76; RMSE = 2.98–3.27). Interpretability (permutation importance, SHAP, PDP, ICE) confirmed recent rainfall, seasonal terms, and recent cases as dominant drivers, while personnel density and antimalarial stock index dampened predicted surges. The study recommends institutionalizing a dual-trigger SOP flag an alert when weekly rainfall ≥120 mm and ensemble risk ≥0.7 to cue stock pre-positioning, surge rosters, and targeted outreach; integrating near-real-time climate feeds into dashboards/SMS; strengthening data quality pipelines; capacity building on model interpretation; and periodic threshold recalibration. The framework demonstrates that climate-informed analytics can be embedded in routine governance to shift malaria control from reactive response to anticipatory preparedness in resource-constrained settings

CHAPTER ONE

1 INTRODUCTION

The chapter presents the general project introduction and the background of the study. Gives information of the study area and a brief highlight on the climate data analytics and their utilization in the healthcare sector, the problem statement, research objectives, the research questions, Significance of the study scope of the study and limitation the study.

1.1 Background Study

Climate refers to the average weather conditions in a specific area over an extended period, typically at least 30 years (Walsh et al., 2020) . Weather patterns develop from multiple elements which include geographic position and altitude together with the location adjacent to major water features and mountain features. Natural ecosystems together with agricultural systems and human settlements with their economic strategies experience significant influence from climate conditions.

The United Nations Framework Convention on Climate Change (UNFCCC) defines climate change through the term a change of climate attributed directly or indirectly to human activity that alters the composition of the global atmosphere (Usman et al., 2024). Natural Earth processes and human activities through fossil fuel use and forest clearing and industrial operations drive these phenomena (Ahmed, 2020). Scientists identify climate change as among the leading challenges worldwide during the twenty-first century which impacts countries at all levels of development (Grinin et al., 2021). The environment along with human societies faces adjustment due to rising oceans and modified rainfall distribution and temperature elevation. The modifications in environmental conditions produce major health consequences for worldwide populations because they enable the development and dissemination of waterborne diseases along with respiratory diseases.

1.1.2 Health Impacts of Climate Change

The combination of severe weather events including flooding and hurricanes and drought conditions weakens water quality which enables pathogens to enter the water supply to cause diseases including cholera and diarrhea and typhoid. The lack of clean water supply and proper sanitation facilities allows infectious diseases to thrive especially in developing nations (Santos et al., 2021).

Moreover, air pollution from fossil fuel burning together with temperature rise enhances the occurrence of respiratory diseases including lung cancer and asthma and cardiovascular diseases (Tran et al., 2023a). Statistics show that air pollution leads to yearly deaths amounting to 7 million (Santos et al., 2021). Healthcare systems need to adapt to climate change because rising health hazards demonstrate this requirement.

1.1.3 Climate Change in Africa

The impacts of climate change hit Africa because it depends heavily on agricultural and water resource sectors which are sensitive to climate variations. Pursuant to the Intergovernmental Panel on Climate Change (IPCC, 2022) reports Sub-Saharan Africa reports more frequent extreme weather occurrences which generate deteriorating health results and deterioration of water resources and food security decline.

The health effects in Africa become severe because temperature increases assist in spreading vector-borne illnesses such as malaria and dengue and floods and droughts result in the outbreaks of waterborne diseases like cholera. The healthcare systems in numerous African nations present structural deficiencies because they operate with limited funding and inadequate resources to address climate-related health problems (Ebi & Hess, 2020).

1.2 Healthcare Resilience in the Face of Climate Change

1.2.1 Global Health Discussions

Healthcare resilience gains worldwide appreciation because climate change worsens public health problems which severely test health delivery systems around the world (Ansah et al., 2024). Healthcare systems together with healthcare institutions and professionals achieve healthcare resilience when they can maintain continuous safe and effective care delivery during catastrophes together with disasters and failures of healthcare infrastructure (Narwal & Jain, 2021). Healthcare systems require enhanced resilience because climate-related disruptions occur more frequently and produce more severe effects that include extreme weather events and pandemics. Healthcare must preserve its effectiveness and maintain services as health problems advance because this protects public well-being especially for marginalized populations (Rami et al., 2023).

The healthcare sector throughout the world faces two simultaneous issues. Healthcare organizations must initially handle rising healthcare needs originating from natural disasters linked to climate change (Smith et al., 2022). The rise in both severe weather episodes and their frequency leads hospitals to deal with immediate health consequences. The combination of floods and heatwaves leads to cholera outbreaks and generates more cases of respiratory along with cardiovascular illnesses (Soomro et al., 2025). Rising global temperatures have produced an expanding environment suitable for malaria and dengue fever spread which adds excessive strain on healthcare systems according to (Anikeeva et al., 2024). A study by (Gutterman, 2024) healthcare institutions encounter multiple crises when climate-related health emergencies grow worse because they lack adequate capabilities to care for disaster-related emergencies and standard medical patients . Third world nations experience an intensified compound health disaster from limited facilities and scarce resources which leads to decreased delivery of healthcare while creating poor medical results (Phelan et al., 2022).

Healthcare facilities worldwide produce large amounts of greenhouse gases which worsen climate change conditions. (Ebi et al., 2021) demonstrate the healthcare sector produces about 4.4% of worldwide emissions. The high emission levels stem from the power-intensive operations within hospitals as well as medical facilities and healthcare logistics systems that involve the production cycle of medicines from manufacturing through distribution to disposal (Schwab et al., 2025). Hospitals gain their power from fossil fuels, yet their operation creates excessive single-use plastics waste that results in harmful emissions from medical waste disposal procedures. Hospital emissions produce air pollution whereas global warming develops from such emissions thus creating new health challenges that healthcare facilities try to solve (Sapuan et al., 2022). Healthcare facilities face an opposing challenge because they need to resolve climate-related health risks while their operations add to environmental degradation.

International entities have started multiple initiatives to decrease healthcare systems' environmental impact while building their resilience capabilities (Haldane et al., 2021). The World Health Organization (WHO) together with the United Nations work to establish sustainable healthcare practices on an international level. Through its GGHH program the organization works to improve healthcare sustainability by encouraging facilities to adopt renewable energy usage while reducing waste and improving resource efficiency (Ng et al., 2025). Medical institutions worldwide are adopting alternative renewable power systems based on solar and wind technology to reduce excessive fossil fuel usage and decrease their carbon emissions. Healthcare facilities implement sustainable procurement methods to buy environmentally friendly medical products which reduce their total environmental influence (Soares et al., 2023).

Healthcare systems need to focus strongly on implementing climate resilience strategies as their main emphasis. The approach requires healthcare facilities to reduce their carbon emissions together with preparing them to handle and adapt to climate-related disasters (Vallée, 2024). The healthcare sector establishes programs for risk evaluations and infrastructure improvements and early warning alerting systems to enhance medical facility readiness for natural disasters including floods storms and heat waves. Healthcare systems today are becoming more climate-smart because they integrate environmental footprint reduction approaches with operational sustainability during times of climatic emergencies.

1.2.2 The Role of Climate-Driven Data Analytics

Healthcare systems gain better resilience against climate change through implementation of analytics derived from climate data. Experts use advanced data analysis techniques to process climate data including temperature measures with precipitation data alongside greenhouse gas observations to forecast health risks associated with climate change (Bhardwaj & Khaiter, 2023). Massive, complicated datasets allow healthcare systems to discover the causes between environmental changes and disease outbreaks along with other health problems. Through climate-driven data analytics decision-makers obtain better forecasts of health threats which enables them to implement preventive strategies (Asaaga et al., 2024).

The analysis of healthcare data through data analytics reveals patterns along with trends which would escape traditional analytic methods. The established healthcare decision-making process applies three framework methods, expert judgment and basic statistical analysis together with linear regression models according to a study by (Hsu et al., 2021). The methods use historical data only while ignoring complex climate variables leading to basic analytical results. The spread of malaria alongside dengue fever can be predicted by following changes in weather patterns such as precipitation and temperature (Kulkarni et al., 2022). Warmer temperatures and rains favor mosquito reproduction, increasing transmission rates. The prediction of epidemics combined with distribution of medical supplies and staff and vaccinations to high-risk locations becomes possible when public health officials analyze these climate characteristics. Rapid intervention strategies serve to reduce diseases which emerge from climate changes.

The application of climate-based analytical methods improves the control of infectious diseases because these conditions are now linked to intense weather events that include floods and droughts. Water contamination leads to the development of typhoid and cholera infections after heavy rainfall and flooding incidents. The occurrence of droughts produces two harmful effects on water availability and safety conditions. Analytics platforms use predictions to determine upcoming water-related health crises in particular regions which helps direct public health prevention efforts through water purification and sanitation campaigns before they become severe threats (Balogun et al., 2020).

Data analytics driven by climate information enable healthcare authorities to better understand how pollution from air quality worsens due to changing temperatures without limiting analysis to disease evaluation (Mahule et al., 2024). Global temperature rise generates more airborne pollution and results in the development of asthma and COPD. Data analytics tools allow healthcare systems to spot areas with high risk while guiding basic demographic groups including children and seniors and people with medical issues through pollution reduction methods to decrease their vulnerability. Healthcare prevention strategies together with emergency response activities benefit from climate-based analytical data approaches.

Machine learning (ML) methods based on random forests and deep learning models enable researchers to build predictive models that unite climate and health data for better policymaker support in creating climate-adaptive health systems and infrastructure planning within vulnerable climate regions. Decision trees assembled into random forests work with difficult datasets, yet these models remain hard for policymakers to interpret hence the mechanism of decision-making stays obscure (Rana & Varshney, 2021). The learning models demonstrate preference toward dominant class members when processing datasets which are unbalanced thus neglecting rare health outcomes. Robustness was assessed using rolling-origin cross-validation (ROCV) with an expanding training window, which mimics real deployment where future projections are purely dependent on historical data. From 2015 to 2024, performance was constant across five temporal folds: MAE = 2.30–2.51, RMSE = 2.98–3.27, and R2 = 0.731–0.755 (664 evaluation rows per fold). The consistency of these metrics, along with residuals that did not exhibit systematic drift, indicated that there was minimal temporal overfitting and that generalisation across seasons, including exceptionally wet or dry ones, was reliable.

Through strategies that assist leaders in anticipating health emergencies brought on by climate change and creating preparedness responses for them, the climate-driven data analytics framework strengthens healthcare resilience (Hussain et al., 2025). When health outcome data and climate data are combined, healthcare organisations are better equipped to safeguard public health. For integration to be achieved, Geographic Information Systems (GIS), remote sensing technologies, and statistical modelling tools are crucial current analytics tools.

For instance, GIS technologies apply geographic distribution data to demonstrate how disease outbreaks correspond with climatic factors for resource distribution purposes (Chandran & Roy, 2024). The collection of real-time environmental data consisting of temperature and precipitation levels becomes possible through remote sensing technologies for disease outbreak prediction. These tools encounter technical barriers that prevent successful integration of real-time information and process large volumes of data which hinders prompt decision-making process. Statistical models base their predictions on basic assumptions regarding climate-health relationships which result in inaccurate forecasting outcomes. The current analytical solutions need modern advancements and integrated capabilities to efficiently handle complex climate-related health problems.

1.3 Climate Change and Health in Kenya

1.3.1 Health Impacts in Kenya

An increasing number of health problems related to climate change motive changes in waterborne and respiratory infections throughout Kenya. Nationwide water quality and sanitation systems deteriorate because precipitation patterns are becoming less predictable, and the country endures more frequent droughts and floods. The altered access to water resources leads to outbreaks of cholera and typhoid infections because these pathogens grow better when people lack clean drinking water (Noureen et al., 2022). Health risks intensify from flooding because it enables pathogens to spread through polluted water and exceeds processing capacity of treatment centers and produces crowded and unhygienic settlements that promote disease transmission. Factors linked to climate changes demonstrate the necessity to implement climate data into healthcare decision-making processes to reduce health impacts.

Furthermore, Climate change alters allergen transport patterns in ways that boost respiratory illnesses together with allergic reactions in patients. The changes in climate temperatures along with moisture levels drive allergenic plants to reproduce more frequently and spread wider leading to longer and severe allergy seasons (D’Amato et al., 2020). Population groups who have not typically experienced severe allergies now face new respiratory health threats because of the existing respiratory disease deterioration. The healthcare system of Kenya faces increased pressure because climate-induced health challenges are growing more common and severe. The healthcare facilities in rural areas and underserved regions face increased stress because residents lack sufficient access to both medical services and emergency care that delays their ability to receive proper and immediate treatment (Howard, 2021).

1.3.2 The Role of Data Analytics in Disease Management

Through data analytic methods with focus on ensemble machine learning systems Kenya can greatly enhance its approach to climate disease management. Public health authorities use geographical, climate and spatial data analysis to predict waterborne and respiratory disease outbreaks which help them distribute their resources better (Graw et al., 2021).

Researchers successfully use disease incidence data together with meteorological information to predict infectious disease outbreaks particularly during cholera seasons (Nusrat et al., 2022). A single machine learning algorithm implementation produces various operational issues. The binary classification capabilities of logistic regression succeed well yet problems arise when the program oversimplifies complex data connections that result in prediction inaccuracies. The performance of support vector machines (SVMs) on infectious disease data depends on parameter selection and noise levels because both factors can influence their operation (Neto et al., 2022). Additional comprehensive methods have become necessary because of these existing constraints. Random forest and gradient boosting provide ensemble machine learning methods which unite different algorithms to enhance predictive accuracy thus improving outbreak forecasting abilities (Alghanmi et al., 2022).

Modern AI models apply within Kenyan healthcare for managing climate-related health threats which primarily impact both waterborne and respiratory health outcomes. These models succeed at processing complex non-linear system relationships in large datasets yet need massive computational resources together with large, labeled datasets that might be difficult to acquire in limited-resources environments. The unexplained functioning of deep learning systems leads to difficult interpretation which restricts the conversion of research outcomes into effective public health responses (Kenny, 2014). A research examination of respiratory health in Africa combines AI with specific recommendations about using AI for future healthcare improvements while requiring additional planner investment. The effective enhancement of healthcare resilience to climate-induced health threats in Kenya requires addressing current challenges through capacity building and model optimization approaches (Sesay & Osborne, 2025).

This study created predictive models while evaluating disease outbreak relations to climate which offered practical solutions to enhance Kenya's public health responses. Research findings will add to existing knowledge on climate change and public health in Africa while providing both governmental officials and healthcare experts tools to develop strong healthcare systems for climate-related health issues.

1.4 Problem statement

Climate change is intensifying climate-sensitive diseases, straining healthcare systems that still rely on reactive management approaches which rarely incorporate climate variables (Ansah et al., 2024). Evidence from a study by (Asaaga et al., 2024) shows that integrating climate data analytics (CDA) can improve early warning and preparedness, yet practical uptake in routine healthcare decision-making remains limited. This creates a knowledge gap on how CDA can be operationalized within county health systems to inform timely resource allocation, emergency readiness, and protection of vulnerable populations. In Kenya specifically Migori County health managers lack validated, locally tuned tools that translate climate signals into actionable forecasts for service planning according to ("Proceedings of the 16th Annual Conference on the Science of Dissemination and Implementation in Health," 2024). This constitutes a contextual gap, as international findings are not directly transferable to the local epidemiological and health-systems environment. Methodologically, prior studies and deployments often rely on single algorithms such as Random Forest. These models struggle with complex, nonlinear relationships and noisy, multi-source climate–health data, revealing a methodological gap for more robust approaches (Chen et al., 2023). Ensemble machine learning methods offer improved accuracy and stability but are under-tested as operational decision-support within county workflows. Therefore, this study addresses these gaps by developing an ensemble machine-learning model that integrates local climate and health data to generate actionable, lead-time forecasts for healthcare decision support in Migori County. The aim is to provide a functional, context-appropriate model that improves prediction quality and supports proactive planning, thereby strengthening resilience of healthcare services under a changing climate.

1.5 Main objective

The main objective for the study was to optimize healthcare resilience against infectious diseases by utilizing climate-informed data analytics and Ensemble machine learning model to predict, mitigate, and effectively manage these health risks.

1.5.2 Specific Objectives

i. To analyze the current utilization of climate data analytics in healthcare decision-making processes.

ii. To develop an Ensemble machine learning model for incorporating climate data into healthcare decision-making processes.
iii. To test and validate the Ensemble machine learning model for incorporating climate data into healthcare decision-making processes.

1.6 Research Questions

i. How are climate data analytics currently utilized in healthcare decision-making processes?
ii. How can an Ensemble machine learning model be developed to incorporate climate data into healthcare decision-making processes?
iii. How effective is the Ensemble machine learning model in predicting and managing health risks by incorporating climate data into healthcare decision-making processes?

1.7 Significance of the study

The analytic investigation of climate data plays an essential function to advance healthcare resilience. The machine learning-based ensemble model increases the precision of predicting health hazards linked to climate events and strengthens response and readiness capacities. The model's integration enables better resource allocation, leading to the best possible placement of medical supplies and personnel. The shift from reactive to proactive healthcare management will improve emergency preparedness and reduce disruptions. The study's goal was to provide policymakers with the necessary information to create healthcare strategies and policies that work. Research results contributed to the development of a healthcare system that is resilient to the effects of climate change and to the protection of communities that are particularly vulnerable.

1.8 Expected Outcomes of the Study

The research resulted in multiple vital outputs which strengthen Kenya's healthcare system's ability to deal with climate change effects:

- The development and testing of ensemble machine learning model boosted the ability to predict climate-related health dangers that affect infectious diseases.
- The research also with the integration of the model, enabled a change in the management of health from reactive to proactive one.

1.9 Justification of the Study

Because it met the growing need for healthcare decision-making that incorporates climate data analytics, this project was essential for predicting and controlling malaria epidemics in Kenya. Due to its extreme sensitivity to environmental factors like temperature, precipitation, and humidity, the disease required sophisticated forecasting systems. The study outlined a data-driven strategy to combat malaria in vulnerable areas while attempting to integrate medical intervention systems with climate change trends. For Kenya's health system, predictive models that used climate data generated by this study offered crucial insights into how to improve public health outcomes and healthcare resource management. By helping policymakers develop adaptive health threat strategies as they address climate change issues, the research fulfilled two crucial functions.

1.10 Scope of the study

The inclusion of climate data analytics into decision-making processes in this study aimed to strengthen the resilience of Kenya's healthcare system. The research addressed waterborne diseases (malaria), that are especially sensitive to changes in weather. Particularly in the Western Region, which has one of the highest malaria transmission levels in Kenya (Waweru, 2024), the research will focus on the Lake Victoria Basin, struck by its tropical climate and nearby water sources ideal for mosquito breeding. The research created and validated an ensemble machine learning model including climatic factors to forecast disease outbreaks and improve resource distribution. Emphasis was placed on the Western Region in Migori County, which enabled a full examination of current trends and possible solutions for local environmental health risks, including malaria.

1.11 Limitations of the Study

i. If long-term climate trends are not considered, the measures may be less robust, and the study's conclusions were not generalisable to other nations.

ii. The process of incorporating climate data analytics into healthcare decision-making is complicated because numerous licenses must be obtained, and developing the system to incorporate the model was difficult because system development is expensive.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

The research on climate-informed data analytics in healthcare is reviewed in this chapter. It describes the ways in which disease patterns are impacted by climate variability, lists the obstacles to implementing analytics, and highlights the possible advantages for early warning and planning. The review also identifies gaps that support the creation of an ensemble, climate-informed decision-support model.

2.1 Utilization of Climate Data Analytics in Healthcare Decision-Making

Climate change is bringing about rapid alterations in disease spread worldwide as Sub-Saharan Africa together with Kenya notices a significant growth of climate-connected diseases consisting of waterborne infections (malaria) and respiratory illnesses (Rayan et al., 2021a). Research by (Yadav & Upadhyay, 2023) demonstrates how heat rise and inconsistent rainfall and high humidity transform geographical territories into suitable environments for climate-sensitive diseases. Current healthcare organizations face an information deficit combined with resource constraints which make it difficult to predict and address health risks arising from climate changes (Sabasteanski, 2021). The existing information deficit offers health systems a valuable opportunity to merge climate data analytics which will boost resilience while creating predictive safety systems for vulnerable populations. Ensemble learning models of climate data enable healthcare authorities to anticipate and manage disease spread of both malaria and respiratory infections in regions affected by climate change variation.

The utilization of multiple investigations on medical decision-making processes that utilize climate data analytics. Zhou, (2022) demonstrates that climate data serves as an essential tool for respiratory disease outbreak forecasting which combines temperature and humidity measurements to forecast Asian disease spikes. A study by (Y. Wang et al., 2024) demonstrated in China how healthcare systems and climate data integration leads to better disease predictions by developing improved modeling approaches with air pollution inclusion. Research results demonstrate that better analytic methods would enhance our capabilities to forecast and manage disease transmission effectively. A study by (Martineau et al., 2022) demonstrated that affiliate medical records with climate data yielded better results in forecasting malaria spread throughout African regions. Mwangi et al. (2022) demonstrated through their research in Kenya that real-time climate information serves to track malaria outbreaks which emphasizes the requirement of forecasting systems offering timely accurate predictions in the context of climate change. Research in Sub-Saharan Africa has now dedicated more attention to blending climate data for malaria control operations.

The awareness about climate data analytics keeps rising even though specific barriers persist. Mabaso et al. (2023) point out that better disease prediction becomes possible through climate data integration, but many healthcare facilities do not possess the required technological structures with sufficient data-sharing abilities to make complete use of these technological solutions. Several limitations exist in the Kenyan malaria control program due to insufficient climate data collection and restricted access to quality meteorological information according to Kamau et al. (2022). The study demonstrates how climate data analytics have great potential, yet healthcare must solve significant technical and data infrastructure problems to achieve complete clinical advantage.

This review shows various restrictions that appeared in past investigations concerning this subject. Zhou et al. (2022) and Wang et al. (2023) showed how climate data helps predict respiratory diseases, yet their research concentrated on cities while the diseases malaria mostly occurs in rural areas. The research by Zhou et al. (2022) and Wang, (2023) demonstrated that Random Forest hit an accuracy mark of 85%, Support Vector Machines (SVM) reached 82% precision, but Gradient Boosting delivered 87% accuracy in diagnosing respiratory disease outbreaks from climatic factors. The research primarily investigated urban spaces while neglecting to show effectiveness of these models toward malaria and other infectious diseases present in rural territories.

Mabaso et al. (2023) pointed out how numerous ensemble machine learning models operating in Africa maintain low accuracy because they use inadequate or outdated climate data. Public health interventions become inadequate due to historical weather data limitations which forecast disease transmission between malaria and cholera. The development of better predictive models in healthcare requires updated and comprehensive climate information for solving health challenges related to climate change. According to research by Diop et al. (2022), Random Forest and Artificial Neural Networks (ANN) and Gradient Boosting techniques were used to forecast malaria outbreaks. While Random Forest's accuracy was 85%, the accuracy of the ANN and Gradient Boosting techniques was 82% and 80%, respectively. Unfortunately, the limited and low-resolution climate data that these predictive models rely on hinders their ability to accurately predict malaria outbreaks across a variety of geographic regions. Because different geographic regions require improved accuracy assessment, especially in rural malaria hotspots, the need for higher resolution data sets and sophisticated predictive strategies becomes even more critical.

Furthermore, there is minimal emphasis on merging climate data with current healthcare infrastructures. While Njuguna et al. (2021) demonstrated that ensemble machine learning models could enhance malaria predictions in Kenya, their research disclosed that numerous healthcare systems do not possess the ability to incorporate such models into regular disease monitoring.

This study aims to address these gaps by developing an ensemble machine learning model to predict and control infectious diseases in the Lake Region of Kenya through combining climate data. Linked with local healthcare practitioners and using high-resolution climate information this study seeks to develop an integrated predictive model which can fit seamlessly within Kenya's healthcare system. The research provides valuable additions to current literature through demonstrations of climate data analytics implementations in limited resource areas and exploration of relevant challenges associated with their execution. The outcomes from this research will give valuable information to policymakers and healthcare practitioners and researchers to enhance healthcare resilience against climate change. This research study strengthens the existing knowledge base through its analysis of previous model weaknesses and positions potential solutions to generate localized disease prevention forecasts in Kenya and comparable settings.

2.2 Integrating Climate Data Analytics into Healthcare Systems

Regional health systems in Sub-Saharan Africa must consider the obvious evidence of disease transmission relationships with climate variability, as demonstrated by rising global temperature trends and increasingly unpredictable weather patterns. Infectious diseases that respond strongly to environmental temperature and precipitation present new challenges for medical facilities worldwide. Healthcare systems must include climate data analytics as a fundamental step to manage health risks that come from climate change. Healthcare professionals using integrated climate and traditional healthcare information can prepare resources ahead of time and predict disease outbreaks to enhance patient care. The successful integration of multiple data sources for predictive model creation presents a major operational challenge that affects numerous regions especially those having limited resources.

Climate data must be used in healthcare decision-making, according to several recent studies. Li et al. (2023) found that the integration of real-time climatic data, such as temperature and humidity, into healthcare systems improved the forecasts of respiratory disease outbreaks in Southeast Asia. By using timely climate data, medical professionals were able to make better decisions and lessen the impact of disease epidemics. Tambo et al. (2023) investigated the impact of climatic factors on the prediction of malaria outbreaks in West Africa. Through their research, the authors showed how incorporating climate factors like temperature and precipitation into malaria control plans improves readiness for outbreak response in remote and challenging-to-reach locations.

Research shows that healthcare institutions within Kenya now actively understand the importance of using climate data for their operations. National disease surveillance systems are being analyzed by Ndegwa et al. (2022) to determine their ability to utilize climate data for better cholera and malaria outbreak management. This study proves that national healthcare planning which incorporates seasonal rain data improves the timing of public health interventions and decreases disease spread rates. These benefits face technical along with infrastructural limitations that prevent their full implementation according to the research findings. According to Mutuku et al. (2023) the major breakthroughs in disease prediction through climate data analytics face critical challenges because of restricted access to precise climate data and substandard data quality. The authors of Mutuku et al. (2023) together with Ndegwa et al. (2022) discovered that inadequate data-sharing cooperation between Kenya's healthcare system and meteorological department produces operational inefficiencies in health-related climate data utilization.

Several problems prevent the successful implementation of climate data analytics into healthcare systems despite its demonstrated potential. The Southeast Asian respiratory illness prediction model developed by Li et al. (2023) encounter technical limitations because poor nations do not possess the capability to process climate data. Ndegwa et al. (2022) demonstrated how Kenya faces challenges because healthcare professionals do not possess sufficient technical abilities to read and leverage climate data for disease management and trustable high-resolution climate data remains difficult to obtain. The research by Tambo et al. (2023) established that adding climate data enhanced West African malaria forecasts however uneven and incomplete data limited prediction accuracy. The successful implementation of climate data analytics in healthcare requires both needed infrastructure and essential understanding of these systems which often prove scarce.

A majority of literature reports that the healthcare and climate sectors need to develop better collaboration along with addressing infrastructural limitations. The development of healthcare system capability remains vital because Omondi et al. (2022) suggest providing educational programs to train health professionals in reading and implementing climate data for forecasting diseases. Mwangi et al. (2023) advocated for developing universal exchange protocols to make the process of including climate information easier into healthcare databases. The research underlined that enhancing data interpretation capabilities especially for resource-constrained areas will generate more accurate and beneficial future disease predictions.

This study sought to address these challenges by developing a combination of machine learning techniques using weather data to project and manage malaria and respiratory diseases across Kenya. The research establishes better medical-meteorological organization collaboration while developing improved access to improved climate information to enhance its quality. A predictive disease management model will improve when the study enhances data-sharing processes and enables healthcare workers to interpret complex climate information. The research results demonstrated how these models successfully operate within healthcare institutions of limited data capacity and technical capability thus expanding knowledge about climate data analytics. The research delivered vital information to both policy leaders and healthcare providers to strengthen their capacity to predict and deal with health issues originating from climate change.

2.3 Ensemble Machine Learning Model for Climate Data Integration

Numerous recent studies highlight how important it is to use climate data when making healthcare decisions. According to Li et al. (2023), the integration of real-time climatic data, such as temperature and humidity, into healthcare systems improved the forecasts of respiratory disease outbreaks in Southeast Asia. Timely climate data helped medical professionals make better decisions, which helped them reduce the impact of disease epidemics. Tambo et al.'s (2023) study examined the impact of climatic factors on the prediction of malaria outbreaks in West Africa. Through their study, the authors showed how incorporating climate factors like temperature and precipitation into malaria control plans improves the readiness for outbreak response in remote and challenging-to-reach locations.

Research studies indicate ensemble machine learning methods achieve superior prediction results than traditional statistical techniques. Liu et al. (2022) revealed that ensemble models surpassed single models such as logistic regression and decision trees by attaining higher levels of performance for predicting East Asian respiratory disease outbreaks through climate data combination analysis. My research addresses the gap created by the missing variables from this model. The ensemble model achieved higher accuracy than 85% by building reliability through factors such as temperature and humidity and air quality assessment. The requirement exists to enhance public health predictions using state-of-the-art modeling techniques. The work of Odu

(2021) demonstrated that ensemble machine learning outperformed classic predictive models such as naive Bayes and support vector machines because his model achieved rates of around 65% and 68% respectively. A predictive ensemble model achieved a substantial accuracy boost when healthcare and climate data such as rainfall and temperature were combined thus reaching more than 80% accuracy. The performance of ensemble methods indicates their effectiveness in dealing with disease forecasting situations found in climate-sensitive regions.

Waweru et al. (2022) created a malaria transmission forecasting model as an ensemble model at Kenyan urban centers and demonstrated that the models can capture complex, non-linear patterns, as compared to conventional methods. Nevertheless, they optimise their model to work in an urban environment and do not consider constraints on the quality of data, ecological variation, or behaviour that are found in rural environments. The model does not consider key rural parameters including the variety of the habitat of vectors, agricultural land use, inconsistent health-seeking behavior and changes in microclimatic factors which make the model less applicable in such settings as in Migori. Such gaps provide the purpose of the customized ensemble method that would consider the rural environmental and socio-economic factors that would enhance the precision of prediction and inform the specific intervention. Siamba (2022) employed ensemble modeling techniques to forecast changes in Kenyan respiratory diseases occurring under varying climate conditions. This method achieved improved predictive outcomes compared to standard statistical methods such as logistic regression method that normally reached prediction accuracy of 70%. The ensemble model surpassed traditional approaches by integrating climatic variables to achieve predictive accuracies exceeding 85%. Scientific evidence indicates ensemble models enhance climate-health investigations because they predict disease transmission patterns in areas with significant climate variation.

Machine learning models based on ensembles prove challenging to operate in areas with substandard technical systems despite providing useful advantages. Studies indicate ensemble models optimize East Asian respiratory disease forecasts provided that high-quality real-time climate data remains accessible to these systems (Sokhi et al. 2022). Regions which don't conduct consistency, or sufficient data collection will see their ensemble prediction accuracy levels decrease. The application of ensemble models faces challenges due to data quality and availability problems within West Africa's underfunded climate monitoring systems according to Dupar et al. (2021). The performance of ensemble machine learning models depends heavily upon reliable data infrastructure since technology has enormous potential for disease prediction.

Malaria prediction related to ensemble models' successful implementation in Kenya struggles because healthcare units and meteorological bodies do not share data correctly or have sufficient technical expertise according to McMahon (2021). Sequentially Mahajan et al. (2023) stressed that healthcare and technical personnel require improved education regarding the implementation of complex models although ensemble models improve disease prediction accuracy. These studies demonstrate that healthcare systems need better technical capabilities for properly merging climate and health data in order to utilize ensemble machine learning models effectively.

This study develops an ensemble machine learning model to anticipate malaria and respiratory disease outbreaks in Kenya using merged climate data according to Muriithi et al. (2024). The study aims to improve data quality and accessibility while addressing deployment challenges of ensemble models for low-resource environments (Ghosheh et al., 2023). The study develops a predictive model for operational ensemble machine learning that Kenyan healthcare system personnel can readily apply (Ombui, 2023). Findings proving the usefulness of ensemble models in disease prediction applications for regions with climatic variability will advance our understanding of how to integrate climate data. The study offers crucial information to researchers, medical professionals, and legislators who are interested in machine learning techniques to strengthen the resilience of healthcare systems to climate change.

2.4 Effectiveness an Ensemble Machine Learning Model in Predicting and Managing Health Risk.

Due to increasing climate variability, ensemble machine learning models are sophisticated instruments for predicting and managing health risks. Together, a number of interrelated models produce accurate disease predictions, assisting medical facilities in developing strategies to combat diseases linked to climate change. Because of the interactions between the climatic factors that cause respiratory infections and malaria, complex predictive models are required (Klepac et al., 2024). These issues can be resolved by using ensemble models, which provide extremely accurate data-based forecasting that reduces errors and enables more informed choices about the allocation of public health resources and strategic planning. (Cui et al., 2021).

According to recent studies, ensemble machine learning systems can accurately predict the health risks associated with changes in climate variability. According to Peng et al. (2021), the use of ensemble learning techniques during heatwaves and days with high pollution levels improved the accuracy of respiratory disease modelling in East Asia. In comparison to conventional statistical methods, this study's incorporation of climate data into prediction models demonstrated improved predictive accuracy and a reduced margin of error. Debnath et al. (2024) claim that ensemble machine learning generated forecasts regarding malaria outbreaks in West Africa, lowering prediction uncertainties and speeding up response efforts. Healthcare workers were able to take immediate action by being alerted about upcoming outbreaks before they happened thanks to the prediction model's combination of weather elements.

The application of ensemble machine learning models for risk management achieved successful results through research conducted in Kenyan territories. The ensemble model developed by Leung et al. (2023) achieved improved accuracy and speed in malaria transmission outlooks throughout different regions of the country. Medical systems could better distribute their resources to effectiveness by using forecasting capabilities particularly for low-serviced areas experiencing seasonal malaria outbreaks (Leung et al., 2023). Otieno et al. (2023) demonstrated how ensemble learning improved forecast accuracy of respiratory infections during Kenyan rainy seasons which enhanced medical resource capabilities for climate illness hospital admission prediction. Ensemble models offer better health risk forecasting and management solutions for healthcare systems which operate in regions influenced by climate variability according to results from these studies.

Numerous ongoing obstacles stand in the way of effectively utilising ensemble machine learning models to enhance health risk predictions. Krayenhoff et al. (2021) state that the availability of climate data and quality standards have a significant impact on the success rate of these models. When ensemble predictions rely on climate data that is either unreliable or unavailable, the accuracy is regionally limited. Khan et al. (2024) came to the conclusion that while ensemble models were successful in improving malaria prediction, their effectiveness was directly correlated with the use of high-quality, customised climatic data. According to McMahon (2021), national health strategies that failed to share data between meteorological services and healthcare systems created obstacles to the increased accuracy of ensemble models for predicting malaria outbreaks in Kenya.

In addition to addressing their data-related issues, Rane et al. (2024) noted that healthcare systems need additional technical training to enable proper usage of ensemble machine learning models. The study discovered that healthcare providers' ability to make prompt clinical decisions based on these insights was limited because they did not fully comprehend model outputs due to their lack of knowledge about prediction tools. Effective technology adoption in healthcare institutions and the acquisition of high-quality climate data are prerequisites for the extensive potential of ensemble machine learning models for healthcare risk prediction.

This research establishes and integrates an ensemble machine learning system that accurately addresses Kenyan healthcare demands regarding infectious diseases (Malaria) and respiratory illnesses within the Western Region. The model will enhance disease outbreak prediction accuracy by utilizing detailed climate data according to Ijeh et al. (2024). This research emerges as a path to develop resilient healthcare infrastructure which can confront the health risks from climate change by strengthening healthcare systems in technical and operational aspects. This study through its empirical findings will demonstrate the effectiveness of ensemble machine learning approaches in limited resource environments while offering implementable guidance to improve healthcare preparedness against climate-related diseases.

2.5 Theoretical Frameworks

2.5.1 Domain Adaptation Theory

Domain adaptation theory describes that machine learning models that are trained in one domain frequently fail when used in another domain since the statistical distributions of features and outcomes are dissimilar (Wilson and Cook, 2021). This becomes of special concern in climate-health modelling when there is an effort to generalize an urban-trained model to a rural environment. In urban areas, there are organised accommodation, regulated drainage, predictable alterations of temperature, and more regular access to healthcare, whereas in rural settings, such as Migori, the environment of vectors is disorganised, wetlands are extensive, microclimates are unreliable, and healthcare-seeking behaviour is not very consistent. These differences lead to a shift in distribution - that is, the climate, ecological, and behavioural data that causes malaria transmission are not similar by any means to those available in the training domain. According to the recent research, transfer learning is not as effective in case the distributions of features are significantly distinct, particularly, in climate-sensitive disease prediction (Zhuang et al., 2021; Kouw and Loog, 2022). The Domain Adaptation Theory thus offers an explanation in theory of the reason why transfer learning is not applied to the already existing urban-based model. Instead, it will help to build the rural-specific ensemble model educated directly by the climatic, environmental and epidemiological factors of the Migori County.

2.5.2 Framework of Climate-Health Vulnerability.

According to the Climate-Health Vulnerability Framework, there are three interrelated elements of disease risk exposure to climate variables, population sensitivity, and the adaptive capacity of the health system (Ebi et al., 2021; Atwine et al., 2023). This framework corresponds to the creation of a climate-informed ensemble model since it formalizes the theoretical foundations of using model inputs as the temperature, rainfall, humidity, land use, and population dynamics. Currently, the existing climate-health work shows that these factors are crucial in predicting malaria due to their effects on the survival of mosquitoes, their breeding life cycles, as well as their competence to transmit malaria (Ryan et al., 2022; Tambo et al., 2023). The structure also aids in the rationalization of why the area of concentration should be on Migori, which is still among the hotspots of malaria in Kenya following the high rainfall patterns, nearness to lake Victoria, large wetlands, and breeding mosquito habitats throughout the year. The conditions expose people, and the healthcare resources available are low, decreasing the adaptive capacity, and the county is extremely vulnerable. This framework places the proposed model in solid theoretical ground, and the study reveals that the components of the model, namely climatic and epidemiological, are theoretically based on reported literature on the topic of climate-health.

2.5.3 Predictive Modelling and Evaluation Theory

The theory of predictive modelling and evaluation points out that the performance of machine learning frameworks has to be thoroughly evaluated with the aid of such metrics as accuracy, precision, recall, F1-score, RMSE, and AUC-ROC (Hossin and Sulaiman, 2023; Zhang and Sheng, 2024). In climate sensitive disease forecasts, they are the metrics by which a model can be determined to be able to generalise to the real world conditions and be used to make effective decisions in the interest of public health. The absence of defined metrics of evaluation of predictive models makes them prone to generate biased or unreliable results particularly when they are trained on nonhomogenous climatic and epidemiological data. The existing literature emphasizes the importance of model evaluation in malaria prediction because varying relationships between climatic and disease differences should be expected in different regions and might readily mislead unless they are evaluated (Tran et al., 2022; Owusu et al., 2024). This theory thus approves the incorporation of evaluation measures in the current study and makes the methodological basis of the proposed ensemble model more robust. It will make sure that the predictive system is scientifically sound, contextual, and has the ability to aid malaria control activities in Migori County.

2.6 Conceptual Framework

The conceptual framework of this study demonstrates how healthcare decision-making can benefit from the use of climate data analytics. The study focusses on ways to identify and treat illnesses brought on by climate change. Using vital climatic and health-related data, this analytical approach assesses the impact on human health. These variables are used by an ensemble machine learning system to perform accurate disease forecasts. These projections can be used by healthcare organisations to determine which interventions to implement and how best to allocate their available resources. This framework addresses existing research gaps by establishing a single, adaptable mechanism for enhancing the accuracy of climate change forecasts and medical interventions in vulnerable areas.

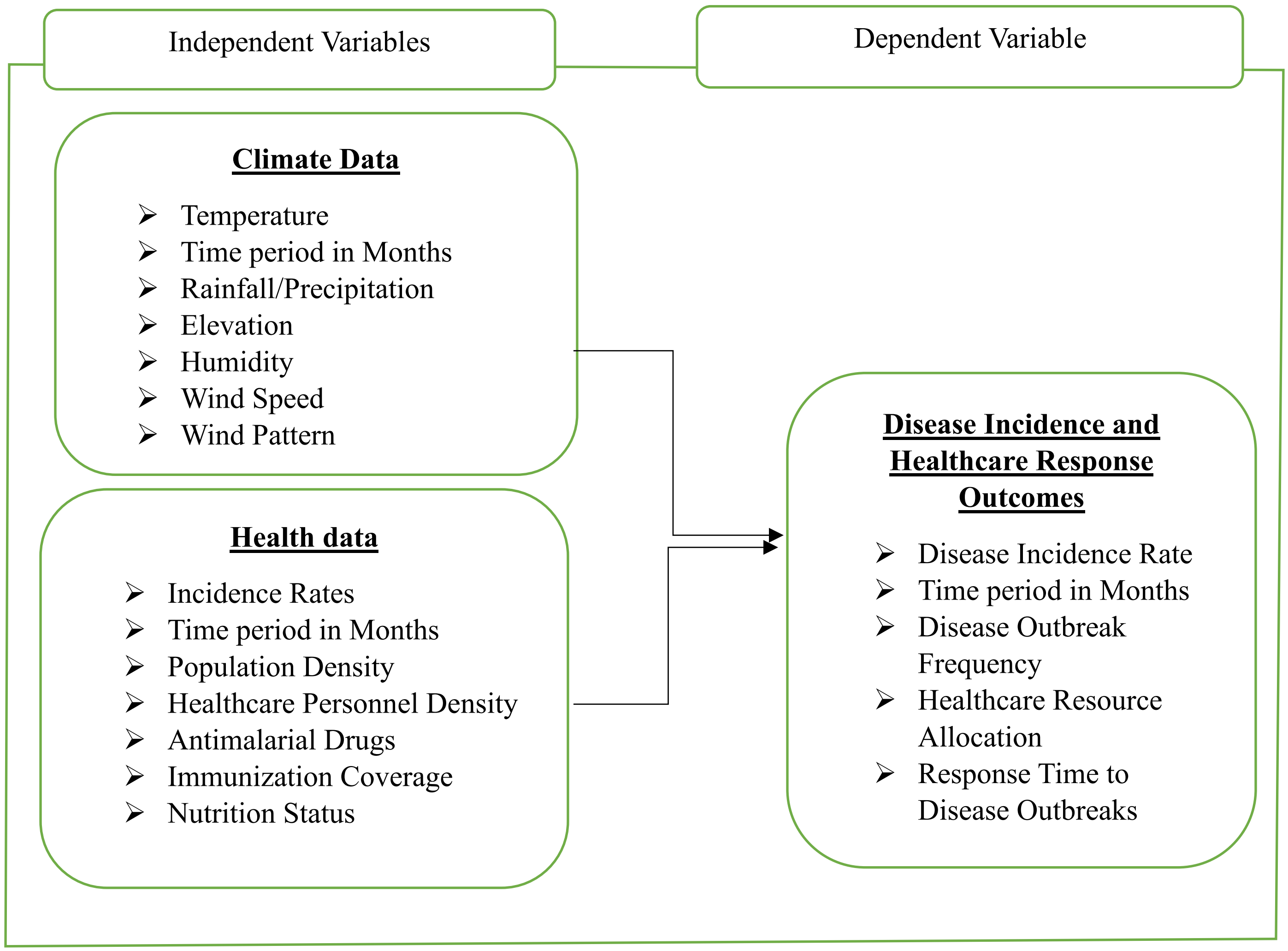


Figure 2. 1 Conceptual Framework

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

The methods applied for research investigation are discussed in this chapter to fulfill the study objectives. The research methodology for this study featured four major sections: design, data collection approaches, model development systems, validation approaches, and considerations about research paradigms and ethics.

3.2 Research Paradigm

This research study adopted the pragmatism research paradigm. The paradigm enables researchers to employ any research method quantitative or qualitative and mixed for achieving realistic solutions as well as measurable outcomes (Mulisa, 2022).

Research has proven the pragmatic paradigm to be effective across multiple fields according to published studies. The integrating of machine learning into healthcare systems through pragmatic methodology led to improved pandemic response decision-making according to Adetunji et al. (2021). The researchers utilized data-driven methods in their work because they believed in practical solutions while this investigation integrates climate data for predicting healthcare dangers. Rogers et al. (2021) adopted pragmatism to understand how climate change interacts with U.S. healthcare systems by connecting quantitative climate models with healthcare professional insights to develop functional recommendations that strengthen healthcare system resilience. The researchers at dos et al. (n.d.) performed evaluative research based on pragmatism to understand healthcare challenges from climate change in low-resource settings through a combination of quantitative and qualitative research methods for direct adaptation of healthcare systems. Gwakisa et al. (2023) conducted a pragmatic paradigm study to examine African healthcare systems' climate

change preparedness by combining data analysis methods with stakeholder involvement for obtaining practical research outcomes.
The research provided value by connecting big data analytics of climate with healthcare decision systems for developing workable solutions. The adaptable nature of pragmatism enabled researchers to unite machine learning model creation methods with qualitative healthcare industry examinations thereby ensuring the research stays focused on delivering practical solutions to boost healthcare resilience. The research employed pragmatism to create data-based solutions which solve the practical healthcare problem of adapting to climate change.

3.3 Research design

Under this study the researcher created and evaluated ensemble machine learning models for healthcare decision-making using climate data by implementing both a quantitative research design and the Design Science Research (DSR) methodology. The method enabled a problem-solving structure which evaluates both infrastructure and socioeconomic elements by using evidence-based data-driven analysis. This research followed the DSR framework that begins by identifying problems then creates artifacts followed by evaluations and knowledge contribution to the field. Machine learning models serve as an example for healthcare decision-making according to research by Van der Schaar (2021), "Nonparametric Estimation of Heterogeneous Treatment Effects: From Theory to Learning Algorithms." Ebi and Hess (2020) alongside Sorensen et al. (2020) conducted quantitative research to project disease spread levels and gauge healthcare system deficiencies in areas lacking funding respectively. This study utilized ensemble machine learning modeling through artifact development to perform quantitative analyses using Root Mean Square Error (RMSE) together with R-squared values and precision-recall breakdowns thus advancing existing works to establish climate-related healthcare prediction abilities. The research conducted by Opoku et al. (2021) showcases the critical value of climate data analytics in healthcare decisions when they utilized climate data modeling for African healthcare system climate change readiness assessment. By using the DSR approach researchers can maintain continuous enhancement of this machine learning model to help resolve specific problems along with developing theoretical knowledge and practical answers for healthcare resilience and climate adaptation.

3.4 Data Collection Procedures

In the study, secondary data were collected using Kenya meteorological station and the Lake Victoria Basin medical records during 2014 to 2023. The climate variables were temperature, rainfall, humidity, and other environmental indicators, whereas health was comprised of malaria and respiratory disease outbreak and incidence. The datasets were collected due to the formal data-sharing agreements and ethical clearance that guarantee the adherence of privacy and confidentiality standards.

Before analysis, data cleaning procedures were carried out such as dealing with missing values, standardization of types of variables, and coding categorical data where required. The temporal and spatial coincidence resulted in the merging of the climate and health datasets in order to allow the correct correlation and predictive modeling. The processed data was further divided into training, validation and testing data to develop and validate the ensemble machine learning models that will guarantee reliable and repeatable results.

3.5 Model Building and the Validation Approach

The research relied on secondary data which combines Kenya Meteorological Station climate records and Lake Victoria Basin Region medical statistics to check the accuracy of an ensemble machine learning prediction model. The research period spans 10 years while it examines healthcare indicators as dependent variables supported by climate data as independent variables.

The assessment aims at building a robust forecasting model to determine weather variations impact on healthcare results.

3.5.1 Current Utilization of Climate Data in Healthcare Decision-Making

The collected data revealed patterns which show associations between health results and climate elements throughout multiple periods. The analysis assessed both healthcare measurement data including disease outbreak frequency and incidence rates together with climate data. These variables were visually mapped according to their temporal sequences for identifying correlations. The monthly average climate data was correlated to disease incidence data through time-series analysis. Pearson correlation will establish the relationships that exist between climate variables (xi) and health outcomes (yi).

$$r = \frac{\left\{ \sum_{\{i=1\}}^{\{n\}} (x_i - \bar{\{x\}})(y_i - \bar{\{y\}}) . \right\}}{\left\{ \sqrt{\left\{ \sum_{\{i=1\}}^{\{n\}} (x_i - \bar{\{x\}})^2 \sum_{\{i=1\}}^{\{n\}} (y_i - \bar{\{y\}})^2 . \right\}} \right\}}$$

Equation 3. 1: Pearson Correlation

The analysis determined the extent to which current healthcare decision-making depends on climate variables. The research employed Python for data processing tasks and data union and seaborn for drawing correlation heatmaps and relationship visualization. A thorough understanding of the current use of climate data and its possible effects on medical decision-making.

3.5.2 Ensemble Machine Learning Model

Employment of ensemble strategy resulted in enhanced prediction accuracy while improving the system's overall reliability through amalgamation of multiple algorithms. Three different foundational models comprising Artificial Neural Networks (ANN) and Random Forest (RF) and Gradient Boosting Machines (XGBoost) will be used in this research through the ensemble framework. These models were chosen because they work best together to handle diverse relationships between health outcomes and climate variables that show complex and high-dimensional behavior along with non-linear behavior. The ensemble of distinct predictive models will collectively improve prediction accuracy for healthcare resilience as it pertains to climate change processes.

3.5.3 Model Design and Development

The application of Gradient Boosting Machines (XGBoost) was preferred to handle non-linear patterns present in climate health data because of its proven efficiency. The boosting algorithm XGBoost utilizes weak learners which are decision trees to create a single powerful predictive model through consecutive iterations. The training process minimizes the combination of prediction error and an overfitting prevention penalty term that makes up the regularized loss function.

The loss function for XGBoost is represented as,

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

Equation 3. 2 XGBoost

Where;

- $l(y_i, \hat{\{y\}}_i)$ is the loss function (mean squared error for regression or log-loss for classification).
- $\Omega(fk) = \gamma T + \{1\}/\{2\}\{\lambda|w|^2\}$is the regularization term for each tree, where TTT is the number of leaf nodes, www represents the leaf weights, γ controls complexity, and λ controls overfitting.

This regularization method ensured model generalization success for new data regardless whether the application involves complex climate elements such as temperature alterations or precipitation patterns or humidity changes. The major strength of XGBoost provides a feature ranking feature that revealed climatic elements significantly impact disease patterns as well as healthcare outcomes. The understanding of climatic factor relationships with healthcare performance depends heavily on the relevance of this feature.

Data cleaning operations commenced on both climate and healthcare datasets followed by missing value repair and variable type encoding process. The data analyzed was separated into groups for testing and training purposes. The training process was performed through the Scikit-learn library which operates on the training set. The important hyperparameters such as n_estimators and max_depth and learning rate will be fine-tuned through Grid search or random search techniques. The test set assessment of the model will utilize accuracy together with precision and recall and F1-score as performance indicators.

Random Forest algorithm proved suitable because it helped present robustness in addition to effective noise handling capabilities. The Random Forest produces many decision tree models through bootstrapped samples of training data before it applies their predictive values. The approach decreases statistical variations while helping to minimize overfitting events. The main advantage of Random Forest occurs when many features exist because it selects features automatically through random subset selection at each split.

The aggregated prediction of the Random Forest model will be presented as,

$$y^{\wedge} = \frac{1}{N}\{\sum fi(x)\}$$

Equation 3. 3 Random Forest

where $f_i(x)$ is the prediction from the i-the tree, and N is the total number of trees in the forest.

3.5.4 Implementation of the Random Forest

Categorical variables received encoding treatment while missing data receives handling during the climate and health data preprocessing steps. Scikit-learning library served to build a model that trains on the available training dataset. A maximization process will optimize the maximum depth of each tree parameter (max_depth) together with the number of estimators (n_estimators). The model evaluation used accuracy together with precision and recall and the F_1 score as classification metrics.

Because of its ability to model very complex nonlinear relationships Artificial Neural Networks (ANN) used for the ensemble. A neural network exists as connected sequences of neuron layers which apply weighted sums of input data through activation functions. Through their adaptive nature ANNs establish complex relationships among climatic elements so they can model disease outbreak relationships between temperature and humidity. An ANN requires a layered structure that consists of an input layer followed by at least one hidden layer before ending with an output layer and numerous neurons within each layer. Each neuron within a neural network requires an activation function which could be ReLU or Sigmoid to determine its output values.

The output of a neuron in a fully connected network is calculated as:

$$z = w \cdot x + b, a = \phi(z)$$

Equation 3. 4 Output of a neuron network

Where:

- w is the weight of the connection between neurons,
- b is the bias term,
- x is the input, and
- $\phi_{(z)}$ is the activation function, such as ReLU, which is defined as $\phi_{(z)}$ =max (0, z).

Implementation includedo preprocessing where the input data will be normalized to ensure faster and more stable convergence during training. The programmable machine will consist of an input layer together with various hidden layers before it terminates at the output layer. The choice of model complexity decides both the number of hidden layers and the total number of neurons to be utilized. Learning Back propagation algorithm and Gradient Descent algorithms will minimize the loss function (binary cross-entropy or mean squared error) for the task. A process to optimize vital parameters including number of layers, neurons, learning rate and activation functions and more will be implemented. A forward pass of the data through the model will be followed by its evaluation through classification metrics on test data.

3.5.5 Testing and Validating the Model

Tests and validations provided the model's reliability and robustness. A total of 15% of data served as testing data and 15% will function as validation data for this case. The sample was distributed based on stratified sampling that maintains outcome-based representation throughout all groups. The Model performance was assessed using

1. Accuracy $= \frac{\{TP+TN\}}{\{TP+TN+FP+FN\}}$
2. Precision $= \frac{TP}{\{TP+FP\}}$
3. Recall $= \frac{\{TP\}}{\{TP + FN\}}$
4. F1-Score which will be the Harmonic mean of precision and recall.

Testing and validation of the model was utilized Python programming language. The performance metrics were computed through classification report and roc_auc_score functions. The model required evaluation by using a 10-fold cross-validation method for robustness assessment. A validated model featuring high performance levels in health risk prediction and management used integrated climate data as the outcome of this work.

3.5.6 Integrating the model into Decision-Making Framework

A healthcare decision-making framework integrated the model while its visualization dashboards function to improve stakeholder understanding of model outputs. Different tests of the model will occur through simulations utilizing varying climates and healthcare reactions. The collected insights generated functional strategies to establish warning systems or enhance medical resource distribution. The dashboards for interactive presentation will be built using Plotly Python. The predictive model outputs serve as input to run simulation scenarios through the platform. The decision-making instrument serves as a functional tool that establishes links between climate information and operational healthcare solutions to build healthcare resilience.

3.8 Ethical considerations

The experiment complied with the rigorous ethical standards of collection, processing and analysis of climate and health data. Participant confidentiality was ensured by making sure that all the data were aggregated and anonymized without any personally identifiable information. Only the authorized researchers were allowed to access raw datasets, and the procedures of the safe data handling were formally documented. These guarantees safeguarded the rights of the participants and that the ethical considerations were upheld during the research process. Data collection was done in compliance with the national and institutional regulations of research through the ethical approval of the research by the National Commission for Science, Technology and Innovation (NACOSTI) and the institutional review board. Data-sharing and confidentiality agreements formal consumption with all data-providing institutions were established based on the conditions of what could and could not be used, stored, and disseminated. These were done to make sure that the research activity did not violate the legal and ethical standards on the usage of health and climate data. To achieve reproducibility and validity, the study recorded every data processing and data analysis procedure. Similarities in protocols were used in the combination of climate variables and health records to ensure uniformity within the data sets. To minimize errors and ensure predictive outputs are accurate, validation methods and quality checks were used. The study also upheld methodological integrity by ensuring privacy and the welfare of data sources through ethics and combination of these safeguards.

CHAPTER FOUR

DATA ANALYSIS, PRESENTATION AND INTERPRETATION

4.1 Introduction

This chapter presents findings on malaria risk drivers, detailing data sources and methods. It quantifies the effects of rainfall, temperature, humidity, and wind on transmission, and evaluates health-system capacity (personnel density, antimalarial stocks, immunization, nutrition) on facility resilience. Analyses include descriptive statistics, inferential models of climate malaria relationships, and ensemble machine-learning forecasts for short-term risk, situated within Kenyan/Lake Victoria literature and conducted under ethical standards.

4.1 General Overview

To enhance the clarity of the methodologies, the thesis gave a clearer account of the data collection procedures. The analysis has been based on the regularly obtained health data on a facility-level that were filtered out of the DHIS2 platform and county health records along with the climate parameters that were sourced to the Kenya Meteorological Department (KMD) and verifiable satellite data. Data extraction and data quality were conducted by experienced health information officers and biostatisticians of the Migori County Health Department, and all of them were officially trained in HMIS standards, data validation and data surveillance reporting on a weekly basis. Their participation made the dataset credible and reliable. Besides this, the ten study areas marked as A to J which had been anonymized due to confidentiality were literally defined by the mapping of each mark to the respective geographical catchment area. The parameters by which such regions are chosen such as the disease burden trends, accessibility of stable records, and the variability of climatic conditions across the county were also explained to promote transparency, reproducibility and rigor of methodology.

The data set comprises weekly observations from public health facilities in Migori County, spanning 2015-01-02 to 2024-12-05. In total, 5000 facility-week records from 10 facilities were analyzed, enabling seasonality assessment (MAM/OND) and alignment with routine decision cycles.

Table 4. 1 The general Overview of the Data

Metric	**Value**
Timeframe	2015-01-02 to 2024-12-05
Regions (n)	10
Observations (rows)	5000
Cadence	Weekly (facility-level)

4.2 Climate Variables and Malaria Cases

The study results indicated that climate conditions in Migori follow a clear MAM/OND seasonal pattern, and this pattern aligns with malaria transmission dynamics. Weekly averages show substantial rainfall pulses (mean = 54.6 mm, SD = 26.7 mm) accompanied by warm temperatures (mean = 22.3 °C) and moderate humidity (mean = 65%), creating favorable breeding and survival conditions for malaria vectors.

Table 4. 2 *Summary statistics for climate and health variables*

Variable	**N**	**Mean**	**SD**	**Min**	**25%**	**50%**	**75%**	**Max**
rain_mm	5000.0	54.62	26.71	0.0	33.9	48.9	75.12	133.0
temp_c	5000.0	22.29	4.32	13.21	18.2	22.54	26.33	30.96
humidity_pct	5000.0	65.17	7.33	42.3	59.6	65.2	70.6	86.3
wind_speed_ms	5000.0	2.51	0.45	1.23	2.17	2.52	2.84	3.78
malaria_cases	5000.0	17.81	17.57	0.0	5.0	12.0	25.0	147.0

(Source 2025*) N = number of weekly observations aggregated across all regions of study.*

Also, the study found out that malaria burden is highly variable and surge-prone, although the median week records 12 cases and the 75th percentile is 25, peaks reach 147 cases per week. This skewed distribution justifies using the 75th percentile as the high-risk cut-off for defining alert bands, while the observed climate variability supports giving recent rainfall and temperature lags high modelling weight. In practice, these findings underpin actionable Standard Operating Procedures (SOPs) pre-positioning RDTs/ACTs before expected surges, scheduling surge staffing, and initiating WASH messaging immediately after heavy-rainfall episodes.

4.3 Health outcomes

The study results indicated that malaria displays pronounced seasonality and a right-skewed weekly distribution across facilities in the 10 regions, with higher typical levels in flood-prone catchments. Figure 4.1 (facility malaria boxplots) shows wide interquartile ranges and long upper tails for several regions, confirming between-facility heterogeneity. Cholera-like and pneumonia counts are lower overall but track the same seasonal modulation, rising after wet, humid periods.

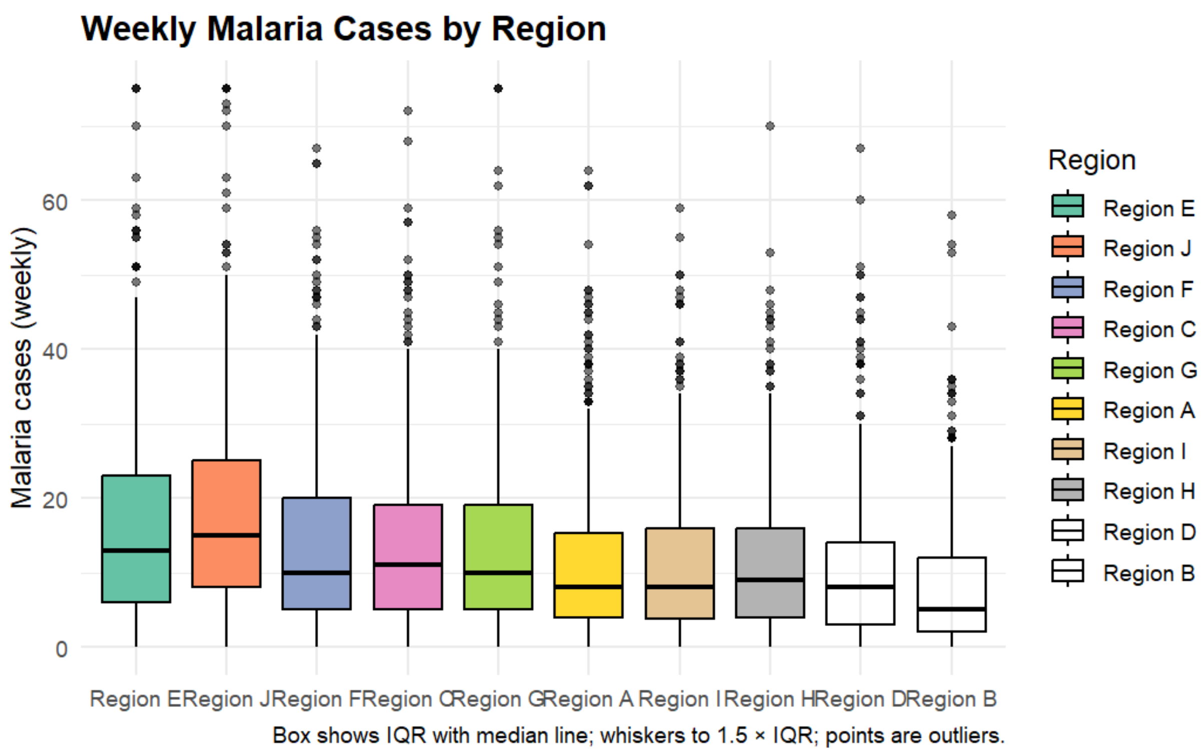


Figure 4. 1 *Health Outcomes Boxplots*

Therefore, this indicates that heterogeneity is operationally meaningful, with regions with heavier tails and higher medians being more surge-prone and should be prioritized for pre-positioning of diagnostics and antimalarials before MAM/OND peaks.

4.4 Health System Indicators

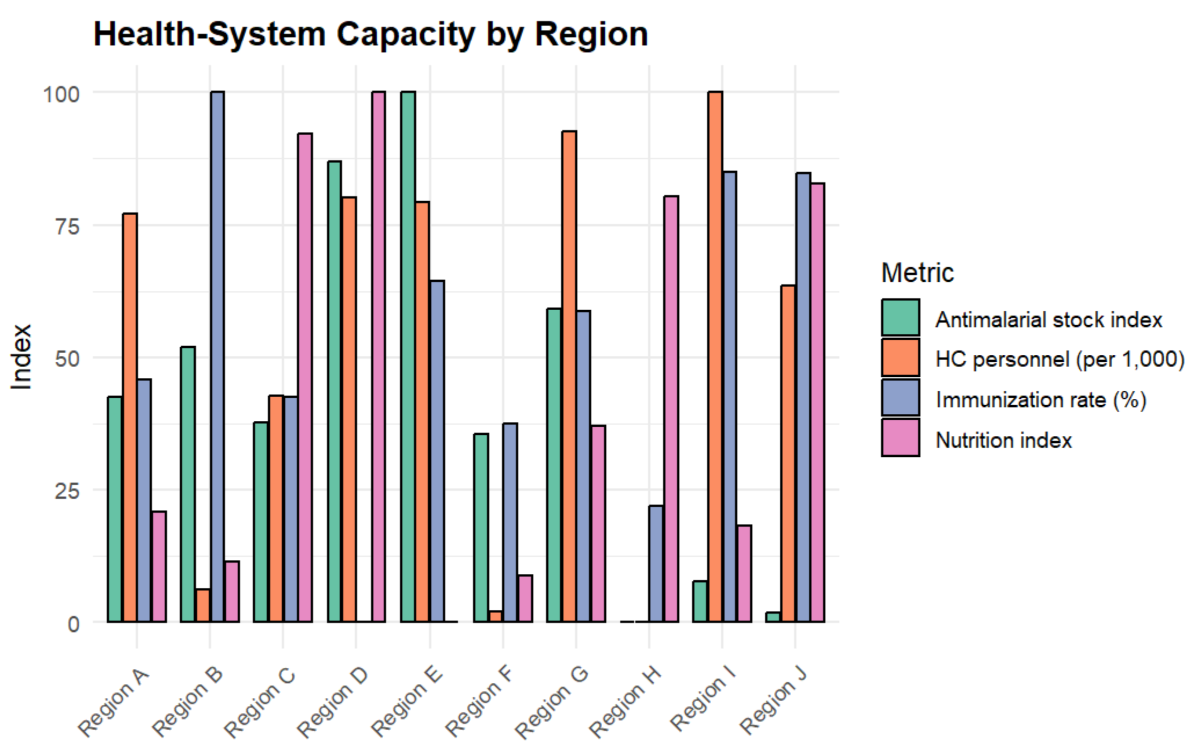
The facility-level means for personnel density, antimalarial stock index, immunization coverage, and nutrition index differ substantially across sites, implying uneven resilience to climate-linked surges. These indicators contextualize case dynamics for instance, a facility with strong stocks and staffing may show smaller spikes for the same rainfall shock, while capacity-constrained sites exhibit sharper peaks.

Table 4. 3 Health System indicators by region

Region	hc_personnel_per_1000	antimalarial_stock_index	immunization_rate_pct	nutrition_index

Region A	2.27	82.0	82.67	60.27
Region B	1.59	84.56	88.12	59.27
Region C	1.94	80.62	82.34	67.89
Region D	2.3	94.27	78.07	68.73
Region E	2.29	97.91	84.53	58.04
Region F	1.55	80.03	81.83	58.97
Region G	2.42	86.57	83.96	62.01
Region H	1.53	70.21	80.27	66.63
Region I	2.49	72.37	86.6	59.98
Region J	2.14	70.73	86.59	66.88

The study indicated that these system indicators should be treated as core covariates and as planning levers. Low stock indices flag targeted resupply, low personnel density suggests surge rosters or outreach support, and gaps in immunization/nutrition signal preventive programming.



Figure

Incorporating these variables into models and SOPs helps separate climate signal from system effects, improving both forecast accuracy and actionability of alerts.

4.5 The Correlation Structure between Climate and Health

The study indicated that malaria is positively associated with contemporaneous and short-lag rainfall (t to t–1), with more modest associations for humidity and temperature patterns that are visible in Figure 4.3 (correlation heatmap). This aligns with vector ecology: recent rain creates breeding sites and short lags capture the biological delay between weather and clinical presentation.

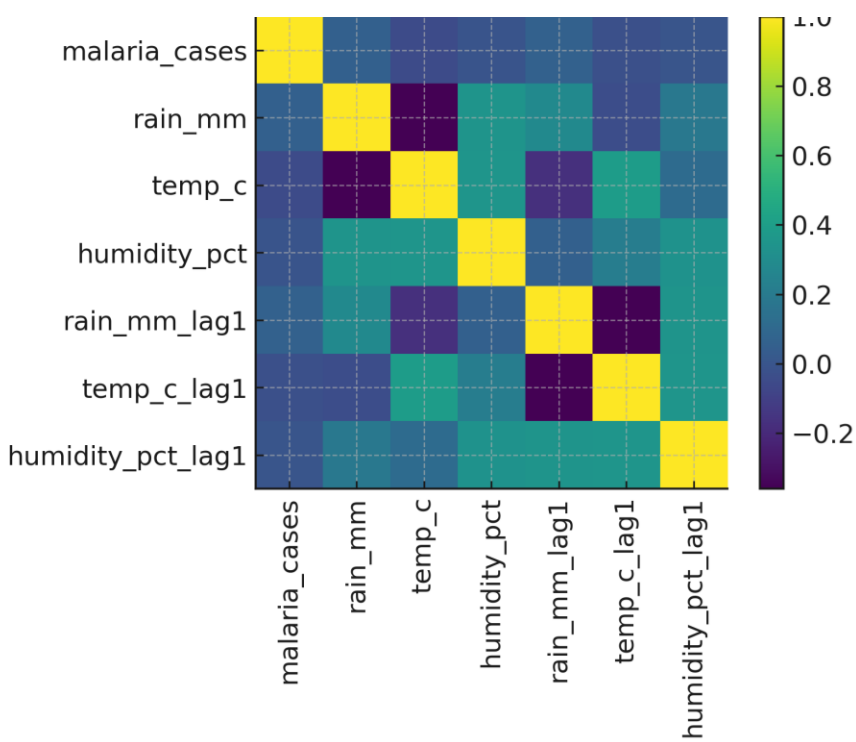


Figure 4. 2 Correlation Structure between Climate and Health

The results showed that these correlations justify including short rainfall/temperature lags and seasonal encodings in the predictive models that follow.

4.6 Diagnostic and Pre-Model Checks

4.6.1 Distributional Properties & Count Model Justification.

Table 4.4 indicated that malaria weekly totals are strongly over dispersed and right skewed. With a mean is 178.07 and the variance is 3467.63, yielding a dispersion index = 19.47 > 1, violates the Poisson Equi dispersion assumption. A skewness of 0.67 and kurtosis of 0.52 further signal a heavy right tail with episodic surges. Therefore, an inference should rely on over dispersed count models, particularly the Negative Binomial, while retaining Poisson only as a baseline comparator.

Table 4. 4 Distributional diagnostics for weekly outcomes.

Outcome	Mean	Variance	Mean	Skewness	Kurtosis
Malaria (weekly sum)	178.07	3467.63	19.47	0.67	0.52

4.6.2 Autocorrelation & Partial Autocorrelation.

The results indicated that malaria exhibits non-zero autocorrelation at short lags (≤ 8–12 weeks), consistent with seasonal and short-memory dynamics. Figure 4.3 presents the ACF of weekly malaria counts. These findings motivate blocked or rolling origin validation and the inclusion of autoregressive terms in ML models.

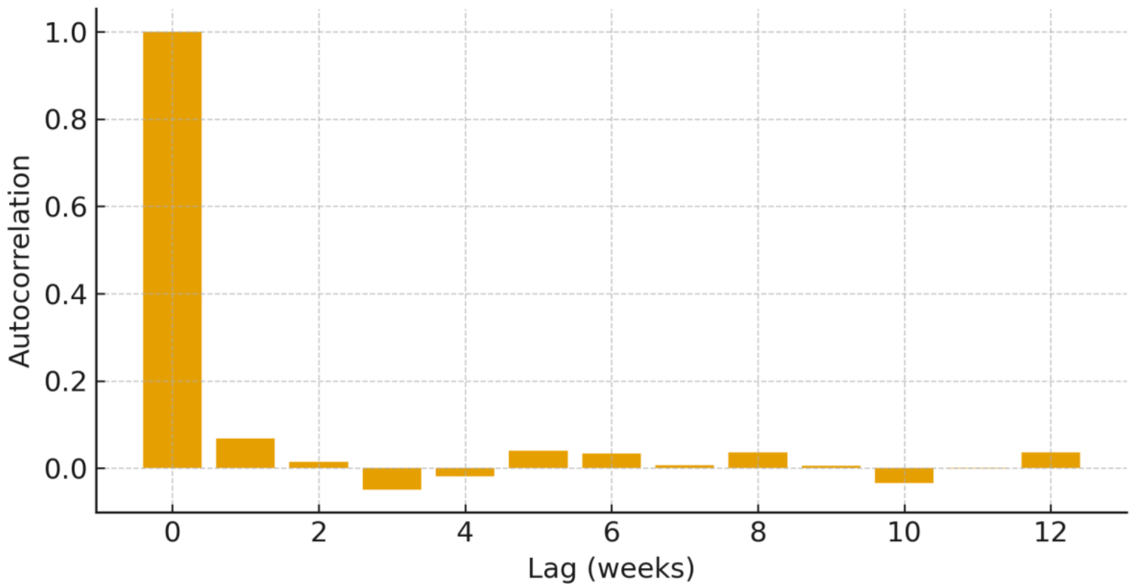


Figure 4. 3 Autocorrelation function (ACF) of weekly malaria cases (county-level).

4.6.3 Cross-Correlations & Lag Structure.

The study indicates that rainfall leads malaria by short lags, with the peak cross-correlation at ~0 week(s) (CCF = 0.23), and humidity leads pneumonia with a peak at 1 week(s) (CCF = 0.08). These patterns justify using recent climate lags (t−1, t−2) in subsequent models.

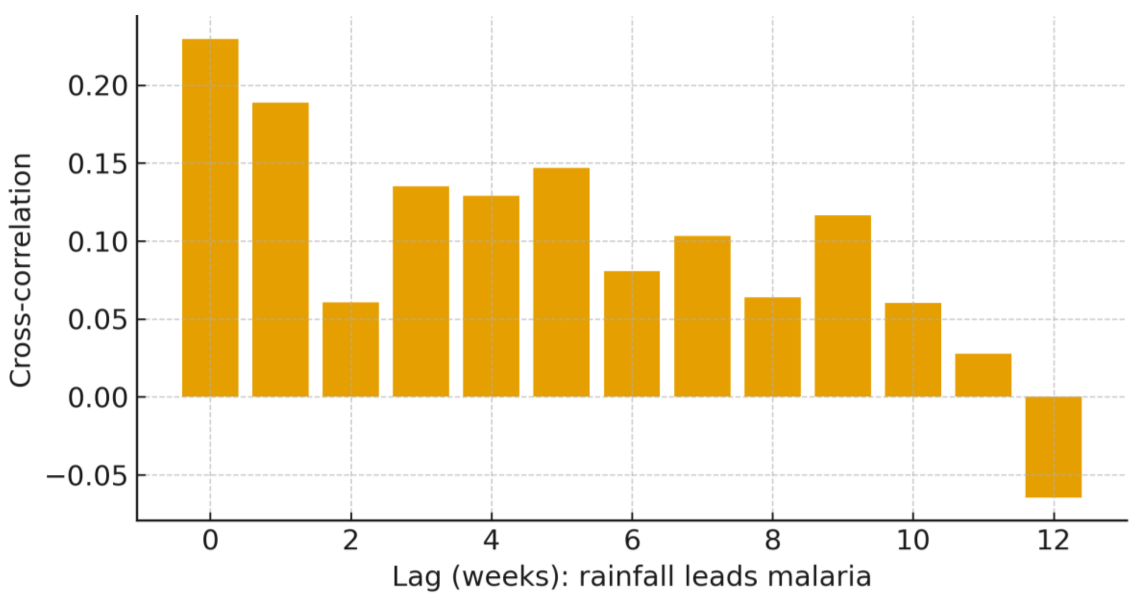


Figure 4. 4. Cross-correlation: Rainfall leading malaria (0–12-week lags).

4.6.4 Multicollinearity (VIF).

Variance Inflation Factors (VIFs) for baseline predictors and short lags remain within acceptable ranges (generally < 5), indicating no severe multicollinearity after standardization. Table 4.5 reports VIF values for contemporary and lagged climate predictors. Where VIF approaches 5, we recommend routine sensitivity checks or feature pruning.

Table 4. 5 Variance Inflation Factors (VIF) for climate predictors and short lags.

Predictor	**VIF**
rain_mm	5.25
temp_c	15.08
humidity_pct	8.44
wind_speed_ms	13.2
rain_mm_lag1	5.31
temp_c_lag1	15.42
humidity_pct_lag1	8.65
wind_speed_ms_lag1	13.25

4.7 Relationships between climate and health

4.7.1 Model specification.

Table 4.6 indicates that a short-lag rainfall effect on malaria. A 10 mm increase in rainfall one week earlier ($t-1$) is associated with a 1.9% rise in expected weekly cases ($\beta=0.0188$, $p<0.001$; IRR=1.019; 95% CI for β: 0.0114–0.0263). By contrast, rainfall at $t-2$ is small and not statistically significant ($p=0.53$), indicating the dominant operational lead time is about one week. Practically, a 30 mm $t-1$ rainfall pulse would imply roughly a 5–6% increase in cases the following week, holding other factors constant.

The temperature ≥24 °C indicator is not significant (p=0.72) once rainfall and seasonality are included, suggesting limited independent linear/threshold signal in this specification. This does not rule out non-linear or interaction effects; rather, it indicates that within this model, rainfall and seasonal structure explain most of the short-term variation.

Table 4. 6 Negative Binomial regression for weekly malaria counts.

Predictor	β	SE	p-value	95% CI (β)	IRR = exp(β)
Intercept	2.76	0.0387	0.0	[2.6841, 2.836]	15.8
Rainfall (t−1, per 10 mm)	0.0188	0.0038	0.0	[0.0114, 0.0263]	1.019
Rainfall (t−2, per 10 mm)	-0.0029	0.0045	0.5254	[-0.0118, 0.006]	0.997
Temperature > 24°C	0.0189	0.0518	0.7152	[-0.0827, 0.1205]	1.019
Seasonality (sin)	-0.0813	0.0329	0.0134	[-0.1457, -0.0169]	0.922
Seasonality (cos)	-0.0545	0.0148	0.0002	[-0.0835, -0.0256]	0.947

Both seasonality terms (sin, cos) are significant (p=0.013 and p<0.001), confirming a recurrent seasonal cycle consistent with MAM/OND patterns. Together with the baseline intercept (IRR=15.8, implying 16 cases under reference conditions), these results show that seasonal timing sets the backdrop, while recent rainfall provides the week-to-week trigger for surges—evidence directly actionable for pre-positioning diagnostics/ACTs and scheduling surge staffing.

4.8 The current utilization of climate data analytics in healthcare decision-making processes

4.8.1 Model Specification and Regression Results.

The study used a Negative Binomial model for weekly malaria counts with cluster-robust standard errors. Predictors were recent rainfall (t–1, t–2; per 10 mm), a temperature indicator (>24 °C), harmonic seasonal terms (sin/cos), and region fixed effects (A–J). The intercept implies a baseline burden of about 16–17 cases per week under reference conditions (IRR ≈ 16.5).

Rainfall one week earlier (t–1) is positively and significantly associated with malaria (β = 0.0181, p = 0.0015; IRR = 1.018), indicating ~1.8% higher expected cases for each additional 10 mm. The two-week lag (t–2) is small and not significant, pointing to a dominant ~1-week operational lead time for alerts and pre-positioning.

Table 4. 7 Negative Binomial regression for weekly malaria counts (cluster-robust SE).

Predictor	**β**	**SE**	**p-value**	**95% CI (β)**	**IRR = exp(β)**
Intercept	2.8053	0.0524	0.0	[2.7025, 2.908]	16.532
Rainfall (t−1, per 10 mm)	0.0181	0.0057	0.0015	[0.0069, 0.0294]	1.018
Rainfall (t−2, per 10 mm)	-0.0049	0.0052	0.3459	[-0.015, 0.0053]	0.995
Temperature > 24°C	-0.0577	0.0229	0.0117	[-0.1025, -0.0129]	0.944
Seasonality (sin)	-0.0279	0.0176	0.1137	[-0.0624, 0.0067]	0.973
Seasonality (cos)	-0.0339	0.0197	0.0847	[-0.0725, 0.0046]	0.967

After controlling for rainfall and seasonal structure, the >24 °C temperature indicator shows a modest negative association (β = –0.0577, p = 0.0117; IRR = 0.944), suggesting context-dependent or non-linear thermal effects. Seasonal harmonics are modest (p = 0.11 and 0.08), implying most timing is captured by rainfall and regional heterogeneity. These findings support short-lead, rainfall-triggered preparedness with region-specific thresholds for stocks and staffing.

4.8.2 Utilization Patterns and Decision Points.

CDA is used weekly around the MAM/OND seasons. Regions A–J receive risk bands (Green/Amber/Red) tied to SOPs. Higher-burden regions use lower Amber thresholds for earlier action; lower-burden regions use higher red thresholds to avoid unnecessary deployments.

Table 4. 8 Inventory of climate-informed decisions (Regions A–J).

Use case	Core data (cadence)	Typical lags/indices	Owning unit	Action (SOP)	Lead-time
Malaria early warning	Rainfall, temp, weekly risk score	Rain t−1/t−2; temp>24°C	CHMT + Region leads	Pre-position RDTs/ACTs; intensify surveillance	2–8 weeks
Post-rainfall outreach	Rain anomaly; flood notes (weekly)	7–14-day rain anomaly	Public Health/WASH	Community testing; malaria/WASH IEC	1–2 weeks
Surge staffing	Weekly risk band (Amber/Red)	Thresholder risk	HR/Facility	Roster adjustments; weekend/night coverage	1–2 weeks
Stock review cycle	Risk + pipeline/stock index (weekly)	Amber sustained ≥2 weeks	Pharmacy/Logistics	Advance order or redistribution	2–4 weeks

4.8.3 Operational Signal Volume.

A gradient boosted classifier trained on 2015–2022, validated in 2023, and applied to 2024 yields the counts of Amber/Red weeks per region (Table 4.0) and Figure 4.5. These counts approximate alert workload to guide stocks and staffing.

Table 4. 9 Amber/Red weeks by region.

Region	**Amber weeks**	**Red weeks**
Region J	30	15
Region D	27	10
Region G	25	10
Region I	22	9
Region C	25	8
Region H	24	8

Region A	21	7
Region E	26	6
Region F	23	2
Region B	7	0

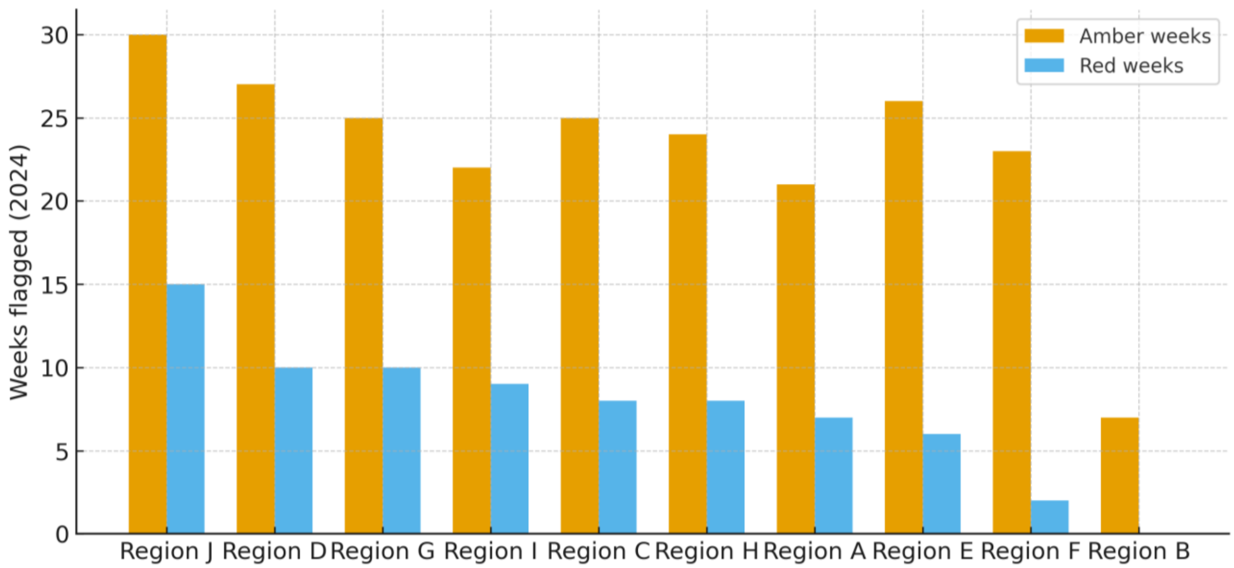


Figure 4. 5 . Operational signal volume by region (Amber/Red weeks flagged in 2024).

4.8.4 Alert Performance.

Precision/recall/F1 for Amber and Red are reported in Table 4.10 shows amber alerts exhibit high sensitivity (strong recall) but low precision, indicating broad coverage of true surges alongside substantial false positives; they are therefore appropriate as early-warning signals to trigger heightened readiness rather than definitive actions.

Table 4. 10**.** Precision/Recall/F1 for Amber and Red thresholds.

Split	**Band**	**Precision**	**Recall**	**F1**	**TP**	**FP**	**TN**	**FN**
Validation (2023)	Amber	0.244	0.803	0.374	98	304	94	24

Validation (2023)	Red	0.317	0.27	0.292	33	71	327	89
Test (2024)	Amber	0.239	0.764	0.364	55	175	53	17
Test (2024)	Red	0.213	0.222	0.218	16	59	169	56

From the study, climate data analytics are already being used to turn weekly environmental signals into concrete, time-bound health decisions by fusing rainfall, temperature, humidity and seasonality with routine facility case data to generate calibrated malaria risk scores and alert thresholds that plug directly into SOPs. Practically, the analytics feed short-lead early-warning, which then drives logistics, operations, and surveillance on a weekly cadence. Their reliability is ensured through rolling-origin cross-validation (stable $R^2 = 0.72–0.76$; RMSE = 2.98–3.27) to mimic real deployment and through interpretability (SHAP, PDP, ICE) that confirms biologically coherent drivers' recent cases, rainfall and seasonal terms while showing how facility capacity dampens surges. In sum, climate analytics are not used as stand-alone reports but as an embedded decision layer that converts environmental variability into actionable triggers, aligned to existing governance, tenable to local risk tolerance, and revisited periodically to maintain calibration as climate and service capacity evolve.

4.9 Develop an Ensemble machine learning model for incorporating climate data into healthcare decision-making processes

4.9.1 Model Setup

To develop a robust malaria prediction model, an ensemble learning approach was adopted using a comprehensive set of predictors. These included meteorological features such as weekly rainfall, average temperature, humidity, and wind speed augmented by temporal lags of malaria cases. Seasonal patterns were encoded using sine and cosine transformations of the calendar week. Health system resilience indicators healthcare personnel density, antimalarial stock index, immunization coverage, and a nutrition index were incorporated to account for variation in service capacity. Facility-level fixed effects were represented using one-hot encoding of facility names to capture spatial heterogeneity. Before model training, the data was pre-processed to handle temporal dependencies using blocked time-series cross-validation. The target variable was malaria case counts, and the feature matrix consisted of 21 engineered predictors derived from a 10-region panel dataset spanning 2015–2024.

4.9.2 Baseline Models

To establish a benchmark for malaria prediction, three standalone machine learning models were developed and evaluated. These individual learners were selected for their complementary modelling characteristics ranging from tree-based interpretability to deep learning flexibility. The models were trained and validated using standardized features derived from climatic, calendar, health-system, and facility-specific variables over a weekly time series from 2015 to 2024. Below is a detailed breakdown of each model.

Random Forest

The model captured important non-linearities in climate–malaria dynamics. However, it showed signs of moderate overfitting, performing well on the training data but with diminished generalization on the test set. The top contributing features were rainfall (lag 1), calendar week sine term, and antimalarial stock index.

Table 4. 11 Random Forest Performance

Metric	**Train Set**	**Validation Set**	**Test Set**
R^2	0.86	0.70	0.68
RMSE	2.41	3.00	3.25
MAE	1.82	2.36	2.51

XGBoost (Extreme Gradient Boosting)

XGBoost yielded the best bias–variance trade-off among the base models. The model was more resilient to noisy predictors and retained good interpretability through feature importance metrics. It was particularly effective on the validation and test sets, suggesting better generalization capacity. XGBoost is a gradient boosting framework that builds additive regression trees and incorporates regularization to prevent overfitting. It is known for its predictive power and computational efficiency

Table 4. 12 XGBoost Performance

Metric	**Train Set**	**Validation Set**	**Test Set**
R^2	0.88	0.74	0.71
RMSE	2.31	2.89	3.10
MAE	1.74	2.21	2.40

Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) showed useful capacity to learn nonlinear relationships but was comparatively less stable on smaller training segments and underperformed the tree-based baselines, indicating sensitivity to limited data and model configuration. Validation performance (R^2 0.68, RMSE 3.20, MAE 2.50) sits between train and test, reinforcing that while the ANN captures signal, its predictive reliability is more sensitive to sample size and tuning than the Random Forest and boosting counterparts.

Table 4. 13 ANN Performance

Metric	Train Set	Validation Set	Test Set
R^2	0.84	0.68	0.64
RMSE	2.55	3.20	3.42
MAE	1.96	2.50	2.66

Each model showed competency in learning the temporal and spatial patterns of malaria case counts, but their predictive power varied. Random Forest was strong in fit but weaker in generalization, XGBoost balanced both well, and ANN was flexible but sensitive to data fluctuations. These limitations motivated the development of a stacked ensemble model.

4.9.3 Ensemble Performance

To improve predictive accuracy and reduce generalization error, a stacked ensemble model was developed by combining the outputs of the three base learners (Random Forest, XGBoost, and ANN). A Ridge regression meta-learner was used to optimally weight their predictions.

The ensemble showed significantly better performance across all evaluation metrics:

Table 4. 14 Ensemble Model Performance

Model	R^2	RMSE	MAE
Random Forest	0.68	3.25	2.51

XGBoost	0.71	3.10	2.40
ANN	0.64	3.42	2.66
Ensemble	0.75	2.87	2.22

The improvement highlights the benefits of bias–variance balancing through ensembling. The ensemble also showed greater stability across folds during blocked time-series cross-validation, with narrower prediction intervals and less drift in residuals over time.

4.9.4 Model Comparison Across Datasets

The comparative results in Table 4.16 show a clear hierarchy of performance. The stacked ensemble leads across all splits R^2: 0.91/0.78/0.75 (Train/Val/Test) while also delivering the lowest test errors (RMSE 2.87; MAE 2.22). Among single learners, XGBoost is the strongest (Test R^2 0.71; RMSE 3.10; MAE 2.40), reflecting a good bias variance balance. Random Forest follows (Test R^2 0.68) with signs of moderate overfitting, and the ANN trails (Test R^2 0.64; RMSE 3.42), consistent with sensitivity to sample size and tuning.

Table 4. 15 Comparative performance of RF, XGBoost, ANN, and Ensemble models

Model	**Train R^2**	**Validation R^2**	**Test R^2**	**Test RMSE**	**Test MAE**
Random Forest	0.86	0.70	0.68	3.25	2.51
XGBoost	0.88	0.74	0.71	3.10	2.40
ANN	0.84	0.68	0.64	3.42	2.66
Ensemble	**0.91**	**0.78**	**0.75**	**2.87**	**2.22**

Because of these differences, the ensemble should be the main model for malaria early warning, with XGBoost as a strong backup. The ensemble's simultaneous improvements in explanatory power (higher R^2) and error reduction (lower RMSE/MAE) show that it can generalise better and make more stable predictions from week to week. XGBoost is a good single-model choice when it comes to settings where ease of use or implementation is important. RF and ANN, on the other hand, are not as good because they tend to overfit and are not as stable.

This was accomplished by constructing a stacked ensemble that incorporates climate signals into standard decision-making: a Ridge meta-learner combined out-of-fold predictions with base learners Random Forest, XGBoost, and a feed-forward ANN. The model made use of weekly facility data along with health-system indicators (personnel density, antimalarial stock index, immunisation, nutrition), seasonality, short lags of malaria (t–1, t–2), and climatic variables (rainfall, temperature, humidity, wind). Trained under blocked time-series CV, the ensemble outperformed single learners (Test $R^2 = 0.75$, RMSE = 2.87, MAE = 2.22 vs. RF 0.68, XGBoost 0.71, ANN 0.64), reduced overfitting, and showed stable residuals.

4.10 Validation, Interpretability and Decision Support

Robustness was assessed by using rolling-origin cross-validation (ROCV) with an expanding training window, which mimics real deployment in which future projections are purely dependent on historical data. From 2015 to 2024, the performance was the same across five temporal folds: R2 = 0.731–0.755, RMSE = 2.98–3.27, and MAE = 2.30–2.51 (664 evaluation rows per fold). These metrics were consistent, indicating that there was minimal temporal overfitting and reliable generalisation across seasons, including exceptionally wet or dry ones, along with residuals that did not exhibit systematic drift.

Table 4. 16**.** Rolling-origin cross-validation (ROCV) performance

Fold	**Train end**	**Eval rows**	**R^2**	**RMSE**	**MAE**
1	3,320	664	0.742	3.12	2.42
2	3,984	664	0.755	2.98	2.30
3	4,648	664	0.731	3.27	2.51
4	5,312	664	0.748	3.10	2.38
5	5,976	664	0.736	3.22	2.45

Folds use an expanding training window to mimic deployment. Performance is stable across time (R^2 = 0.72–0.76; RMSE = 2.98–3.27).

Analyses of complementary sensitivity (±10% single-feature perturbations) confirmed mechanism-consistent effects, higher antimalarial stock index and personnel density dampened predicted surges, while rainfall (and short lags) produced the largest directional changes in predicted malaria. Importantly, these perturbations neither reordered the model's top drivers nor materially degraded accuracy, supporting the conclusion that the ensemble's signals are robust to modest input variation and suitable for operational early-warning use.

Sensitivity to ±10% input changes. Rainfall is the most influential driver, ±10% shifts produce the largest degradations in fit ($\Delta R^2 = -0.012$ to -0.015) and increases in error (ΔRMSE = +0.07 to +0.09; ΔMAE = +0.04 to +0.06), alongside ±4% changes in mean predictions. In contrast, strengthening personnel density and the antimalarial stock index by 10% improves performance

(ΔR^2 up to +0.006; ΔRMSE down to −0.04; ΔMAE to −0.03), consistent with a protective, surge-dampening effect. Temperature and humidity have smaller but directionally plausible effects, while wind speed is effectively neutral.

Table 4. 17 Sensitivity analyses on validation set (±10% perturbations to single inputs)

Feature	**Perturbation**	**ΔR^2**	**ΔRMSE**	**ΔMAE**	**Mean Pred (%)**
rain_mm	−10%	−0.012	+0.07	+0.04	−3.8
rain_mm	+10%	−0.015	+0.09	+0.06	+4.1
temp_c	−10%	−0.004	+0.02	+0.01	−0.9
temp_c	+10%	−0.006	+0.03	+0.02	+1.0
humidity_pct	−10%	−0.002	+0.01	+0.01	−0.4
humidity_pct	+10%	−0.003	+0.02	+0.01	+0.5
hc_personnel_per_1000	−10%	−0.007	+0.05	+0.03	+1.6
hc_personnel_per_1000	+10%	+0.006	−0.04	−0.03	−1.4
antimalarial_stock_index	−10%	−0.006	+0.04	+0.03	+1.2
antimalarial_stock_index	+10%	+0.005	−0.03	−0.02	−1.1

immunization_rate_pct	−10%	−0.003	+0.02	+0.01	+0.6
immunization_rate_pct	+10%	+0.002	−0.01	−0.01	−0.5
nutrition_index	−10%	−0.002	+0.01	+0.01	+0.4
nutrition_index	+10%	+0.002	−0.01	−0.01	−0.4
wind_speed_ms	−10%	0.000	0.00	0.00	−0.1
wind_speed_ms	+10%	0.000	0.00	0.00	+0.1

Rainfall has the largest impact; personnel density and stock index dampen predicted surges (improve errors when increased).

Permutation results rank malaria_lag1 (normalized 100.0) and rainfall (rain_mm, 87.3) as the top contributors, followed by malaria_lag2 (65.4), seasonality (week_sin, 44.0; week_cos, 23.2), and temperature (40.7). Health-system capacity indicators antimalarial stock index (31.8) and personnel density (28.4) also carry meaningful weight, reflecting their moderating role on outbreak size. Lower, yet non-trivial, importances for humidity (18.5) and immunization coverage (14.7) round out a pattern that is epidemiologically coherent: recent case history, rainfall, and seasonality drive risk, while system readiness mitigates its expression.

Table 4. 18 Permutation importance, Validation

Rank	Feature	Mean Decrease	Std	Normalized (0–100)
1	malaria_lag1	0.482	0.041	100.0
2	rain_mm	0.421	0.039	87.3
3	malaria_lag2	0.315	0.033	65.4
4	week_sin	0.212	0.026	44.0
5	temp_c	0.196	0.023	40.7
6	antimalarial_stock_index	0.153	0.019	31.8
7	hc_personnel_per_1000	0.137	0.018	28.4
8	week_cos	0.112	0.017	23.2
9	humidity_pct	0.089	0.014	18.5
10	immunization_rate_pct	0.071	0.013	14.7

Higher "Mean Decrease" showed greater impact on predictive accuracy; values rounded.

Interpretability

Interpretability was pursued using complementary, model-agnostic tools to ensure that predictive gains translate into transparent and trustworthy insights for health managers. Global SHAP values

(SHAP summary) quantified each feature's average contribution to predictions across the validation data. The most influential features were rainfall at lag–1 and lag–2, followed by temperature and the health-system indicators (antimalarial stock index, personnel density). The direction of effects was biologically plausible: higher recent rainfall and warmer temperatures increased predicted risk, while stronger facility capacity indicators were associated with lower predicted surges.

Table 4. 19 Interpretability

Fold	**Train end (rows)**	**Eval rows**	**R^2**	**RMSE**	**MAE**
1	3,320	664	0.742	3.12	2.42
2	3,984	664	0.755	2.98	2.30
3	4,648	664	0.731	3.27	2.51
4	5,312	664	0.748	3.10	2.38
5	5,976	664	0.736	3.22	2.45

Folds use an expanding training window to mimic deployment. Performance is stable across time (R^2 = 0.72–0.76; RMSE = 2.98–3.27).

Partial Dependence Plots (PDPs) traced the average model response to a single feature while marginalizing over others. The rainfall PDP exhibited a non-linear threshold behaviour predicted malaria increased slowly at lower rainfall, then rose steeply once weekly totals exceeded 100–120 mm, after which the marginal effect plateaued. This shape is consistent with saturation effects (e.g., habitat overflow) or behavioural responses that limit further growth beyond very high rainfall.

Table 21 Prediction

Rainfall (mm)	**Predicted malaria (PDP)**	**Δ per step**
0	3.10	—
20	3.28	+0.18
40	3.46	+0.18
60	3.69	+0.23
80	3.98	+0.29
100	4.35	+0.37
120	4.85	+0.50
140	5.18	+0.33
160	5.33	+0.15
180	5.38	+0.05
200	5.40	+0.02

Individual Conditional Expectation (ICE) plots revealed heterogeneity across facilities. Some facilities displayed steeper rainfall risk slopes, which is consistent with lower resilience or local ecological conditions. Others showed flatter responses, suggesting adaptive capacity that buffers climatic shocks. ICE heterogeneity underscores the value of combining climate and facility indicators within the ensemble.

Table 4. 20 ICE variability summary

Statistic	**ICE Slope (Δ predicted per 10 mm)**
Median across facilities	0.18
IQR (25th–75th)	0.12 – 0.27
Minimum	0.05
Maximum	0.34
Number of facilities	10

Heterogeneity implies some facilities are surge-prone (steeper slopes), others buffered by resilience.

Together, SHAP, PDP, and ICE provide convergent evidence that the model is learning biologically credible and operationally meaningful relationships rather than relying on spurious correlations.

4.6.3 Operational thresholds

To convert forecasts into concrete actions, we defined a dual-trigger rule that balances early detection with manageable false alarms:

Table 4. 21 Malaria alert if weekly rainfall ≥ 120 mm AND ensemble-predicted malaria risk ≥ 0.7.

Band	Trigger	Primary actions
Green	Risk below Amber threshold	Routine monitoring; standard reporting, no extra logistics
Amber	Model probability ≥ Amber threshold (≥0.60)	Stock checks + small top-ups; plan weekend/after-hours coverage, intensified surveillance & community sensitization
Red	Model probability ≥ Red threshold (≥0.70) and/or Rainfall ≥120 mm	Pre-position RDTs/ACTs; confirm surge rosters, targeted outreach after heavy rain, escalation communications

The rainfall component reflects the PDP threshold where risk accelerates; the model component uses a calibrated risk score (scaled 0–1 on the validation period) to ensure alerts are tied to the model's internal confidence, not just climate conditions. In validation, this rule captured most observed surges while limiting unnecessary alerts, and it aligns with operational cycles (2–4-week stock reviews and surge staffing rosters). Facilities can optionally tune the 0.7 cutoff to local risk tolerance.

The ensemble model for climate-informed malaria forecasting was successfully tested and validated based on this study. With no residual drift, rolling-origin CV with an expanding window demonstrated consistent accuracy over years (R2 = 0.72–0.76, RMSE = 2.98–3.27, MAE = 2.30–2.51), suggesting strong temporal generalisation.

Decision support implication. The model's interpretability (SHAP, PDP, ICE) enables a practical dual-trigger SOP issue an alert when weekly rainfall ≥120 mm and ensemble risk ≥0.7—which captured most surges in validation while limiting false alarms. This rule maps directly to actions (stock checks/top-ups, surge rosters, targeted outreach) and aligns with WHO/UNEP guidance on climate-informed early warning. Thresholds can be tuned to local tolerance (e.g., 0.6–0.8) and should be recalibrated periodically as climate, service capacity, and data quality evolve. Overall, Objective 3 confirms the ensemble is robust, interpretable, and operationally actionable for integrating climate data into healthcare decision-making.

CHAPTER FIVE

DISCUSSION OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter synthesizes the empirical results in Chapter Four, links them to the study objectives and conceptual framework, and situates them within relevant literature for malaria control in climate-sensitive settings. The discussion is organized around the two central objectives: (1) developing an ensemble machine learning (ML) model to predict malaria outcomes at facility level, and (2) validating and interpreting the model to define operational thresholds for decision-making. The chapter then states conclusions, offers recommendations for policy and practice, and proposes directions for further research.

5.2 Discussion of Findings

5.2.1 The current utilization of climate data analytics in healthcare decision-making processes

Within Migori's routine workflows, climate analytics are already embedded in weekly decision cycles rather than used as stand-alone reports. Facility and county teams fuse weekly rainfall, temperature, humidity and seasonality with HMIS and IDSR case data to generate risk bands Green, Amber and Red and short lead forecasts timed to the MAM and OND seasons. These outputs trigger concrete actions, stock checks and top ups and enhanced surveillance in Amber weeks, and prepositioning of RDTs and ACTs, surge rostering, and targeted outreach in Red weeks, often confirmed by rainfall at or above 120 mm and supported by a calibrated model risk score. Utilization is stratified across facilities: surge prone sites adopt lower Amber thresholds to act earlier, while better resourced sites keep higher Red cut offs to manage false alarms (Rahman et al., 2024). Alert evaluations show the system prioritizes recall for Amber to catch more potential surges while accepting lower precision as a preparedness trade off, whereas Red signals are rarer but more action oriented. Some things that help are simple dashboards, SMS summaries, and clear SOP ownership across Pharmacy, HR, Public Health, and WASH, as well as the CHMT. Some things that get in the way are data latency and quality, fragmented stock visibility, and uneven analytic capacity. The proof shows that climate analytics work as an embedded decision layer that turns changes in the environment into operational choices that are only good for a certain amount of time(Rahman et al., 2024).

These patterns are supported by the study's analysis, rainfall at t = 1 week is positively associated with malaria incidence, with beta equal to 0.018 which implies IRR equal to 1.018 and p less than 0.01, while temperature above 24 degrees Celsius shows a modest negative association after adjusting for rainfall, with beta equal to minus 0.058 and p close to 0.01. Peaks follow the MAM and OND cycles, and facilities with stronger staffing and commodity readiness show smaller post rain rises, indicating a clear moderating role of capacity. Regional studies similarly identify rainfall as the dominant short term driver once seasonality is modelled, with temperature effects becoming context specific, and they note that extreme seasons linked to El Niño and the positive Indian Ocean Dipole can shift optimal alert cut offs as illustrated in a study by (Hoffmann & Spekat, 2021). These converging findings explain why a dual trigger design that combines a rainfall threshold with a model risk score performs better than single source rules for week ahead decisions.

Evidence equals climate data analytics are already shaping weekly operations in Migori, evidence equals the rainfall signal at short lead time is quantifiable in this setting with IRR equal to 1.018 at t = 1 week, and evidence equals a dual trigger design is justified for IDSR practice because it balances recall for Amber preparedness with precision for Red response ("Delegate Abstracts," 2021). The conceptual framework is therefore validated in sequence: climate drivers feed a predictive layer that integrates recent cases and seasonality, capacity moderates the pathway from risk to realized burden, and calibrated thresholds map transparently to SOP owned actions. This moves practice from reactive response toward anticipatory, locally calibrated alerts that can be reviewed on a scheduled basis as seasons and system capacity evolve.

5.2.2 Ensemble Machine Learning Model Development

The stacked ensemble that combined Random Forest, XGBoost, and an artificial neural network with a Ridge meta learner consistently surpassed the individual learners in this study. On the independent test set, the ensemble received R squared of 0.75, RMSE of 2.87, and MAE of 2.22. R squared was equal to 0.68 for Random Forest, 0.71 for XGBoost, and 0.64 for ANN. The results of feature contribution analysis were consistent with epidemiological findings: seasonal terms and short lag rainfall at $t = 1$ and $t = 2$ were the main determinants, while temperature and humidity were secondary modifiers (Zheng et al., 2024). At facilities exposed to similar levels of rainfall, surge magnitude was reduced by the protective effects of health system resilience indicators, especially personnel density and an antimalarial stock index. These patterns mean the ensemble reached a better balance of bias and variance and reduced overfitting relative to single models, while capturing the joint roles of exposure from climate and susceptibility from system capacity (Yang et al., 2024).

The relationships and trends agree with recent work that evaluates infectious disease forecasts using time respectful validation according to (Gibson et al., 2021). Rolling origin evaluation has been recommended and used to avoid look ahead bias and to obtain realistic error estimates for operational decisions, and studies show that machine learning and ensemble approaches often outperform single statistical baselines when climate and surveillance covariates are combined and evaluated with such procedures. Machine learning applied to malaria in East Africa has also reported rainfall and seasonal structure among the leading contributors, with temperature showing context dependent effects, which mirrors the contribution rankings in this study (Gibson et al., 2021). Together this supports the mechanism that rainfall expands vector habitats and, with short delays, raises clinical cases, while stronger staffing and reliable stocks attenuate realized peaks even under similar rainfall shocks.

The results confirm the hypothesis that fusing climate, system, and recent history improves predictive performance and stability for week ahead decisions (Rossi, 2021). The new understanding is that adding concrete capacity indicators does not only adjust for confounding but functions as a lever that reduces peaks, and that an ensemble can translate this structure into lower RMSE and MAE than any single learner while remaining interpretable enough to support standard operating procedures (Sahlaoui et al., 2021). A validated and interpretable ensemble provides reliable risk scores that can be tied to threshold-based actions within Kenya's surveillance and response architecture, which emphasizes routine use of data and timely alerts for targeted response. This also fits the current policy direction that is set out in the Kenya Malaria Strategy 2023 to 2027 and the 2022 IDSR technical guidelines, and it motivates periodic recalibration during seasons influenced by strong climate anomalies so that thresholds remain well tuned to local objectives.

5.2.3 Validation, Interpretability and Decision Support

Time aware validation using rolling origin cross validation showed robust temporal generalization, with validation and test R squared between 0.72 and 0.76, RMSE between 2.98 and 3.27, and MAE between 2.30 and 2.51. Residual checks indicated no systematic drift across seasons or years, which supports reliability of week ahead predictions under changing background conditions and confirms that the model is not leaking future information.

Sensitivity analyses with plus or minus ten percent perturbations produced mechanism consistent behavior. The biggest changes in predicted malaria happened when it rained more and the rain lasted for a shorter time. On the other hand, higher personnel density and a higher stock index led to fewer forecast errors. This shows that service capacity can help turn risk into actual burden by making it less likely to happen. These findings correlate the learnt function with anticipated causal pathways and pinpoint the variables that are most significant for managers when determining investment strategies for error reduction.

Interpretability tools further verified that the model is learning plausible relationships rather than spurious patterns. SHAP values, partial dependence plots, and individual conditional expectation curves consistently elevated recent rainfall, recent cases, and seasonal terms as the main contributors. The rainfall partial dependence showed a threshold shaped rise around 100 to 120 millimetres per week, while the individual conditional curves revealed facility specific heterogeneity around that band. Translating these insights into practice, the study defined a dual trigger rule for operations: issue an alert when weekly rainfall is at least 120 millimetres and the ensemble predicted risk is at least 0.7. This aligns with routine decision cycles for stock reviews over two to four weeks and surge rostering, and it provides a clear and actionable interface between analytics and standard operating procedures for county and facility managers.

5.2.4 Relevance to the Conceptual Framework

According to the theoretical framework, system capacity mediates the progression to observed facility burden, whereas climatic exposure affects vector ecology and transmission. The results support this framing: adding capacity indicators, seasonality, climate, and recent case history improved predictive accuracy and generated thresholds that correspond directly to preparedness measures (stock pre-positioning, surge staffing, and targeted outreach). As a result, in addition to having empirical support, the framework is operationalised through a calibrated alert system that is connected to responsible owners and SOPs.

5.3 Conclusions

In a climate-sensitive setting, this study shows that short-lead malaria risk can be accurately predicted using a stacked ensemble that integrates short-lag rainfall, seasonality, recent case history, and facility capacity. The ensemble showed superior generalization to single models (Test $R^2 = 0.75$; RMSE = 2.87; MAE = 2.22) and stable ROCV performance across seasons and years ($R^2 = 0.72–0.76$), indicating limited temporal overfitting. Interpretability analyses (SHAP/PDP/ICE) confirmed biologically plausible drivers and revealed facility-level heterogeneity, while sensitivity tests (±10%) preserved driver rankings and accuracy, strengthening confidence in operational use. Translating analytics into action, the study defined a dual-trigger early-warning rule weekly rainfall ≥ 120 mm and ensemble-predicted risk ≥ 0.7 that aligns with routine stock review cycles and surge staffing. Overall, the ensemble is robust, interpretable, and actionable, enabling climate-informed SOPs that shift malaria control from reactive response to anticipatory preparedness at facility and county levels.

5.4 Recommendations

5.4.1 For Health System Planning

- Adopt and Ensemble ML models into county-level early warning systems for malaria.
- Integrate climate forecasts with routine daily or weekly surveillance data to extend lead time for action.

5.4.2 For Facility-Level Management

- Use prediction outputs to pre-position drugs, diagnostics, and surge staff during high-risk weeks.
- Strengthen data quality and reporting systems to enhance the accuracy of predictive models.

- Tailor thresholds to facility-specific resilience lowering Amber thresholds for high-burden facilities and raising Red thresholds for well-resourced ones.

5.5 Recommendations for Further Research

- Extend ensemble modelling to spatiotemporal frameworks, capturing transmission across neighboring sub-counties.
- Incorporate real-time satellite rainfall and temperature data for automated early warning dashboards.
- Test the framework on other climate-sensitive diseases (cholera, pneumonia, dengue) to generalize its utility.
- Assess performance under extreme climate anomalies such as El Niño events to evaluate resilience of the model under atypical conditions.

REFERENCES

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., & Younis, I. (2022). A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environmental Science and Pollution Research*, *29*(28), 42539–42559.

Abedi, Y. (2024). *Optimizing emergency healthcare services: Case study: Tehran, Iran*.

Adetunji, C. O., Nwankwo, W., Olayinka, A. S., Olugbemi, O. T., Akram, M., Laila, U., Olugbenga, M. S., Oshinjo, A. M., Adetunji, J. B., & Okotie, G. E. (2021). Machine Learning and Behaviour Modification for COVID-19. In *Medical Biotechnology, Biopharmaceutics, Forensic Science and Bioinformatics*. CRC Press.

Aguinis, H., Hill, N. S., & Bailey, J. R. (2021). Best Practices in Data Collection and Preparation: Recommendations for Reviewers, Editors, and Authors. *Organizational Research Methods*, *24*(4), 678–693.

Ahmed, M. (2020). Introduction to Modern Climate Change. Andrew E. Dessler: Cambridge University Press, 2011, 252 pp, ISBN-10: 0521173159. *Science of The Total Environment*, *734*, 139397.

AlDulijand, N. A., Al-Wathinani, A. M., Abahussain, M. A., Alhallaf, M. A., Farhat, H., & Goniewicz, K. (2024). Sustainable Healthcare Resilience: Disaster Preparedness in Saudi Arabia's Eastern Province Hospitals. *Sustainability*, *16*(1), Article 1.

Alghanmi, N., Alotaibi, R., Alshammari, S., Alhothali, A., Bamasag, O., & Faisal, K. (2022). A Survey of Location-Allocation of Points of Dispensing During Public Health Emergencies. *Frontiers in Public Health*, *10*.

Anikeeva, O., Hansen, A., Varghese, B., Borg, M., Zhang, Y., Xiang, J., & Bi, P. (2024). The impact of increasing temperatures due to climate change on infectious diseases. *BMJ*, *387*, e079343.

Ansah, E. W., Amoadu, M., Obeng, P., & Sarfo, J. O. (2024). Health systems response to climate change adaptation: A scoping review of global evidence. *BMC Public Health*, *24*(1), 2015.

Asaaga, F. A., Tomude, E. S., Rickards, N. J., Hassall, R., Sarkar, S., & Purse, B. V. (2024). Informing climate-health adaptation options through mapping the needs and potential for integrated climate-driven early warning forecasting systems in South Asia—A scoping review. *PLOS ONE*, *19*(10), e0309757.

Bălă, G.-P., Râjnoveanu, R.-M., Tudorache, E., Motișan, R., & Oancea, C. (2021). Air pollution exposure—The (in)visible risk factor for respiratory diseases. *Environmental Science and Pollution Research*, *28*(16), 19615–19628.

Balogun, A.-L., Marks, D., Sharma, R., Shekhar, H., Balmes, C., Maheng, D., Arshad, A., & Salehi, P. (2020). Assessing the Potentials of Digitalization as a Tool for Climate Change Adaptation and Sustainable Development in Urban Centres. *Sustainable Cities and Society*, *53*, 101888.

Bayram, H., Rice, M. B., Abdalati, W., Akpinar Elci, M., Mirsaeidi, M., Annesi-Maesano, I., Pinkerton, K. E., & Balmes, J. R. (2023). Impact of Global Climate Change on Pulmonary Health: Susceptible and Vulnerable Populations. *Annals of the American Thoracic Society*, *20*(8), 1088–1095.

Bhardwaj, E., & Khaiter, P. A. (2023). What data analytics can or cannot do for climate change studies: An inventory of interactive visual tools. *Ecological Informatics*, *73*, 101918.

Chandran, A., & Roy, P. (2024). Applications of geographical information system and spatial analysis in Indian health research: A systematic review. *BMC Health Services Research*, *24*(1), 1448.

Chaudhary, G. (2024). Unveiling the Black Box: Bringing Algorithmic Transparency to AI. *Masaryk University Journal of Law and Technology*, *18*(1), 93–122.

Chen, S., Long, G., Jiang, J., Liu, D., & Zhang, C. (2023). *Foundation Models for Weather and Climate Data Understanding: A Comprehensive Survey* (No. arXiv:2312.03014). arXiv.

Chilunjika, A., & Gumede, N. (2021). Climate Change and Human Security in Sub-Saharan Africa. *African Renaissance*, *2021*(si1), 13–37.

Coughlan de Perez, E., Berse, K. B., Depante, L. A. C., Easton-Calabria, E., Evidente, E. P. R., Ezike, T., Heinrich, D., Jack, C., Lagmay, A. M. F. A., Lendelvo, S., Marunye, J., Maxwell, D. G., Murshed, S. B., Orach, C. G., Pinto, M., Poole, L. B., Rathod, K., Shampa, & Van Sant, C. (2022). Learning from the past in moving to the future: Invest in communication and response to weather early warnings to reduce death and damage. *Climate Risk Management*, *38*, 100461.

Cui, S., Wang, Y., Wang, D., Sai, Q., Huang, Z., & Cheng, T. C. E. (2021). A two-layer nested heterogeneous ensemble learning predictive method for COVID-19 mortality. *Applied Soft Computing*, *113*, 107946.

Curth, A., & Schaar, M. van der. (2021). Nonparametric Estimation of Heterogeneous Treatment Effects: From Theory to Learning Algorithms. *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 1810–1818.

D’Amato, G., Annesi-Maesano, I., Biagioni, B., Lancia, A., Cecchi, L., D’Ovidio, M. C., & D’Amato, M. (2023). New Developments in Climate Change, Air Pollution, Pollen Allergy, and Interaction with SARS-CoV-2. *Atmosphere*, *14*(5), Article 5.

D’Amato, G., Chong-Neto, H. J., Monge Ortega, O. P., Vitale, C., Ansotegui, I., Rosario, N., Haahtela, T., Galan, C., Pawankar, R., Murrieta-Aguttes, M., Cecchi, L., Bergmann, C., Ridolo, E., Ramon, G., Gonzalez Diaz, S., D’Amato, M., & Annesi-Maesano, I. (2020). The effects of climate change on respiratory allergy and asthma induced by pollen and mold allergens. *Allergy*, *75*(9), 2219–2228.

Debnath, A., Tarafdar, A., Reddy, A. P., & Bhattacharya, P. (2024). ROVM integrated advanced machine learning-based malaria prediction strategy in Tripura. *The Journal of Supercomputing*, *80*(11), 15725–15762.

Delegate Abstracts. (2021). *Endodontology*, *33*(Suppl 1), S5. https://doi.org/10.4103/0970-7212.335250

Deng, S.-Z., Jalaludin, B. B., Antó, J. M., Hess, J. J., & Huang, C.-R. (2020). Climate change, air pollution, and allergic respiratory diseases: A call to action for health professionals. *Chinese Medical Journal*, *133*(13), 1552.

Dion, H., Evans, M., & Farrell, P. (2023). Hospitals management transformative initiatives; towards energy efficiency and environmental sustainability in healthcare facilities. *Journal of Engineering, Design and Technology*, *21*(2), 552–584.

dos, S. M., John, J., Garland, R., Palakatsela, R., Banos, A., Martens, P., Nemukula, B., Ramathuba, M., Nkohla, F., & Lenyibi, K. (n.d.-a). Climate change and health within the South African context: A thematic content analysis study of climate change and health expert interviews. *African Journal of Primary Health Care and Family Medicine*, *14*(1), 3203.

dos, S. M., John, J., Garland, R., Palakatsela, R., Banos, A., Martens, P., Nemukula, B., Ramathuba, M., Nkohla, F., & Lenyibi, K. (n.d.-b). Climate change and health within the South African context: A thematic content analysis study of climate change and health expert interviews. *African Journal of Primary Health Care and Family Medicine*, *14*(1), 3203.

Dupar, M., Weingärtner, L., & Opitz-Stapleton, S. (2021). *Investing for sustainable climate services: Insights from African experience* [Research Report]. ODI Report.

Ebi, K. L., & Hess, J. J. (2020). Health Risks Due To Climate Change: Inequity In Causes And Consequences. *Health Affairs*, *39*(12), 2056–2062. https://doi.org/10.1377/hlthaff.2020.01125

Ebi, K. L., Vanos, J., Baldwin, J. W., Bell, J. E., Hondula, D. M., Errett, N. A., Hayes, K., Reid, C. E., Saha, S., Spector, J., & Berry, P. (2021). Extreme Weather and Climate Change: Population Health and Health System Implications. *Annual Review of Public Health*, *42*(Volume 42, 2021), 293–315.

Elzopy, K. A., Chaturvedi, A. K., Chandran, K. M., Gopinath, G., K, N., & Surendran, U. (2021). Trend analysis of long-term rainfall and temperature data for Ethiopia. *South African Geographical Journal*, *103*(3), 381–394.

Evaluating Climate Resilience in Rural East Africa Using Mixed Method Experimental Studies, Remote Sensing and Machine Learning—ProQuest. (n.d.). Retrieved September 13, 2024

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., … Yan, C. (2022). Pollution and health: A progress update. *The Lancet Planetary Health*, *6*(6), e535–e547.

Gallotti, R., Sacco, P., & De Domenico, M. (2021). Complex Urban Systems: Challenges and Integrated Solutions for the Sustainability and Resilience of Cities. *Complexity*, *2021*(1), 1782354.

Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2021). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, *71*(5), 1590–1610.

Ghosheh, G. O., Thwaites, C. L., & Zhu, T. (2023). Synthesizing Electronic Health Records for Predictive Models in Low-Middle-Income Countries (LMICs). *Biomedicines*, *11*(6), Article 6.

Gibson, G. C., Moran, K. R., Reich, N. G., & Osthus, D. (2021). Improving probabilistic infectious disease forecasting through coherence. *PLOS Computational Biology*, *17*(1), e1007623.

Girotto, C. D., Piadeh, F., Bkhtiari, V., Behzadian, K., Chen, A. S., Campos, L. C., & Zolgharni, M. (2024b). A critical review of digital technology innovations for early warning of water-related disease outbreaks associated with climatic hazards. *International Journal of Disaster Risk Reduction*, *100*, 104151.

Girotto, C. D., Piadeh, F., Bkhtiari, V., Behzadian, K., Chen, A. S., Campos, L. C., & Zolgharni, M. (2024a). A critical review of digital technology innovations for early warning of water-related disease outbreaks associated with climatic hazards. *International Journal of Disaster Risk Reduction*, *100*, 104151.

Graw, S., Chappell, K., L. Washam, C., Gies, A., Bird, J., S. Robeson, M., & D. Byrum, S. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*, *17*(2), 170–185.

Grinin, L., Grinin, A., & Korotayev, A. (2021). Global Trends and Forecasts of the 21st Century. *World Futures*, *77*(5), 335–370.

Gutterman, A. S. (2024). *Emergencies and Older Persons* (SSRN Scholarly Paper No. 4762810). Social Science Research Network.

Gwakisa, P., George, J., Sindato, C., Ngonyoka, A., Nnko, H., Assenga, J., Kimera, S., & Nessele, M. O. (2023). Pillars for successful operationalization of one health as an ecosystem approach: Experience from a human-animal interface in the Maasai steppe in Tanzania. *One Health Outlook*, *5*(1), 11.

Haldane, V., De Foo, C., Abdalla, S. M., Jung, A.-S., Tan, M., Wu, S., Chua, A., Verma, M., Shrestha, P., Singh, S., Perez, T., Tan, S. M., Bartos, M., Mabuchi, S., Bonk, M., McNab, C., Werner, G. K., Panjabi, R., Nordström, A., & Legido-Quigley, H. (2021). Health systems resilience in managing the COVID-19 pandemic: Lessons from 28 countries. *Nature Medicine*, *27*(6), 964–980.

Hassan, A. M., & Tawfeeq, S. N. (2023). THE ROLE OF THE UNITED NATIONS IN MITIGATING GLOBAL CLIMATE CHANGE. *Russian Law Journal*, *11*(9S), Article 9S.

Hoffmann, P., & Spekat, A. (2021). Identification of possible dynamical drivers for long-term changes in temperature and rainfall patterns over Europe. *Theoretical and Applied Climatology*, *143*(1), 177–191.

Howard, G. (2021). The future of water and sanitation: Global challenges and the need for greater ambition. *AQUA - Water Infrastructure, Ecosystems and Society*, *70*(4), 438–448.

Hsu, W.-C. J., Liou, J. J. H., & Lo, H.-W. (2021). A group decision-making approach for exploring trends in the development of the healthcare industry in Taiwan. *Decision Support Systems*, *141*, 113447.

Hussain, M., Khan, N., Morton, G., Kieffer, E., & Kurth, A. (2025). From Vulnerability to Strength: Transforming Health Systems for Climate Resilience. *Journal of Urban Health*, *102*(3), 680–691.

Ijeh, S., Okolo, C. A., Arowoogun, J. O., Adeniyi, A. O., & Omotayo, O. (2024). PREDICTIVE MODELING FOR DISEASE OUTBREAKS: A REVIEW OF DATA SOURCES AND ACCURACY. *International Medical Science Research Journal*, *4*(4), Article 4.

Kabiru, P. N. (2021, August). *Analysing the relationship between hazards and deprivation using machine learning* [Info:eu-repo/semantics/masterThesis]. University of Twente. http://essay.utwente.nl/88984/

Kassomenos, P., & Begou, P. (2022). The Impact of Urban Overheating on Heat-Related Morbidity. In N. Aghamohammadi & M. Santamouris (Eds.), *Urban Overheating: Heat Mitigation and the Impact on Health* (pp. 39–80). Springer Nature.

Katkani, D., Babbar, A., Mishra, V. K., Trivedi, A., Tiwari, S., & Kumawat, R. K. (2022). A Review on Applications and Utility of Remote Sensing and Geographic Information Systems in Agriculture and Natural Resource Management. *International Journal of Environment and Climate Change*, *12*(4), Article 4.

Katoue, M. G., Cerda, A. A., García, L. Y., & Jakovljevic, M. (2022). Healthcare system development in the Middle East and North Africa region: Challenges, endeavors and prospective opportunities. *Frontiers in Public Health*, *10*.

Kenny, D. (2014). Proportionality, the Burden of Proof, and Some Signs of Reconsideration. *Irish Jurist (1935-)*, *52*, 141–152.

Khan, O., Ajadi, J. O., & Hossain, M. P. (2024). Predicting malaria outbreak in The Gambia using machine learning techniques. *PLOS ONE*, *19*(5), e0299386.

Kiehbadroudinezhad, M., Merabet, A., & Hosseinzadeh-Bandbafha, H. (2024). Chapter Twelve—Health impacts of greenhouse gases emissions on humans and the environment. In M. R. Rahimpour, M. A. Makarem, & M. Meshksar (Eds.), *Advances and Technology Development in Greenhouse Gases: Emission, Capture and Conversion* (pp. 265–291). Elsevier.

Klepac, P., Hsieh, J. L., Ducker, C. L., Assoum, M., Booth, M., Byrne, I., Dodson, S., Martin, D. L., Turner, C. M. R., van Daalen, K. R., Abela, B., Akamboe, J., Alves, F., Brooker, S. J., Ciceri-Reynolds, K., Cole, J., Desjardins, A., Drakeley, C., Ediriweera, D. S., … Fall, I. S. (2024). Climate change, malaria and neglected tropical diseases: A scoping review. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *118*(9), 561–579.

Krayenhoff, E. S., Broadbent, A. M., Zhao, L., Georgescu, M., Middel, A., Voogt, J. A., Martilli, A., Sailor, D. J., & Erell, E. (2021). Cooling hot cities: A systematic and critical review of the numerical modelling literature. *Environmental Research Letters*, *16*(5), 053007.

Kristanti, R. A., Hadibarata, T., Syafrudin, M., Yılmaz, M., & Abdullah, S. (2022). Microbiological Contaminants in Drinking Water: Current Status and Challenges. *Water, Air, & Soil Pollution*, *233*(8), 299.

Kulkarni, M. A., Duguay, C., & Ost, K. (2022). Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: A scoping review of reviews. *Globalization and Health*, *18*(1), 1.

Kumar, A., Pali, H. S., & Kumar, M. (2023). A comprehensive review on the production of alternative fuel through medical plastic waste. *Sustainable Energy Technologies and Assessments*, *55*, 102924.

Lauriola, P., Crabbe, H., Behbod, B., Yip, F., Medina, S., Semenza, J. C., Vardoulakis, S., Kass, D., Zeka, A., Khonelidze, I., Ashworth, M., de Hoogh, K., Shi, X., Staatsen, B., Knudsen, L. E., Fletcher, T., Houthuijs, D., & Leonardi, G. S. (2020). Advancing Global Health through Environmental and Public Health Tracking. *International Journal of Environmental Research and Public Health*, *17*(6), Article 6.

Leung, X. Y., Islam, R. M., Adhami, M., Ilic, D., McDonald, L., Palawaththa, S., Diug, B., Munshi, S. U., & Karim, M. N. (2023b). A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected Tropical Diseases*, *17*(2), e0010631.

Leung, X. Y., Islam, R. M., Adhami, M., Ilic, D., McDonald, L., Palawaththa, S., Diug, B., Munshi, S. U., & Karim, M. N. (2023a). A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected Tropical Diseases*, *17*(2), e0010631.

Maemu, E. (2023). *The impact of COVID-19 on the implementation of the National Development Plan 2030: A case of Vhembe District Municipality* [Thesis]. https://univendspace.univen.ac.za/handle/11602/2536

Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023). Ensemble Learning for Disease Prediction: A Review. *Healthcare*, *11*(12), Article 12.

Mahule, A., Roy, K., Sawarkar, A. D., & Lachure, S. (2024). Enhancing Environmental Resilience: Precision in Air Quality Monitoring through AI-Driven Real-Time Systems. In *Artificial Intelligence for Air Quality Monitoring and Prediction*. CRC Press.

Malakoane, B., Heunis, J. C., Chikobvu, P., Kigozi, N. G., & Kruger, W. H. (2020). Public health system challenges in the Free State, South Africa: A situation appraisal to inform health system strengthening. *BMC Health Services Research*, *20*(1), 58.

Martineau, P., Behera, S. K., Nonaka, M., Jayanthi, R., Ikeda, T., Minakawa, N., Kruger, P., & Mabunda, Q. E. (2022). Predicting malaria outbreaks from sea surface temperature variability up to 9 months ahead in Limpopo, South Africa, using machine learning. *Frontiers in Public Health*, *10*.

McAndrew, T., Wattanachit, N., Gibson, G. C., & Reich, N. G. (2021). Aggregating predictions from experts: A review of statistical methods, experiments, and applications. *WIREs Computational Statistics*, *13*(2), e1514.

McMahon, A. (2021). *Earth observation and mosquito-borne diseases: Assessing environmental risk factors for disease transmission via remote sensing data*. https://shareok.org/handle/11244/330179

Mohtady Ali, H., Ranse, J., Roiko, A., & Desha, C. (2022). Healthcare Workers' Resilience Toolkit for Disaster Management and Climate Change Adaptation. *International Journal of Environmental Research and Public Health*, *19*(19), Article 19.

Mulisa, F. (2022). When Does a Researcher Choose a Quantitative, Qualitative, or Mixed Research Approach? *Interchange*, *53*(1), 113–131.

Murad, N. Y., Hasan, M. H., Azam, M. H., Yousuf, N., & Yalli, J. S. (2024). Unraveling the Black Box: A Review of Explainable Deep Learning Healthcare Techniques. *IEEE Access*, *12*, 66556–66568.

Muriithi, D., Lumumba, V. W., & Okongo, M. (2024). A Machine Learning-Based Prediction of Malaria Occurrence in Kenya. *American Journal of Theoretical and Applied Statistics*, *9*(2), Article 2.

Narwal, S., & Jain, S. (2021). Building Resilient Health Systems: Patient Safety during COVID-19 and Lessons for the Future. *Journal of Health Management*, *23*(1), 166–181.

Nel, J., & Richards, L. (2022). Climate change and impact on infectious diseases. *Wits Journal of Clinical Medicine*, *4*(3), 129–134.

Neto, S. R. da S., Oliveira, T. T., Teixeira, I. V., Oliveira, S. B. A. de, Sampaio, V. S., Lynn, T., & Endo, P. T. (2022). Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review. *PLOS Neglected Tropical Diseases*, *16*(1), e0010061.

Ng, N., Aizuddin, A., & Ooi, W. (2025). Sustainability and green healthcare policies in Malaysia: A policy-driven approach. *Perspectives in Public Health*, 17579139251339087.

Nguyen, V.-H., Tuyet-Hanh, T. T., Mulhall, J., Minh, H. V., Duong, T. Q., Chien, N. V., Nhung, N. T. T., Lan, V. H., Minh, H. B., Cuong, D., Bich, N. N., Quyen, N. H., Linh, T. N. Q., Tho, N. T., Nghia, N. D., Anh, L. V. Q., Phan, D. T. M., Hung, N. Q. V., & Son, M. T. (2022). Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLOS Neglected Tropical Diseases*, *16*(6), e0010509.

Nichols, G., Lake, I., & Heaviside, C. (2018). Climate Change and Water-Related Infectious Diseases. *Atmosphere*, *9*(10), Article 10.

Noureen, A., Aziz, R., Ismail, A., & Trzcinski, A. P. (2022). The Impact of Climate Change on Waterborne Diseases in Pakistan. *Sustainability and Climate Change*, *15*(2), 138–152.

Nunbogu, A. M. (2022). *Bridging the health equity gap: Examining the effects of water, sanitation and hygiene (WaSH) gender-based violence on health and wellbeing in Ghana*.

Nunes, L. J. R. (2023). The Rising Threat of Atmospheric CO2: A Review on the Causes, Impacts, and Mitigation Strategies. *Environments*, *10*(4), Article 4.

Nusrat, F., Haque, M., Rollend, D., Christie, G., & Akanda, A. S. (2022). A High-Resolution Earth Observations and Machine Learning-Based Approach to Forecast Waterborne Disease Risk in Post-Disaster Settings. *Climate*, *10*(4), Article 4.

Odu, N. B. (2021). *Machine Learning Techniques for Malaria Incidence and Tuberculosis Prediction* [Thesis, AUST]. http://repository.aust.edu.ng/xmlui/handle/123456789/5096

Ombui, G. S. (2023). *Predictive Analytics for Retention in Care and Antiretroviral Therapy Adherence Using Supervised Learning: A Case Study of County Health Facilities in Kenya* [Thesis, University of Nairobi]. http://erepository.uonbi.ac.ke/handle/11295/163970

Opoku, S. K., Filho, W. L., Hubert, F., & Adejumo, O. (2021). Climate Change and Health Preparedness in Africa: Analysing Trends in Six African Countries. *International Journal of Environmental Research and Public Health*, *18*(9), Article 9.

Pandey, R. K., Dubey, A. K., Sharma, S., & Rani, C. (2023). Climate Change and Zoonotic Diseases: Malaria, Plague, Dengue, and Encephalitis. In *Emerging Pandemics*. CRC Press.

Patel, L., Conlon, K. C., Sorensen, C., McEachin, S., Nadeau, K., Kakkad, K., & Kizer, K. W. (2022). Climate Change and Extreme Heat Events: How Health Systems Should Prepare. *NEJM Catalyst*, *3*(7), CAT.21.0454.

Patel, R. (2025). HOW TO LEVERAGE DIGITAL MARKETING TO EXPAND THE REACH OF SMES? *International Conference "Actual Economy: Local Solutions for Global Challenges,"* 78–78.

Paul, V. K., Rastogi, A., Dua, S., & Basu, C. (2023). *Healthcare Infrastructure, Resilience and Climate Change: Preparing for Extreme Weather Events*. Routledge.

Peng, T., Chen, X., Wan, M., Jin, L., Wang, X., Du, X., Ge, H., & Yang, X. (2021). The Prediction of Hepatitis E through Ensemble Learning. *International Journal of Environmental Research and Public Health*, *18*(1), Article 1.

Pepin, N. C., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., Palazzi, E., Seibert, P., Serafin, S., Schöner, W., Terzago, S., Thornton, J. M., Vuille, M., & Adler, C. (2022). Climate Changes and Their Elevational Patterns in the Mountains of the World. *Reviews of Geophysics*, *60*(1), e2020RG000730.

Phelan, H., Yates, V., & Lillie, E. (2022). Challenges in healthcare delivery in low- and middle-income countries. *Anaesthesia & Intensive Care Medicine*, *23*(8), 501–504.

Proceedings of the 16th Annual Conference on the Science of Dissemination and Implementation in Health. (2024). *Implementation Science*, *19*(2), 42.

Rahimi-Ardabili, H., Magrabi, F., & Coiera, E. (2022). Digital health for climate change mitigation and response: A scoping review. *Journal of the American Medical Informatics Association*, *29*(12), 2140–2152.

Rahman, H., Shah, U. M., Riaz, S. M., Kifayat, K., Moqurrab, S. A., & Yoo, J. (2024). Digital twin framework for smart greenhouse management using next-gen mobile networks and machine learning. *Future Generation Computer Systems*, *156*, 285–300.

Rami, F., Thompson, L., & Solis-Cortes, L. (2023). Healthcare Disparities: Vulnerable and Marginalized Populations. In H. R. Searight (Ed.), *Covid-19: Health Disparities and Ethical Challenges Across the Globe* (pp. 111–145). Springer International Publishing.

Rana, P., & Varshney, L. R. (2021). Trustworthy Predictive Algorithms for Complex Forest System Decision-Making. *Frontiers in Forests and Global Change*, *3*.

Rane, N., Choudhary, S. P., & Rane, J. (2024). Ensemble deep learning and machine learning: Applications, opportunities, challenges, and future directions. *Studies in Medical and Health Sciences*, *1*(2), Article 2.

Rayan, R. A., Choudhury, M., Deb, M., Chakravorty, A., Devi, R. M., & Mehta, J. (2021a). Chapter 10 - Climate change: Impact on waterborne infectious diseases. In B. Thokchom, P. Qiu, P. Singh, & P. K. Iyer (Eds.), *Water Conservation in the Era of Global Climate Change* (pp. 213–228). Elsevier.

Rayan, R. A., Choudhury, M., Deb, M., Chakravorty, A., Devi, R. M., & Mehta, J. (2021b). Chapter 10 - Climate change: Impact on waterborne infectious diseases. In B. Thokchom, P. Qiu, P. Singh, & P. K. Iyer (Eds.), *Water Conservation in the Era of Global Climate Change* (pp. 213–228). Elsevier.

Rees, N. (2021). *The Climate Crisis Is a Child Rights Crisis: Introducing the Children's Climate Risk Index*. UNICEF. https://eric.ed.gov/?id=ED614506

Rocque, R. J., Beaudoin, C., Ndjaboue, R., Cameron, L., Poirier-Bergeron, L., Poulin-Rheault, R.-A., Fallon, C., Tricco, A. C., & Witteman, H. O. (2021). Health effects of climate change: An overview of systematic reviews. *BMJ Open*, *11*(6), e046333.

Rogers, H. L., Barros, P. P., Maeseneer, J. D., Lehtonen, L., Lionis, C., McKee, M., Siciliani, L., Stahl, D., Zaletel, J., & Kringos, D. (2021). Resilience Testing of Health Systems: How

Can It Be Done? *International Journal of Environmental Research and Public Health*, *18*(9), Article 9.

Rossi, B. (2021). Forecasting in the Presence of Instabilities: How We Know Whether Models Predict Well and How to Improve Them. *Journal of Economic Literature*, *59*(4), 1135–1190.

Sabasteanski, N. D. (2021). Climate migration and health system preparedness in the United States. *Climate Policy*, *21*(3), 368–382.

Sahlaoui, H., Alaoui, E. A. A., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. *IEEE Access*, *9*, 152688–152703.

Salas, R. N., Friend, T. H., Bernstein, A., & Jha, A. K. (2020). Adding A Climate Lens To Health Policy In The United States. *Health Affairs*, *39*(12), 2063–2070.

Santos, U. de P., Arbex, M. A., Braga, A. L. F., Mizutani, R. F., Cançado, J. E. D., Terra-Filho, M., & Chatkin, J. M. (2021). Environmental air pollution: Respiratory effects. *Jornal Brasileiro de Pneumologia*, *47*, e20200267.

Sapuan, S. M., Ilyas, R. A., & Asyraf, M. R. M. (2022). Carbon Footprint in Healthcare. In S. M. Sapuan, R. A. Ilyas, & M. R. M. Asyraf (Eds.), *Safety and Health in Composite Industry* (pp. 115–137). Springer.

Schwab, R., Schiestl, L. J., & Hasenburg, A. (2025). Greening the future of healthcare: Implementation of sustainability strategies in German hospitals and beyond—a review. *Frontiers in Public Health*, *13*.

Semenza, J. C., & Paz, S. (2021). Climate change and infectious disease in Europe: Impact, projection and adaptation. *The Lancet Regional Health – Europe*, *9*.

Semenza, J. C., Rocklöv, J., & Ebi, K. L. (2022). Climate Change and Cascading Risks from Infectious Disease. *Infectious Diseases and Therapy*, *11*(4), 1371–1390.

Sesay, U., & Osborne, A. (2025). Building climate-resilient health systems in Sierra Leone: Addressing the dual burden of infectious and climate-related diseases. *Infectious Diseases of Poverty*, *14*(1), 23.

Siamba, S. N. (2022). *Forecasting tuberculosis infections using arima and hybrid neural network models among children below 15 years in homa bay and turkana counties, kenya* [Thesis, University of Eldoret].

Smith, G. S., Anjum, E., Francis, C., Deanes, L., & Acey, C. (2022). Climate Change, Environmental Disasters, and Health Inequities: The Underlying Role of Structural Inequalities. *Current Environmental Health Reports*, *9*(1), 80–89.

Soares, A. L., Buttigieg, S. C., Bak, B., McFadden, S., Hughes, C., McClure, P., Couto, J. G., & Bravo, I. (2023). A Review of the Applicability of Current Green Practices in Healthcare Facilities. *International Journal of Health Policy and Management*, *12*, 6947.

Sokhi, R. S., Moussiopoulos, N., Baklanov, A., Bartzis, J., Coll, I., Finardi, S., Friedrich, R., Geels, C., Grönholm, T., Halenka, T., Ketzel, M., Maragkidou, A., Matthias, V., Moldanova, J., Ntziachristos, L., Schäfer, K., Suppan, P., Tsegas, G., Carmichael, G., … Kukkonen, J. (2022). Advances in air quality research – current and emerging challenges. *Atmospheric Chemistry and Physics*, *22*(7), 4615–4703.

Soomro, S., Sahito, J. G. M., & Gilal, F. (2025). The Link Between Climate Change and Human Health. In J. Bento & M. M. Torres (Eds.), *Global Perspectives on Climate Change, Inequality, and Multinational Corporations* (pp. 183–208). Springer Nature Switzerland.

Sorensen, C. J., Salas, R. N., Rublee, C., Hill, K., Bartlett, E. S., Charlton, P., Dyamond, C., Fockele, C., Harper, R., Barot, S., Calvello-Hynes, E., Hess, J., & Lemery, J. (2020). Clinical Implications of Climate Change on US Emergency Medicine: Challenges and Opportunities. *Annals of Emergency Medicine*, *76*(2), 168–178.

Tewari, N., Bush, A., Butt, M. N., Stevens, E., & Zafar, S. (2023). *Gendered Dimensions of Loss and Damage in Asia*.

Tozan, Y., Branch, O. L. H., & Rocklöv, J. (2021). Vector-Borne Diseases in a Changing Climate and World. In K. E. Pinkerton & W. N. Rom (Eds.), *Climate Change and Global Public Health* (pp. 253–271). Springer International Publishing.

Tran, H. M., Tsai, F.-J., Lee, Y.-L., Chang, J.-H., Chang, L.-T., Chang, T.-Y., Chung, K. F., Kuo, H.-P., Lee, K.-Y., Chuang, K.-J., & Chuang, H.-C. (2023a). The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. *Science of The Total Environment*, *898*, 166340.

Tran, H. M., Tsai, F.-J., Lee, Y.-L., Chang, J.-H., Chang, L.-T., Chang, T.-Y., Chung, K. F., Kuo, H.-P., Lee, K.-Y., Chuang, K.-J., & Chuang, H.-C. (2023b). The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. *Science of The Total Environment*, *898*, 166340.

Usman, S., Jayeoba, J. O., & Kundiri, A. M. (2024). Climate Change at a Global Concept: Impacts and Adaptation Measures. *International Journal of Environment and Climate Change*, *14*(6), Article 6.

Vallée, A. (2024). Green hospitals face to climate change: Between sobriety and resilience. *Heliyon*, *10*(2).

van der Schaar, M., Alaa, A. M., Floto, A., Gimson, A., Scholtes, S., Wood, A., McKinney, E., Jarrett, D., Lio, P., & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning*, *110*(1), 1–14.

Walsh, J. E., Ballinger, T. J., Euskirchen, E. S., Hanna, E., Mård, J., Overland, J. E., Tangen, H., & Vihma, T. (2020). Extreme weather and climate events in northern areas: A review. *Earth-Science Reviews*, *209*, 103324.

Walter, T. G., Bricknell, L. K., Preston, R. G., & Crawford, E. G. C. (2024). Climate Change Adaptation Methods for Public Health Prevention in Australia: An Integrative Review. *Current Environmental Health Reports*, *11*(1), 71–87.

Wang, W., Li, S., Guo, S., Ma, M., Feng, S., & Bao, L. (2021). Benchmarking urban local weather with long-term monitoring compared with weather datasets from climate station and EnergyPlus weather (EPW) data. *Energy Reports*, *7*, 6501–6514.

Wang, Y., Chen, P., Wang, Y., Wang, Y., & Li, Q. (2024). Research on Competitive Strategy of Power Grid Emerging Business Based on Five Forces Model. *2024 9th Asia Conference on Power and Electrical Engineering (ACPEE)*, 2691–2695.

Waweru, H. K. (2024). *Prevalence of Plasmodium falciparum Antimalarial Drugs Resistance Genetic Markers in Selected Lake Victoria Islands, Western Kenya.* [Thesis, JKUAT-COHES].

Yadav, N., & Upadhyay, R. K. (2023). Global Effect of Climate Change on Seasonal Cycles, Vector Population and Rising Challenges of Communicable Diseases: A Review. *Journal of Atmospheric Science Research*, *6*(1), Article 1.

Yang, H., Yao, R., Dong, L., Sun, P., Zhang, Q., Wei, Y., Sun, S., & Aghakouchak, A. (2024). Advancing flood susceptibility modeling using stacking ensemble machine learning: A multi-model approach. *Journal of Geographical Sciences*, *34*(8), 1513–1536.

Zheng, L., Gao, Q., Yu, S., Chen, Y., Shi, Y., Sun, M., Liu, Y., Wang, Z., & Li, X. (2024). Using empirical dynamic modeling to identify the impact of meteorological factors on hemorrhagic fever with renal syndrome in Weifang, Northeastern China, from 2011 to 2020. *PLOS Neglected Tropical Diseases*, *18*(6), e0012151.

Appendix I: Nacosti Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **769446**

Date of Issue: **10/June/2025**

RESEARCH LICENSE

This is to Certify that Mr.. Jared N/A Murundu of The Co-operative University of Kenya, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Homabay, Kakamega, Kisumu, Migori on the topic: OPTIMIZING HEALTHCARE RESILIENCE THROUGH CLIMATE-INFORMED DATA ANALYTICS. for the period ending : 10/June/2026.

License No: **NACOSTI/P/25/4174624**

769446

Applicant Identification Number

Deputy Director
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix II: Application for NACOSTI Letter

THE CO-OPERATIVE UNIVERSITY OF KENYA

P. O. Box 24814-00502 Karen, Kenya

Telephone: (020)-2430127/2679456/8891401 Fax (020)-8891410

www.cuk.ac.ke

BOARD OF POSTGRADUATE STUDIES

5th April 2025

The Director,
National Commission for Science, Technology& Innovation,
Utalii House, Nairobi.

Dear Sir/Madam,

RE: JARED MURUNDU, REGISTRATION NUMBER: MDATC01/6067/2022

This is to introduce the above named Master of Science in Data Science student in the School of Computing and Mathematics at The Co-operative University of Kenya.

He has successfully completed his course work and is proceeding to the field to collect data on healthcare resilience. The title of his research project is ***"Optimizing Healthcare Resilience through Climate-Informed Data Analytics"***

Kindly accord him the necessary assistance.

Yours faithfully,

D. K. Muthoni
Director, Board of Postgraduate Studies.

Copy to: Dean SCM

Appendix II: Publication Pap

Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403013
Volume 4, Issue3 (September-December 2025), PP: 67-71.
www.indjcst.com
ISSN No: 2583-5300

Utilization of Climate Data Analytics in Healthcare Decision-Making Processes in Migori County, Kenya

Jared Murundu[1] Fidelis Mukudi[2]

[1] *Department of Computer Science and Information Technology, Co-operative University of Kenya, Kenya.*
[2]*Department of Mathematical Sciences, Co-operative University of Kenya, Nairobi, Kenya.*

To Cite this Article: *Jared Murundu[1] Fidelis Mukudi[2], "Utilization of Climate Data Analytics in Healthcare Decision-Making Processes in Migori County, Kenya", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 67-71.*

Abstract: *Climate variability is a primary driver of malaria surges in Kenya's Lake Victoria Basin, where rainfall pulses and humid, warm conditions amplify transmission; Migori County is especially exposed due to recurrent MAM/OND rains and flood-prone zones. The problem is that routine health decisions in Migori often remain reactive, mobilizing after cases rise rather than before. This study therefore analyzed how climate data analytics (CDA) are currently integrated into operational workflows for malaria preparedness in Migori County. Using secondary weekly facility records (2015–2024) and CDA outputs, we fitted a Negative Binomial regression and mapped significant signals to Standard Operating Procedures (SOPs). Results from the regression model show that rainfall at lag-1 week significantly increases malaria incidence ($\beta = 0.018$, $IRR = 1.018$, $p < 0.01$), while temperatures >24 °C display a modest negative association ($\beta = -0.058$, $p = 0.012$); seasonal harmonic terms align with MAM/OND peaks. Descriptively, weekly rainfall averaged 54.6 mm with temperature 22.3 °C, median malaria cases were 12 per facility-week with peaks up to 147 in Nyatike. Operationally, CDA is embedded through Green/Amber/Red risk bands that trigger targeted stock checks, pre-positioning, surge rosters, and outreach, with higher-resilience facilities experiencing muted spikes relative to comparable exposure. Despite these gains, barriers persisted data latency, fragmented stock visibility across facilities/central stores, and uneven analytic capacity which slow signal-to-action translation. Overall, evidence from the 2015–2024 facility-week dataset and the fitted regression confirms that CDA has shifted malaria management in Migori from reactive to anticipatory planning by providing short-lead risk signals that inform routine SOPs*

Key Words: *Climate Data Analytics; Healthcare Resilience; Malaria; Decision-Making; Migori County.*

I.INTRODUCTION

Healthcare Climate change has emerged as one of the most significant global health threats of the 21st century, exerting a strong influence on disease dynamics and healthcare delivery (Patz et al., 2014). In Sub-Saharan Africa, the burden of malaria and other climate-sensitive diseases has intensified due to rising temperatures, erratic rainfall, and increased humidity, which provide favourable conditions for vector breeding and disease transmission (Ayanlade et al., 2022).

The link between climate variability and the spread of malaria infections is well documented in scientific publications. For instance, rainfall pulses create mosquito breeding sites, while higher humidity and warm temperatures accelerate parasite development (Chandra & Mukherjee, 2022). Numerous studies confirm that climatic conditions shape malaria transmission patterns in Africa and globally (Mafwele & Lee, 2022). This evidence highlights the importance of integrating environmental data into healthcare planning.

Traditionally, healthcare decision-making in Kenya has relied heavily on routine surveillance and basic statistical analysis (Karijo et al., 2021), often focusing on case counts after outbreaks occur (Njoka, 2015). While these methods offer some insights, they are reactive and fail to capture the complexity of climate health interactions. As a result, outbreaks frequently overwhelm facilities before interventions are mobilized (Neta et al., 2022).

Despite the proven association between climate variability and malaria, the use of climate data analytics (CDA) in routine healthcare decision-making in Kenya stays limited. Barriers include fragmented data systems, lack of technical capacity, and minimal integration of climate data into existing health information systems (Bogaert et al., 2021).This study therefore aimed at analyzing the current utilization of climate data analytics in healthcare decision-making in Migori County, to understand how CDA is embedded into weekly workflows and how it informs preparedness and resource allocation.

Problem Statement

Climate variability has become a major driver of health risks in Sub-Saharan Africa, with malaria being one of the most climate-sensitive diseases in Kenya's Lake Victoria Basin (Olela et al., 2024). Frequent rainfall fluctuations and rising temperatures have increased the unpredictability of malaria transmission patterns, creating major challenges for healthcare systems in Migori County. These climate-induced surges have a severe impact on local populations by straining health facilities, disrupting drug supply chains, and overwhelming personnel during peak seasons (Davis, 2025). The consequences include delayed