PREDICTIVE MODEL TO FORECAST DESERT LOCUST OUTBREAKS IN KENYA USING MAXIMUM ENTROPY

BY

NOAH KIPYEGON CHEPKWONY

MASTER OF SCIENCE IN INFORMATION SYSTEMS MANAGEMENT

KCA UNIVERSITY

2025

PREDICTIVE MODEL TO FORECAST DESERT LOCUST OUTBREAKS IN KENYA USING MAXIMUM ENTROPY

BY

NOAH KIPYEGON CHEPKWONY

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN INFORMATION
SYSTEMS MANAGEMENT IN THE SCHOOL OF TECHNOLOGY AT KCA
UNIVERSITY

OCTOBER, 2025

DECLARATION

I declare that this research dissertation is my original work and has not been previously published or submitted elsewhere for award of a master's degree. I also declare that this contains no material written or published by other people except where due reference is made, and author duly acknowledged.

NAME: Noah KIPYEGON Reg No:**22/08079**

Sign: Date: 16TH October 2025

This dissertation has been submitted for examination with my approval as the appointed university supervisor.

Digitally signed by Prof.
Dr. Lucy W. Mburu
Date: 2025.10.26
14:01:16 +03'00'

Dr. Lucy W. Waruguru Date: 26/10/2025

ABSTRACT

Desert locusts (*Schistocerca gregaria*) are one of the most destructive transboundary pests, posing significant threats to food security, livelihoods, and vegetation. In Kenya, a severe outbreak of desert locust outbreak occurred between December 2019 and June 2021, causing extensive damage to crop and vegetation, specially in the eastern and northeastern parts of the country. Using forecasted climate and environmental data as well as historical occurrence data, it is possible to Predict possibility of an outbreak which can facilitate relevant stakeholders to put in place necessary measures to mitigate the effects. This prediction can help enhance early warning systems by facilitating timely intervention towards mitigating risks efforts. This research study aimed at coming up with a predictive model for desert locust outbreaks in Kenya using the MaxEnt algorithm and historical presence data together with environmental variables such as precipitation, soil moisture, temperature, and vegetation indices to identify areas susceptible to infestations. The research used used Maxent algorithm and latest technologies of GIS and machine learning techniques to generate maps that classify areas in terms of risks levels (low, medium, high) based on climate data and historical locust occurrence data. The output will help enhance locust monitoring and forecasting, providing critical insights for policymakers, stakeholders, and farmers. The output includes a validated prediction model, maps, and recommendations for locust control strategies. The findings revealed that precipitation and soil moisture were the strongest predictors of habitat suitability, followed by temperature and vegetation indices. The MaxEnt model produced validated habitat suitability maps, classifying areas into low, medium, and high-risk zones. High-risk areas were concentrated in northeastern and eastern Kenya, aligning with regions historically affected by locust invasions. These results demonstrate that combining presence-only data with climatic and environmental predictors provides reliable forecasts of potential outbreak zones. The study concludes that the predictive model and generated risk maps can strengthen early warning systems, guide surveillance and control operations, and support policymakers, stakeholders, and farmers in mitigating the impact of desert locust outbreaks.

ACKNOWLEDGMENT

First and foremost, I would like to thank God Almighty for giving me the strength, good health, will and for sustaining me throughout my course of study.

I especially would like to thank my family and specifically my wife for her belief in me and her tireless encouragement, our children, Nayla and Nolan, whom I sacrificed my usual attention and parental care during my study. This work is dedicated to them for the sacrifice, care, love, concern, encouragement, support, and enthusiasm.

I want to thank my supervisor Dr. Lucy Mburu, who guided, encouraged, and led me though the journey to completing this dissertation.

I want to express my gratitude to everyone who, in some way or another, helped me to attempt and complete this project.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ACRONYMS AND ABBREVIATIONS

DLIS - Desert Locust Information Service

ENT - Ecological Niche Theory

GIS – Geographic Information System

ISSM - Information Systems Success Model

MaxEnt - Maximum Entropy

SDM - Species Distribution Modeling

TAM - Technology Acceptance Model

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Desert locusts (Schistocerca gregaria) are some of the most destructive migratory pests that causes extensive damage to crop and vegetation. Their ability to move across boundaries over long distances as well as their ecology and biology makes them among the world's most devastating pests. Outbreaks of these locusts pose a severe threat to food security, livelihoods, and ecosystems with lasting socio-economic impacts (Maeno et al., 2021). Desert locust infestation has in the past caused great agricultural losses resulting in famine, displacement, and economic losses in the affected regions.

Historically, major plagues have occurred periodically, and small plagues have been recorded between 1926 and 1934, 1940 and 1948 and 1986 and 1989 which caused food insecurity and economic losses to a large extent (FAO, 2020). The last outbreak of 2019-2021 was one of the worst in recent decades, affecting more than 23 countries and having a severe negative impact on agriculture in the Horn of Africa (World Meteorological Organization, 2021).

Countries in Eastern Africa region has in the past been one of the most vulnerable areas to desert locust invasions due to their favorable weather conditions, rainfall patterns and being located adjacent to regions considered to have frequent outbreaks of desert locust. The 2019-2021 locust invasion was particularly devastating to countries like Ethiopia, Somalia and Kenya which it affected severely, and which only helped to worsen food insecurity and economic instability. In Kenya, more than 70,000 hectares of cropland and pasture were ravaged, and the government and international organizations launched emergency control operations (DLCO-EA, 2021). The FAO together with the relevant authorities dispatched surveillance teams, educated farmers on locust surveillance, and conducted large-scale pesticide application. However, the response efforts were mainly backward-looking, and the focus was on damage control rather than prevention and prediction (FAO, 2021).

In Kenya, the most recent and devastating outbreak was reported between year 2019 and 2022 and it affected mainly the eastern and northeastern regions of the country. This outbreak caused significant destruction where destruction of crops and animal pasture was reported, leading to food insecurity and exposing the country to economic vulnerabilities. Kenya geographic location of closeness to common breeding sites in Ethiopia and Somalia makes it more susceptible to desert locust invasion.

However, the severity of this outbreak highlighted critical gaps in Kenya's ability to respond proactively. Specifically, the country has not put in place robust real-time monitoring infrastructure for monitoring desert locust infestation, with tracking often done using manual field observations which leads to delay in notification and hence reactive control measures, which struggle to mitigate the scale of damage caused by rapidly spreading locust swarms.

While there exists a monitoring system supported by FAO called eLocust3 tool, which provide reliable data collection for monitoring infestation, it lacks predictive modeling capabilities that would make use of climatic and environmental data to provide forecast information on the anticipated outbreaks (FAO, 2021).

These events highlight the importance of a sustained monitoring approach to deal with the negative effects of a desert locust outbreak (World Meteorological Organization, 2021). These recurring outbreaks emphasized the necessity of better surveillance and more efficient rapid-response actions to control locusts. Managing the spread posed huge difficulties to control agencies and towards the end of the outbreak, around $400 million was estimated to have been spent on control operations, whereas agricultural damages were estimated to be close to $2.5 billion.

Today we are seeing exciting new technologies to manage desert locust infestations using data as the technology evolves more and more. As a result of global scholars working together, a wealth of insights on how climate and environmental conditions influence behavior has allowed this progress to flourish. Based on machine learning and sophisticated climatic data analysis, researchers have created predictive models that may be used to track and even anticipate locust outbreaks. My study aimed to create a predictive model specifically for desert locust occurrences in Kenya. I used the Maximum Entropy (MaxEnt) algorithm to integrate historical locust presence data with key climate indicators such as temperature, precipitation, soil moisture, and vegetation indices. The goal was to pinpoint and map high-risk areas, which will enhance our ability to respond effectively to potential infestations. This approach is a collaborative, technology-based model that represents a significant step forward in our efforts to manage these agricultural threats.

The primary objective of this study is to develop a predictive model to predict and characterize local desert locust outbreaks in Kenya. We seek to integrate historical locust presence information with key environmental and climatic features. Data will be vital for establishing early warning systems and for developing timely measures that can combat any possible outbreaks.

The model studies important factors, such as rainfall, soil moisture, temperature, and vegetation health, to understand how desert locust behavior is shaped: reproductive and movement patterns. These factors, besides the fundamentals of ecosystem ecology of desert locusts, also measure on satellite data and climate records, therefore makes the approach data-driven and applicable to further the research in a bigger way. The objective of this predictive model is to transform complex data into actionable insights, generating risk maps showing areas at risk for locust swarms—low, medium, or high risk for locust activity.

This will allow policymakers, farmers, and other community members to have a tool to predict locust outbreaks and take preventative action. This aligns with better food security and resilience and goes beyond responding to outbreaks to developing preventive approaches that minimize the damage. By doing so, the study corresponds to the wider climate adaptation and disaster risk reduction aspirations in Kenya and the wider Horn of Africa.

1.2 Statement of the Problem

The agricultural productivity and food security in Kenya is still threatened by Desert Locust Outbreaks. These infestations are particularly damaging since the locusts can destroy and consume a tremendous number of crops and pastures within a very short amount of time. The outbreak of the year 2019-2021 was particularly devastating as it resulted in the worst invasion in Kenya in several years. To this date, Desert Locusts have infested roughly seventy thousand hectares of land. This infestation posed a severe food security threat especially in the Northern part of Kenya where billions of people were in danger (Kazeem, 2020). By mid-May of 2020, the afflicted regions suffered losses of crops and pastures in aggregate of 5-15 percent in northern Kenya and 1-5 percent in southeastern regions (ICPAC, 2022). These exacerbated vulnerabilities further worsened the economic hardship and food insecurity for the afflicted communities, leading to these communities facing dire consequences.

In response to previous outbreaks, the government of Kenya with the assistance of international partners have undertaken massive aerial and land-based pest control measures. These activities were contributory to progressive diminutions in the population of locusts from June through September 2020, noting decreased swarm numbers and a more concentrated distribution. As of October 2020, only minor swarms were present in the northwestern regions of Baringo, Laikipia and Samburu counties (Germany Trade & Invest, 2021).

Despite these and other control measures, they were not enough to prevent the impact of the infestation on livelihoods from being severe as they were largely passive, focusing on containment rather than early detection and prevention. Manual and chemical control methods are neither sufficient in preventing locusts from breeding due to their moderation plan that is too slow to catch up with the locusts, allowing swarms to cause significant damage before intervention (FAO, 2021).

Use of conventional locust monitoring methods result in drastic increase in the cost incurred and decreased effectiveness of the control techniques, as they must be implemented only after the outbreak has already occurred (Salih et al., 2020). Too much use of pesticides can endanger specious of the non-targeted kind while also ruining the quality of soil and water and soil (Lecoq, 2020). Because of the rapid reproduction and migration patterns of the desert locusts, as well as their uncontrolled breeding, it is difficult to control the damage done by them. The difficulty in controlling locusts is enhanced by climate change as well (van Huis, 2019).

There is no recent research focusing on the response to desert locust outbreaks in Kenya on that matter with the latest research of Piou et al. in 2021 where they demonstrated the use of remote sensing and climate data (e.g., temperature, precipitation, and soil moisture) to map locust breeding sites. This study was built from static or retrospective models, and no predictions are possible in real time based on real time modeling.

For better early warning and response to locust epidemics, it will be useful to have a predictive model that includes historical locust occurrence and modern environmental and climate information (Piou et al., 2021). Although the FAO-developed eLocust3 tool does a good job of collecting dependable information, it would never be able to predict or provide us with early warning alerts, either (FAO, 2021; Piou et al., 2021). Our monitoring infrastructure also do not sufficiently integrate such climate variables that are critical to desert locust breeding and migration (Piou et al., 2021).

This highlights a specific research gap—the lack of an integrated framework that combines presence-only locust occurrence data with climatic and environmental variables to enable real-time predictive modeling. While past systems have emphasized monitoring and data collection, they fall short in linking these observations with dynamic environmental drivers to forecast potential outbreaks. Bridging this gap is essential to shift from reactive responses to proactive forecasting in Kenya’s locust management systems.

Although it has been tried many times over to use machine learning models predicting the infestation of desert locust, most of them fail in creating accurate results. For instance, Ren et al. (2020) studied past locust invasions in China that were based on ANN with climate and environment factors. Although their model could accurately model infestation patterns, it faced issues with enough labeled training samples. This disadvantage made predicting desert locust occurrences with accuracy impossible.

Using Support Vector Machines (SVMs) to classify locust habitats and find areas of moisture, we concentrated on the relatively dry and semi-dry regions. Zhang et al. (2019) developed a locust forecast model in North Africa using an SVM process with the remote sensed data to choose suitable breeding as well as feeding point. While SVMs were effective on clean, low dimensional data, they were inappropriate with the complex character of climate interactions. Due to the intricate and non-linear relationships among many environmental factors and locust distribution, the SVMs' steep decision boundaries could lead to lack of generalization and large prediction errors in unknown locations.

Random Forest (RF) models have also been tried out for locust predictions, especially in Pakistan and India during the 2019-2021 locust invasion (Kumar, et al. 2021). The RF model combined data from different vegetation indices, temperature and wind speed to attempt to classify possible locust breeding grounds. While RF models are known to withstand overfitting and are efficient in the presence of missing data, they also do not have enough sensitivity to time and do not incorporate sequential changes in the movement of locusts over time. These models completely ignore the dynamics of locust movement and therefore, become incapable of making accurate predictions in real time.

The highlighted limitations indicate the need for a combination of current situations and past information in a single unit. An integrated, data-driven predictive model that would ascertain locust movements and breeding patterns to enable locust control before the outbreaks reach severe levels is sorely needed.

In this research study, we build a prediction model using the MaxEnt algorithm for predicting desert locust infestations. One major advantage of the MaxEnt algorithm is precisely predicting the range of species from the location of their observation, making it ideal for determining the probable risk zones for desert locust outbreaks based on several ecological and climatic points (Elith et al., 2011). The Maximum Entropy (MaxEnt) model gives great performance in scenarios where absence data is less or none—something that very few other machine learning models handle. This is particularly important if we are to predict the behaviors of desert locusts in Kenya, where there are more presence-only datasets available.

The study extends its coverage past occurrence records to future patterns of precipitation, temperature, and soil moisture; it uses predicted climatic data. Utilizing these components, the research can make models that illustrate how habitats that are most suitable for the species might change in the future. Not only does this approach help us gain insights into why infestations occurred before, but it also reveals how future outbreaks may unfold as climate changes. It’s a kind of step that can help advance the model’s predictive capabilities and can yield a crucial benefit in early warning systems.

MaxEnt is being used in this context to implement species distribution modeling techniques to improve forecasting systems in the field of desert locust surveillance. The model uses historical and projected environmental data which enhances the operation of early warning systems for desert locusts. This helps officials more proactively respond to and prevent locust outbreaks from being catastrophically intense. This approach prioritizes not only food security, but also livelihoods and ecosystems.

1.3 Main Objective

The main aim of this study was to build and validate a predictive model for the Maximum Entropy (MaxEnt) method to predict the expected distribution and prevalence of desert locusts (Schistocerca gregaria) in the region of Kenya. The model combines the historical presence-only dataset of locust swarms with the most important environmental and climate features to ascertain ecological suitability of distinct parts of the country.

This study is novel in controlling desert locusts in Kenya by focusing on predictive modelling. Rather than only responding to outbreaks as they're occurring, the aim is to predict breeding and migration hotspots. This allows us to allow parties to anticipate threats to be faced, to plan interventions in a timely fashion and to allocate scarce resources to better the extent available.

An important part of this effort is ensuring the predictive model performs well not just in theory but in action. We assessed this model through independent test data for 2023 to 2024 to assess its performance to predict locust behavior based on the evolution of climate data. In terms of validation, this process can be a proof to be a reliable early warning tool that can be advantageous in the control efforts. Finally, the goal fits within the larger food security and climate resilience agenda for Kenya.

By adopting machine learning methods, climate science, and geospatial analysis, the model is set to improve locust monitoring systems as well as inform national and regional preparedness planning. The aim is to address the devastating socio-economic outcomes associated with locust invasions, ensure people survive in arid and semi-arid areas and achieve sustainable ways of farming.

1.4 Specific Objectives

i. Identify key environmental and climate factors that affect reproduction, mobility and the spread of the desert locust in Kenya.
ii. Develop a predictive model that integrates presence-only desert locust occurrence data with environmental and climatic variables to forecast the likelihood and spatial distribution of desert locust infestations in Kenya.
iii. Validate the model's predictive performance by comparing its outputs with historical desert locust outbreak data and, ground truth observations to assess its accuracy and reliability.
iv. Provide an early warning tool that supports timely intervention strategies to mitigate locust infestations in high-risk areas.

1.5 Research Questions

This study attempts to address the following questions:

i. What environmental and climatic factors influence desert locust breeding, migration and infestation in Kenya?
ii. How can historical and real-time data be integrated to develop a predictive model for forecasting desert locust outbreaks?
iii. How accurate and effective is the developed predictive model in categorizing locust risk areas?
iv. How can the predictive model support early warning systems and inform intervention strategies?

1.6 Significance of the Study

This study is aimed at developing and validating a predictive model using the MaxEnt algorithm to forecast desert locust outbreaks. We're drawing on historical presence data of desert locusts and looking closely at environmental and climate factors, such as precipitation, soil moisture, and surface temperature. By doing this, we aim to create reliable habitat suitability maps that highlight areas with a high likelihood of locust outbreaks. Our findings show that climate factors have a significant impact on desert locust infestations. These include precipitation levels, soil moisture, and surface temperature, among other environmental influences. The tools and maps generated from this study will be valuable for policymakers and stakeholders, helping them make informed decisions about how to manage and control desert locust populations. By categorizing areas into low, medium, and high-risk zones, we provide a clear framework that can guide surveillance and intervention strategies. Ultimately, this study showcases how technology can enhance predictive analytics, improving early warning systems for desert locust outbreaks.

The study, by correlating environmental and climate datasets with historical presence-only data, can add to the application of data-driven technologies in pests management. By using climate and environmental data in MaxEnt algorithm to generate maps of desert locust breeding sites it describes how technology could be used to address agricultural issues. This new development supports advanced predictions of other problems in the environment or agriculture and extends machine learning in information technology to new heights.

The predictive modelling system integrates real-time and forecasted climate data processing with GIS tools and predictive modelling, helping to bring IT solution into a more active role in decision making and resource management through both forecasting and decision making. Research results from this study contribute to the advancement of developing technologies within agricultural landscapes for proactive pest management through the development of platforms. Its systems are based on real-time data analytics and geographic information systems (GIS), as technology has developed to build flexible frameworks. The scope of these innovations can expand into disaster preparedness and climate monitoring, just to name a few applications of information technology.

This study emphasizes the need for integrating environmental and climate data together with historical presence information. This can serve as the basis for strategic recommendations aimed at countering the dangers of outbreaks of desert locust. This study provides evidence that helps us develop sound tools for monitoring these pests and regional cooperation. Essentially, this method will encourage building sustainable food systems in light of climate-specific pest epidemics.

1.7 Motivation of the Study

Desert locusts endanger people's food, income, and the environment, particularly in dry and semi-dry areas. The 2019–2021 locust invasion in Kenya epitomized the devastating consequences of these pests, including the obliteration of crops, food insecurity, and the economic decline that ensued.

Typically, response strategies to monitor and control locusts have the upshot that widespread damage is left uncontrolled as reconnaissance for and response to imminent threats comes too late. Traditionally, locust monitoring and control operations have mostly been reactive in nature, using ground surveillance and spraying once the infestation is already underway. Although these actions can deliver short-term relief, they are usually limited by delayed reporting, limited resources, and logistical challenges in reaching affected areas quickly. As a result, by the time interventions are implemented, significant damage has already occurred. This reactive approach underscores the critical gap in proactive, predictive systems that can anticipate locust outbreaks before they escalate into full-scale invasions. Without reliable forecasting tools, the opportunity to contain swarms early is frequently missed, leaving millions of people exposed to the devastating consequences of unchecked locust infestations.

With the advancement of technology and the availability of remote sensing and climate data provide the chance to transition from a responsive approach to proactive management of locusts. Machine learning techniques, such as MaxEnt, with predictive modeling enables the enhancement of climate and environmental variables within the warning systems for a more efficient response. This study focuses on model development that assesses potential breeding and infestation areas so appropriate action can be taken by decision makers, farmers and other stakeholders before outbreaks worsen.

MaxEnt is particularly applicable to presence-only data, which is usually the most reliable locust occurrence information available. In this way, elements of environmental and climatic characteristics can be integrated into early warning systems so that potential breeding and infestation zones are better predicted. The goal of this study is to develop a reliable model to analyze changes in desert locust habitat over time and in different areas in Kenya. Thus, the goal is to provide farmers, decision makers, stakeholders and others with actionable information that can spur early action, focused monitoring and timely control measures to mitigate the potential for small locust outbreaks to become larger invasions.

By leveraging predictive analytics to current locust surveillance programs, we aim at bolstering food security, protecting livelihoods and minimizing long-term economic and environmental damage due to locust swarms. Also, by leveraging past presence-only and key environmental data, we obtain information on the behavior of desert locusts. Such knowledge not only augments their surveillance, but also allows the effective optimization of efforts and resources for control activities. This study aims to improve our readiness to handle future locust incursions to provide for food security with informed decision making.

1.8 Scope of the Study

1.8.1 Content Scope

This study involves modeling and validation of a predictive model for desert locusts based on the Maximum Entropy (MaxEnt) algorithm. We seek to integrate historical records of desert locust locations with significant climatic and environmental trends. Hopefully in doing that you can get a better sense of what areas these locusts could thrive and where their habitats evolve. Selected predictors included (specifically precipitation, soil moisture, surface temperature, vegetation indices (e.g., NDVI)) have been ecologically justified since they represent the most important influencing factors for the reproduction, development and swarm movement of locusts. Rainfall and vegetation index denote availability of food and breeding environments, soil moisture directs egg survival and hatching, and surface temperature regulates the physiological processes of incubation, development, and flight mobility. Overall, these considerations offer a comprehensive ecological baseline for mapping locust distribution and spread.

The study is divided into several intertwined phases. First, it includes its data collection and preprocessing stages, where the historical outbreak years of locust presence are aggregated from monitoring databases, FAO reports, and national surveillance sources and environmental data layers are extracted from trusted climate and remote sensing repositories including ERA5 and MODIS. Preprocessing activities consist of correcting coordinate inconsistencies, removing duplicates from data, aligning spatial scales between each other, and ensuring formats make sense for analysis.

Second, in the study we start with generating a model in MaxEnt (which is pretty good for understanding where species might live when we don't have absence data). This algorithm helps us forecast habitat suitability by deducing a probability distribution of situations that resemble areas where we've seen locusts. This results in outputs that would be continuous and would indicate areas of the world like locust regions in the history of its discovery.

Third, we will try to validate our predictions on separate datasets, which are not in our first training. We'll validate how robustly our model is using multiple metrics like AUC, omission rates, precision, recall and F1-scores. We will also concentrate on the question of the reliability of the model to be applied over different climate years and environmental conditions.

Finally, we will generate risk maps that outline where desert locusts may flourish. The maps will provide critical information for decision makers, guiding early warning activities and tailored measures regarding the suppression of potential epidemic outbreaks. The maps will help identify high-risk regions, even in dry and semi-arid counties susceptible to infestation, as well as allow for the effective use of limited resources for surveillance, monitoring, and control campaigns. Furthermore, those outputs could be used within regional early warning models to mitigate the transboundary nature of desert locust invasions within the Horn of Africa.

In conclusion, the research creates in addition to generating a model, it validates and operationalizes it to be a viable tool for risk mapping and management for fostering more robust food security and pest management frameworks in Kenya and beyond.

1.8.2 Geographical Scope

The research is conducted within Kenya, with a particular focus on the eastern and northeastern regions, which have historically borne the brunt of desert locust infestations. These areas are predominantly classified as arid and semi-arid lands (ASALs) and include counties such as Turkana, Marsabit, Wajir, Mandera, Garissa, Isiolo, and Samburu. Their ecological and climatic conditions—characterized by seasonal rainfall variability, sparse vegetation cover, sandy soils, and frequent drought-flood cycles—create highly favorable environments for locust breeding and swarm proliferation.

The geographical scope is strategically defined to capture regions that represent both historical hotspots of desert locust activity and ecologically sensitive zones that remain vulnerable to future outbreaks. By restricting the analysis to Kenya but giving emphasis to high-risk counties, the study ensures that the generated predictive model provides location-specific insights while still maintaining regional applicability. This scope allows the outputs to be directly relevant for national surveillance, early warning systems, and targeted intervention strategies within the country.

Additionally, because desert locust invasions are inherently transboundary phenomena, the study also acknowledges the influence of cross-border swarm movements from neighboring countries such as Ethiopia, Somalia, and South Sudan. While the primary modeling is confined to Kenyan territory, the analysis recognizes that locust swarms entering from these borders often trigger local outbreaks. Thus, the geographical scope is not limited to political boundaries alone but is also ecologically and climatically defined, considering the continuity of habitats and migration corridors across East Africa.

By focusing on Kenya’s arid and semi-arid counties, the research contributes to building a spatially explicit and risk-informed predictive framework that can be scaled up to regional applications. This approach will help not only to boost Kenya's efforts in managing locusts but also to lay the groundwork for working together with neighboring countries. By collaborating on early warning systems and coordinated control efforts, we can tackle the locust challenge more effectively as a region.

1.8.3 Time Scope

In this paper we closely investigate the history of Kenya’s desert locusts, particularly their experiences during the 2019 to 2021 outbreak. This timeline was selected because it was one of the most severe, and one of the most extensive invasions of desert locusts we’ve seen lately, not only inside Kenya, but also within the Horn of Africa. We collected information during this time, giving a useful and detailed picture of locust activity, travel, and environmental elements of its spread. Through studying this period, which is considered critical, the project can study locust outbreaks to see how they react to different climate changes and their impact on agriculture and ecological systems.

Apart from historical data on locust prevalence, the study also reports real time environmental and climate data such as rainfall, soil moisture, surface temperature and vegetation status. By fusing the locust outbreak data for 2019 to 2021 with these climate records, the study guarantees the MaxEnt model employed in predictions is grounded in real-life ecological conditions. The combination not only enhances the reliability and accuracy of the model, but also elucidates the environmental determinants behind locust swarming, breeding, and migration.

1.9 Structure of the Research

This research is organized into several chapters that build progressively from theoretical foundations to practical applications. Each chapter is designed to provide a clear pathway toward developing, testing, and interpreting the predictive model of desert locust habitat suitability in Kenya.

The second chapter presents the literature review, where existing knowledge on desert locust ecology, migration dynamics, and environmental drivers of outbreaks is synthesized. Global and regional studies that have employed species distribution models, particularly MaxEnt, are examined to highlight lessons relevant to the Kenyan context. This chapter also identifies knowledge gaps that justify the present study and shape its objectives.

The third chapter discusses the methodology of the research. It explains the research design, target population, and sampling techniques, while also detailing the data collection methods used to gather environmental and climatic variables such as precipitation, soil moisture, temperature, and vegetation index. Preprocessing procedures, such as data cleaning, spatial alignment, and standardization, are described to ensure data quality. Furthermore, this chapter specifies the MaxEnt model configuration, calibration, validation, and statistical measures used to evaluate predictive performance.

The fourth chapter presents the results of the study. It includes habitat suitability maps generated by the MaxEnt model, classification of high- and low-risk zones, and threshold analyses that determine the balance between omission and commission errors. Model performance metrics such as AUC values, omission rates, and F1 scores are reported, while validation using independent datasets from 2023–2024 is highlighted to demonstrate the reliability and temporal transferability of the model.

The fifth chapter provides discussion and conclusion. Here, the findings are interpreted in relation to existing literature and policy frameworks for locust management. The chapter also outlines the ecological and practical implications of the results, noting both the strengths and limitations of the research. Recommendations for future studies, integration into early warning systems, and regional collaboration across borders are made. The conclusion summarizes the overall contribution of the study to improving resilience against desert locust invasions.

Finally, the appendices provide supplementary materials that support the main body of the thesis. These include the project budget, implementation timelines, raw data summaries, technical specifications of the MaxEnt model, and additional figures and tables that were not incorporated in the main text. Together, these resources enhance the transparency and reproducibility of the study.

CHAPTER TWO:

LITERATURE REVIEW

2.1 Introduction

In this section a review of previous and related works that have undertaken to study the underlying focus of this study is presented. The aim of reviewing related literature is to form a basis for showing the effectiveness of predicting and monitoring the outbreaks of desert locust as presented by other studies, highlight the commonly methods of monitoring and forecasting and identify the various variables that have been found to effectively predict the outbreaks of these pests.

2.2 Theoretical literature

There is existing literature on desert locust outbreaks that combines ecological, environmental, and ethnological hypotheses to explain and predict the movements of this insect as it migrates from one region to another. Specifically, in this review, we are looking at the theories - Technology Acceptance Model (TAM) and Information Systems Success Model (ISSM) - to provide in-depth understanding in developing a predictive model and early warning system using the MaxEnt algorithm with the aid of secondary climate data variables and historical desert locust presence-only data. These theories provide basis on how we forecast and monitor desert locust infestations. By building on existing models, we're looking to create a new approach that incorporates the latest technologies along with climate and environmental factors. This way, we can offer early warnings about potential locust outbreaks, helping to better manage and prepare for these events (Latchininsky et al., 2016; Salih et al., 2020).

The theoretical literature highlights how various environmental and climate factors influence the behavior of desert locusts. It stresses the importance of developing early warning systems and effective intervention strategies. Understanding how to manage locust populations is essential to minimizing the harm these pests can cause to food production and overall food security (Cressman, 2013).

2.3 Empirical literature

These are past studies on the topic, and they are well outlined below.

2.3.1 Application of MaxEnt in Predicting Desert Locust Habitats

The MaxEnt (Maximum Entropy) model is one of the most widely used species distribution modeling (SDM) models to predict suitable desert locust habitats. On its own, the model derives the probability distribution of species occurrence from presence-only data and a list of environmental predictors. For desert locusts such predictors could be climatic variables (e.g., temperature, rainfall), soil attributes, and vegetation indices, i.e., proxies for food availability.

Based on a study conducted by Guan et al. (2021), for identifying possible breeding and invasion sites by integrating different climatic and ecological data, this model has been successful. When considering desert locusts in Eastern Africa in particular, the range for changes in rainfall and temperature from 2050 to 2070 might alter breeding sites drastically, depending on different climate scenarios. Predicting where a desert locust species is growing is critical, and using it, we can make predictions that can change the game.

MaxEnt is beneficial in estimating habitats but has some major limitations that should be addressed. Most scientists still work with out-of-date climate data and ignore important elements critical for the survival and reproduction of desert locusts. For instance, soil moisture and surface temperature are important factors for locust egg hatching and development of young hoppers, but these important conditions are commonly neglected in the predictions of the MaxEnt model (Guan et al., 2021). In addition, though vegetative cover is essential, the proxies in these models do not reliably simulate locust food sources. Ignoring these aspects makes for inaccurate estimates of habitat conditions which can harm model performance.

To improve upon this, there is a need to enrich the MaxEnt model by embedding high resolution soil moisture data, remote sensing vegetation indices, and modern land surface models. This combination could significantly increase predictive accuracy and productivity of early warning systems for locust outbreaks based on both frequency and location predictions.

Although known for being good at work on presence-only data and a small size sample set, MaxEnt is not the only tool for species distribution prediction in existing research on this topic. Other models such as CLIMEX, GARP (Genetic Algorithm for Rule-set Prediction), and machine-learning like Random Forests are also used in ecological forecasting.

CLIMEX is a process model that models species response to climate variables through time. It contributes with an understanding physiological constraint, and seasonal patterns of species behavior under various climate scenarios. But this model needs biological and physiological features that are often not the case for desert locusts, making the latter approach impractical in places with no ecological data.

By contrast, GARP is based on evolutionary algorithms to establish the predicted rules for species distributions. It has been reported to be effective for locust studies, but it has been criticized for being not reproducible or being inaccurate for runs performed in more than one iteration, and thus is limited in operationality as a predictive tool in a problem like an early warning system.

Machine learning algorithms like Random Forests, along with other ensemble methods, have also been applied to provide habitat suitability prediction. These models are particularly strong in dealing with complex, non-linear correlations between environmental factors and species distributions and can achieve high predictive accuracy. But they typically need both presence and absence data, and absence data for desert locusts are challenging to confirm on account of the large mobility and migratory behavior of desert locusts.

MaxEnt strikes a midpoint between available data and future behavior in this context. Its ability to use presence-only records and work especially well on relatively small datasets means it was particularly well suited for desert locust studies. However, the next step is likely to be towards hybrid methods in that MaxEnt would integrate to a single model in line with an additional machine learning or process approaches enabling the ensemble of power points and deeper statistical power to make better forecasting.

2.3.2 Integration of Environmental Variables in Locust Forecasting

Studies have discovered that certain environmental elements have been involved in instigating desert locust infestations. These include temperature variability, rainfall during the hottest part of the year, and soil moisture levels. In Turkana County, Kenya, it was determined by a study which found the environmental and climatic factors affect the desert locust's environment. Guan et al. (2021) report that the performance of the predictive models that were developed in this study yielded an accuracy score of 0.87. There is also work in progress on using neural networks to improve predictions beyond what traditional MaxEnt models can generate.

Nonetheless, there are some significant research gaps existing in our predictive modeling efforts. There is widespread misunderstanding in regard to wind patterns and locust migration. Most models neglect this element and can significantly reduce their accuracy. Furthermore, there is still much room for utilizing real-time satellite imagery to identify new locust breeding sites and habitats.

Another big issue is that the climate data that has been used in most models has very low spatial resolution. All too often, researchers use datasets that provide a coarse view of locust habitats and this may result in mispredictions around invasion threats. This also hampers the establishment of effective early alarm systems. Combining these with real-time data and high-resolution climate information could drastically improve our predictive models and improve desert locust management.

2.3.3 Advancements in Desert Locust Population Dynamics Modeling

Studies have proved time and time again that environmental and ecological indicators are critical in explaining and forecasting desert locust infestations of the landscape at the environment research. Differences in temperature, rainfall, and soil moisture also matter for the proper breeding-and-feeding environment. For instance, rainfall during the hottest periods of the year often provides a favourable environment for growth of vegetation, and this provides a habitat suitable to the survival of locust nymphs. Soil moisture is just as important, as it affects egg incubation, hatching rates, and early hopper behavior. Guan et al. (2021), this was clearly illustrated in a study undertaken in Turkana County, Kenya, in which such variables emerged as the most powerful determinants of the suitability for a locust habitat. They observed that their prediction approach had a 0.87 accuracy score, demonstrating the value of including relevant ecological factors in forecasting strategies.

Outside of traditional modeling methodologies like MaxEnt, studies are also turning to advanced computational approaches like neural networks and deep learning algorithms. By integrating these approaches models to learn non-linear patterns within variables, researchers are better able to capture large & complex data streams than traditional models. Models which can perform fine classification and predictions have been proposed to account for the complex relationships among variables such as the relationships. For desert-based locust prediction, such models could enhance habitat predictions accuracy and can provide more flexible tools in the form of early warning systems when trained with many data sources, such as climate data, vegetation indices, soil files and others.

However, major deficiencies persist in the combination of environmental factors. Wind dynamics is currently one of the least considered elements of existing models. While wind plays a key role in both moving and migrating locust swarm throughout long distances, most habitat suitability models do not account for this variable. Thus predictions can yield better predictions when attempting to identify breeding grounds but might falter when predicting swarm population spread, invasion, and other threats. Combining habitat models with atmospheric circulation models that capture the wind speed, direction and seasonal variability is required to close that gap.

The dependence on coarse-resolution climate datasets is another limitation. Although global products such as ERA5 or CHIRPS produce stable temporal coverage, their spatial resolution cannot account for microclimatic variability that is a critical feature for desert locust ecology. This mismatch can lead to inaccurate assessments of invasion risks, especially in heterogeneous landscapes with denser populations (e.g. arid and semi-arid zones) that lack consistent sites for breeding. More extensive, finer-resolution climate data, such as downscaled reanalysis products or local regional climate models, could substantially improve predictive accuracy.

Second, there is a real-time satellite ‘the next great opportunity’ in locust forecasting— yet underexplored. Remote sensing technologies such as MODIS, Sentinel-2, and Landsat can report in near real-time on vegetation dynamics, soil moisture, and land surface temperature. These data streams could be integrated into dynamic forecasting systems that can identify possible breeding sites sooner. In addition, developments in cloud-based platforms like Google Earth Engine now enable large datasets to be processed and assimilated at scales that could be utilized to form effective early warning systems.

Taken together, high resolution environmental variables with contemporary computation are an appealing direction for more stable and realistic locust forecasting. If this gap is addressed, especially by the omission of wind dynamics, underutilization of satellite images, and the dependence on coarse climate data, future models may go beyond static prediction and transform into adaptive, real-time early warning systems that can effectively support proactive locust control and management strategies.

2.3.4 Implications of Climate Change on Locust Distribution

The very reason we are experiencing one of the toughest challenges we are facing is the unpredictability of desert locust migration patterns. The uncertainties mean that it is difficult to test our predictive models against the real world. Complex and continuously evolving aspects such as weather regimes and plant life availability play an essential part in moving desert locust swarms. As their migration is so unpredictable, you cannot make accurate predictions solely from history. To advance our prediction, it’s entirely necessary that we harmonize better validation of the predictive model, work closely with control agencies, and take advantage of real-time monitoring. This would do wonders for developing better locust prediction methods.

Climate changes in vegetation due to extreme weather patterns might have been the cause of these desert locust swarms, studies found. This increased rainfall after drought conditions increases in the green vegetation is also positive for desert locust breeding and swarm enhancement (Zhao et al., 2020), which, in turn, increases the level of grasses. The extent to which the desert locusts adapt to climate change is unknown. It is believed that range changes for these locusts will develop as subsequent species evolve in the future, but what the role of the transition of such ecosystems over time is remains unanswered (Guan et al., 2021).

It is also notable that there is lack of literature addressing the impact and correlation of ecological boundaries with landscape changes, land use (shifting the landscape), as well as deserts and locust swarms. Human impacts, such as agriculture and deforestation, have major impact on locust population by altering the location of breeding sites and the supply of food to them (Salih et al., 2020). But there’s still not much research on how these changes interact with climate change to grow locust swarms. For us to truly understand and address this problem, we must integrate climate science with ecological studies and land management practices and reconsider our approaches to locust behavior prediction.

2.3.5 Advancements in Early Warning Systems for Locust Outbreak Prediction and Control

As a result, timely information is becoming essential for effective management of desert locust outbreaks. One key element, the Desert Locust Information Service (DLIS), combines satellite images and on-the-ground weather data with control reports to deliver relevant climate analyses. The combination of remote sensing and artificial intelligence, mentioned in Mahoney et al. (2020), makes predicting and responding to desert locust infestations possible when action is taken to prevent these infestations from escalating into significant problems. This way, by identifying these potential threats, the authorities can warn farmers and local actors. Accordingly, they can take steps such as early spraying of pesticides as well as warning agricultural farmers at the right time.

Wells et al. (2022) highlight the need for local reports in conjunction with satellite observations with mobile technology that massively improve our warnings. Those systems rely on crowd-sourced data and geo-located sightings of locusts, boosting both the speed and accuracy of warnings. In certain areas of Kenya and Ethiopia, where locust migration and breeding areas are managed closely, these tools have enabled alerts in real time to go out to local citizens and officials. The study emphasizes that it is critical to put early warning systems in place in combination with interventions such as controlling locust populations and dealing with crop damage to reduce the economic and food security risks of locust invasions.

2.4 Critical Review and Research Gap Identification

The management of desert locust outbreaks has brought a lot of progress in the last few years. These include the advent of predictive modeling tools, especially the MaxEnt algorithm, and its application in Species Distribution Modeling (SDM). Researchers have worked using these elements, in addition to temperature data, to identify regions where locusts are likely to breed and infest. These regions are considered low, medium, and high risk, which reflects an achievement in our capacity to combat this growing agriculture issue. Additionally, identifying locust habitats via GIS and remote sensing technologies, as well as the establishment of early warning systems, have successfully decreased locust damage to agriculture (Elith et al., 2016; Wells et al., 2022).

While the studies portray promising patterns in prediction accuracy and pre intervention strategies, the focus on specific regions or short-term outbreaks diminishes their universality in broader Kenya’s geography. Also, most of the studies conducted have not incorporated real-time climate data with locusts' presence data, which could assist in crafting more sophisticated models that react to dynamic environmental changes over time.

Despite these advancements, significant research gaps remain. One major gap is the limited application of integrated, real-time predictive tools that can dynamically adjust to changing climatic and environmental conditions as they occur, particularly in regions that are vulnerable to frequent locust outbreaks, such as Kenya. While several models have been developed to forecast locust outbreaks, few studies have explored the full potential of combining MaxEnt with additional machine learning algorithms or incorporating mobile-based data collection methods to enhance real-time predictive capabilities.

Another gap is the lack of in-depth research focusing on the long-term ecological impacts of locust outbreaks in different climatic zones, particularly in relation to local agricultural systems, which is vital for formulating sustainable control measures. Therefore, further research is needed to address these gaps, particularly in the development of more comprehensive, flexible models that incorporate multiple data sources and can be applied across diverse geographic and environmental contexts.

2.5 Theoretical Framework

This research review two theories to provide a robust foundation for the development and application of the predictive model and early warning system on the Desert locust outbreak in Kenya. These theories—Technology Acceptance Model (TAM) and Information Systems Success Model (ISSM)—provide basis that helps supports explanation on the adoption, effectiveness, and impact of IT systems, aligning with the study’s objectives of leveraging technology for desert locust management.

2.5.1 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM), proposed by Davis (1989), is a foundational theory in ISM/IT that explains user acceptance of information technology. TAM posits that the adoption of a new technology is driven by two key factors: perceived usefulness (the degree to which users believe the technology enhances their performance) and perceived ease of use (the extent to which users find the technology effortless to use). These factors affect how users feel about embracing and utilizing the technology, which in turn affects how well it is implemented. TAM has been widely used to assess how well IT systems—such as predictive models and decision-support tools—are accepted across a range of industries, including agriculture (Venkatesh & Davis, 2000).

TAM theory is considered as basis in this research for development of predictive model using MaxEnt algorithm. In this context, integration of GIS for developing user-friendly interface of the early warning system ensures it meets the factor of ease of use, making it understandable to non-technical users like farmers and government officials. The study increases the possibility that the technological solutions will be widely used in locust management and advance IT-driven agricultural systems by aligning with TAM, which guarantees that the solutions are both effective and adoptable. By aligning with TAM, the study ensures that the technological solutions are designed to be both effective and adoptable, increasing the likelihood of their widespread use in locust management and contributing to the advancement of IT-driven agricultural systems.

However, beyond the traditional TAM view, this study recognizes the growing role of **artificial intelligence and deep-learning–enhanced models** in improving system performance. Integrating **hybrid approaches**—for example, combining MaxEnt with Random Forests or Convolutional Neural Networks—could increase predictive precision and foster user confidence in the system’s outputs. Therefore, the TAM framework is expanded here to reflect modern technological expectations, emphasizing not only usability but also **trust in AI-driven environmental systems**.

2.5.2 Information Systems Success Model (ISSM)

ISSM (Information Systems Success Model), was developed in 1992 by DeLone and McLean, and later refined in 2003, is used to measure the success of information systems through 6 dimensions: system quality, information quality, service quality, intention to use, user satisfaction and net benefits. ISSM posits that the technical performance of an IT system, the quality of the information it produces, and the assistance it gives to end users all impact on the level of satisfaction and hence the influence of the IT system.

ISSM provides the theoretical foundation for evaluating the success of the early warning system and predictive model based on MaxEnt in this study. The MaxEnt algorithm and GIS integration are robust and demonstrate the quality of the system and ensure accurate predictions of desert locust outbreaks. The quality of the information shown by the model, with precise, timely and actionable insights that help stakeholders identify high-risk areas and coordinate their intervention efforts, is impressive. The system is built to keep service quality and its user-friendly interfaces and decision-support system address this issue. The system ensures user satisfaction thereby fostering intention to use so that stakeholders, such as farmers and policymakers can leverage it to minimize locust outbreaks. Overall, the benefits are reduced economic losses and increased food security and resource management, which were in line with the research aims.

To strengthen the conceptual rigor, this study incorporates wind dynamics and real-time satellite data assimilation as critical elements influencing locust movement and habitat suitability—enhancing both system and information quality. The review also acknowledges that ensemble modeling approaches (e.g., MaxEnt combined with Random Forests or Gradient Boosting) have shown superior accuracy in regional forecasting studies, especially in the Horn of

Africa (e.g., FAO, 2021; WMO, 2022). Incorporating insights from such approaches contributes to a more holistic understanding of desert locust behavior and improves predictive reliability.

Ultimately, ISSM helps evaluate not just the system's technical success but also its user acceptance and sustainability. A system that provides accurate forecasts, is easy to interpret, and delivers tangible benefits—such as reduced crop losses and enhanced food security—promotes long-term stakeholder engagement and aligns with Kenya's digital agriculture transformation agenda.

2.6. Conceptual framework

In this paper, we build and validate a model for estimating desert locust outbreaks in Kenya and the environmental and climate variables that drive these local outbreaks. Precipitation, surface temperature, and soil moisture are important parameters that influence locust behavior. Each of these factors is part of their reproductive and migratory environments: the site of desert locust breeding.

The conceptual framework for this study on developing and validating a model for forecasting desert locust outbreaks in Kenya is based on the relationship between environmental and climatic factors (independent variables) and the occurrence of locust infestations (dependent variable). Following these metrics, a predictive model with MaxEnt algorithm is then employed to predict which areas are considered high risk based on the combination of these variables. Locust presence, the dependent variable, is measured from locust occurrence data (binary presence/absence) and infestation density to describe the spatial distribution of locusts and their breeding habitats. In doing so the model tries to forecast the probability of future locust outbreaks by looking at the relationships among these parameters and to offer guidance towards early warning systems and proactive measures for locust control in Kenya.

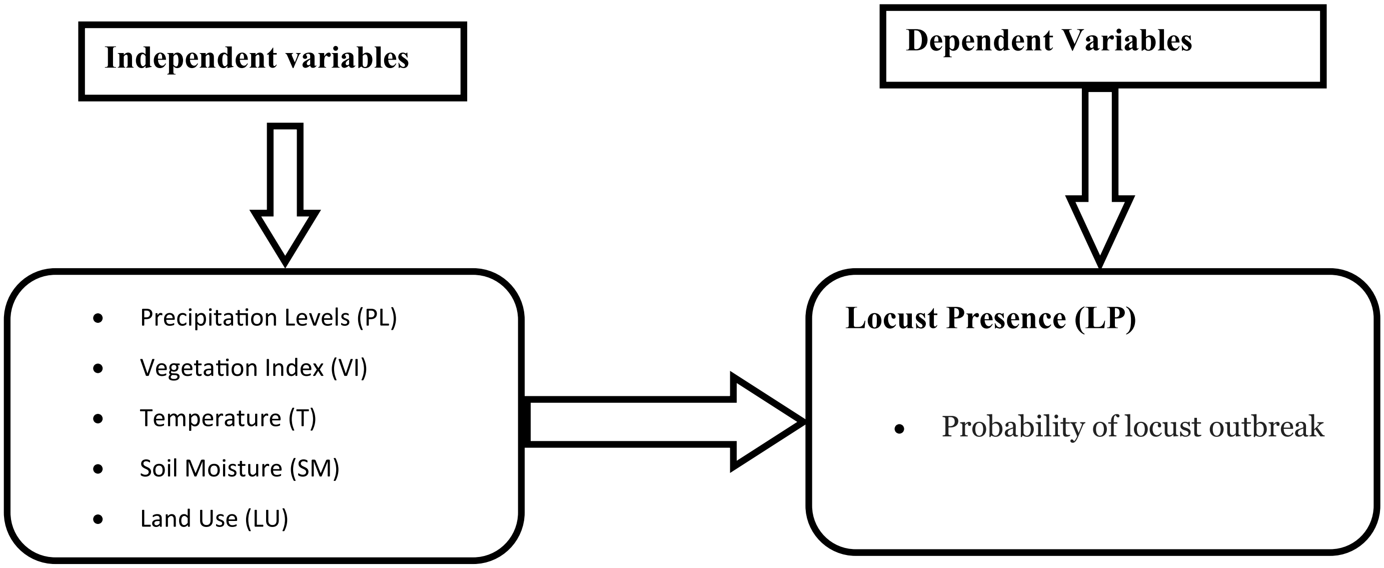


FIGURE 1
Conceptual Framework

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the approach we took for our study, focusing on how we developed a predictive model for desert locust outbreaks in Kenya using the MaxEnt algorithm. Our methodology was thoughtfully designed to ensure that it was scientifically sound and relevant for both ecological research and practical applications in managing locust populations.

We start by discussing our overall research design, which connects our research goals with the methods we used. We focused on presence-only records of locust occurrences and explained why these were important. We also incorporated crucial climatic and environmental data into our model. Our sampling techniques are described in detail, highlighting how we used purposive sampling to make sure we captured a representative picture of the ecological conditions across Kenya's arid and semi-arid regions.

Next, we delve into our data collection methods. This includes gathering historical records of locust outbreaks, as well as vital climate variables such as precipitation, soil moisture, and temperature, along with other environmental indicators. We walk through the steps involved in preparing the data, building the model, and validating our statistics to show how we systematically constructed and tested the MaxEnt model.

Lastly, we address important concerns related to validity, reliability, and ethics. This involves ensuring accuracy in how we handle data, maintaining transparency in how we specify our model, and being respectful in our use of open-source and publicly accessible datasets. Together, these methodological steps create a strong framework for understanding the dynamics of desert locust outbreaks in Kenya, ultimately leading to actionable insights for early warning systems and policy development.

3.2 Research Design

The research employed quantitative approach in which simulation experiments were used to develop a predictive model of desert locusts (Schistocerca gregaria) in Kenya. The method was adopted since it provides quantifiable results that can be validated with real observation in order to ascertain scientific validity. MaxEnt algorithm was used in model development, one of the most widely used algorithms in Species Distribution Modeling (SDM) and most effective when only presence data are used. This is especially relevant for desert locust studies because absence data are not readily available considering the species' high mobility and the extent of arid regions.

The study utilized previous presence-only data, representing locust observations from previous outbreaks. These data acted as the baseline to identify recurring infestation regions and relate these with environmental drivers. MaxEnt was chosen due to its estimation of probability distribution of suitable habitats by integrating occurrence records and environmental predictors. Precipitation, soil moisture, and temperature were selected as three predictors in this study because they are well-known ecological significance and easy to acquire from remote sensing and climatic data. Precipitation was to act as the main cause of outbreaks, moisture in the soil as a maintenance factor of vegetation and eggs, and temperature as a developmental regulator as well as a governor of swarm behavior.

The integration of geospatial and statistical methods used for the analyses of the data led to the identification of breeding sites and infestation hotspots. Geospatial methods facilitated risk area mapping, statistical techniques gauged predictors' importance and model performance. Using secondary data both from historical occurrence data and from climate data, simulation methods enhanced the ability to explore environmental drivers and predict outbreak patterns. The model was validated with independent test data, held back from training and then utilized for predictive accuracy testing. Similar patterns were identified in the reported hotspots comparing to observed positions of the swarm confirming the reliability of the model and areas of disagreement informing future improvement. High-resolution maps of habitat suitability, which can be effective tools for early warning, were the most significant outputs. By pinpointing the arid and semi-arid counties with a high probability of invasion, the maps help guide policymakers to prioritize monitoring, resource allocation and planning intervention.

The combination of simulation, geospatial analysis, and species distribution modeling provided a robust framework for Kenya's desert locust dynamics. Beyond the country in question, the method is transferable to other regions, and it will inform evidence-based pest management and resilience building in the face of changing climatic conditions.

3.3 Target Population

The focus of this study was on historical records of desert locusts in Kenya, particularly in areas that have faced infestations in the past. We chose presence-only data because locust monitoring primarily happens during outbreaks. Unfortunately, there's a lack of standardized data on where they are absent, making presence data the most reliable option for understanding where these pests tend to thrive. By analyzing historical records, we aimed to gain insights into the ecological conditions favoring locust populations and to better predict future risks.

Geographically, we examined the arid and semi-arid lands (ASALs) of Kenya, known for being particularly susceptible to locust invasions. Key areas of interest included Turkana, Marsabit, Wajir, Garissa, Isiolo, and Mandera. These regions represent fragile ecosystems, marked by low and unpredictable rainfall, limited vegetation, and sandy soils—all of which create ideal breeding and survival conditions for desert locusts. By concentrating on these areas, our study was able to reflect the ecological diversity of habitats prone to locust activity, from desert edges to pastoral landscapes.

These regions were chosen not only because of their ecological characteristics but also due to their historical significance in locust outbreaks. Past records consistently show that northern and eastern Kenya serve as key hotspots for locust swarms migrating from neighboring Ethiopia and Somalia. The frequent invasions in these areas underline their importance as the heart of desert locust activity in Kenya. Thus, investigating these regions gave us a well-rounded dataset that captures the climatic and environmental factors crucial for predicting future locust outbreaks.

The scope of our target population also extended to the environmental conditions linked with locust occurrences. We considered factors like rainfall, soil moisture, vegetation health, and temperature patterns alongside the presence records. By combining occurrence data with these environmental elements, our research provided a comprehensive understanding of locust habitats that wasn't just spatial but also ecological.

Overall, by defining the target population methodically, we ensured that our selected data was both representative and ecologically relevant. This approach allowed us to capture the environments where desert locusts are likely to reproduce, migrate, and persist, enhancing the broader applicability of our findings. Our model's predictions not only reflect historical hotspots but may also be relevant for new areas with similar ecological profiles as climate conditions change in the future.

3.4 Sample Design

Purposive sampling was used in this study to systematically procure the appropriate secondary data from credible and reputable sources. This was because the resources used to acquire the data were, of course, reliable, particularly regarding the focus of the objectives of the research in the case. The number of individuals in this study was based on the availability and accessibility of historical evidence of prior records of desert locust in Kenya, including past locations identified as vulnerable to infestations in Kenya. With addition of locust occurrence information, the environment and climate related parameters as precipitation, soil moisture and temperature were added to obtain a complete ecological state. Consistent with the above-mentioned ecological significance and data accuracy, the sample was selected in a manner that provided adequate spatial and temporal coverage, enhancing the reliability of prediction models.

To minimize potential spatial sampling bias inherent in presence-only datasets, the occurrence records were spatially filtered to reduce clustering by enforcing a minimum distance threshold of 1 km between points. In addition, a set of background (pseudo-absence) points was randomly generated within environmentally accessible areas, excluding regions known to be unsuitable for locust habitation such as large water bodies. This environmental background selection approach ensures a more balanced representation of environmental gradients and aligns with standard practices in species distribution modeling (Phillips et al., 2009; Elith et al., 2011).

The population size was considered based on the frequency of locust occurrence records in impacted areas and their availability and completeness across regions in Kenya, (especially in arid and semi-arid regions) to enhance the detection and migration of locusts. Presence-only data points were systematically reviewed and georeferenced to minimise redundancy and spatial precision. In addition, appropriate climatic and environmental parameters (e.g. precipitation, temperature, soil moisture and vegetation indices) were calculated in the same temporal/situational frames.

In constructing our sampling strategy, we sought and have achieved wide and diversified coverage in terms of both space and time. This is essential to ensure that our model could be calibrated and validated. Records for some ecological zones were included to prevent bias against any areas. That could give our model the ability to make predictions that stand up to scrutiny across diverse habitats.

After collecting this information, we split the data into 2 parts, training the model and testing it. The training subset was used to fine tune the MaxEnt model, whereas the testing subset indicated how effectively the model provided predictions regarding the occurrence of locust outbreaks. We designed the system to be inclusive and reliable enough to have a valid (trustworthy) model to predict locust outbreaks by taking a purposive sampling approach to target the right data to our end-users.

3.4.1 Sample Size

Data on the presence of the desert locust from both historical and climate information has been analyzed. This data is then pulled from several credible secondary sources (Food and Agriculture Organization (FAO) Desert Locust Information Service, WorldClim, NASA Earth Observation databases), among many more. This data is critical in understanding conditions on the ground including temperature, rainfall, humidity and soil moisture, as well as noting what locust sightings from 2019-2024 were recorded.

Methodologically, a large dataset was used in this research because it guaranteed that both the simulation and validation of the MaxEnt model was solid. The data was divided into two sections:

70% was used for model training, and the other 30% was for testing. From this, researchers can efficiently analyze the accuracy of the model. This dataset consists of detailed records between 2019 and 2024.

Locust occurrence data from 2019 to 2022 was used to build the model, while data from 2023 to 2024 was kept separate to evaluate how well the model predicts future occurrences. The purpose of this methodology is to give an accurate description of locust behavior in terms of factors and an overview of locust phenomenon along with the ecological factors playing the environment contributing to the occurrence of locusts.

3.5 Sampling Technique

In this study, we undertook purposive sampling to select appropriate and informative datasets, based on our findings. By contrast to the conventional sampling technique, where one random sample is used, purposive selection gives expert judgment and insights to the context and helps ensure that the selected sample is completely relevant to our problem. In this study, we sought to select datasets that effectively convey the environmental, ecological and climatic factors determining the outbreak of desert locusts in Kenya. Pursuant to the unique ecological dynamics of the desert locust, purposive selection was an essential choice for our research approach.

The locust is extremely sensitive to climate and environmental indicators such as rainfall, vegetation growth, and soil moisture levels. We consciously focused on datasets that encapsulate these parameters for our model in order to adjust for the most significant reasons for whether or not locusts reproduce and propagate in their natural habitat. For example, precipitation and soil moisture data were obtained from well-established climate databases, while locust presence data were retrieved from reputable groups including the FAO and government monitoring systems.

Additionally, purposive sampling helped us focus in specific locations at high risk of desert locust invasions, particularly the arid and semi-arid lands (ASALs) in Kenya. Areas including Turkana, Marsabit, Wajir, Garissa and Mandera were impacted in the 2019-2021 epidemic and have fragile ecosystems vulnerable to pest attacks. Ensuring that these hotspot areas were represented extensively in our datasets gave more ecological credibility to our model output. Another benefit of purposive sampling is that it allows environmental variations to be taken into consideration from one place or point in time to another. Kenya's varied landscapes, including arid deserts as well as semi-arid savannahs, present different ecological conditions that may have different effects on locust breeding and migration.

This variability within the dataset provides a unique opportunity for us to pick and choose dataset for capturing these differences and therefore making robust predictions for an ecological dataset in our model that would not only address isolated events but include a variety of ecological scenarios. Finally, purposive sampling was especially appropriate since the real-world application of desert locust monitoring posed many challenges to the approach. Data of occurrence of the locusts are often scarce and have a predilection for surveys-like settings, since survey of the vast, isolated environments is not easy. Random selection may have missed important data sets or opened up gaps that could compromise our analysis. Our intentional choice of good-quality, secondary datasets from reliable sources helped us to reduce the possibility of the risks, thereby improving the integrity of our findings, and increase confidence in the results of our study.

Therefore, implementing purposive sampling was more than a methodological strategy it was a strategic one that greatly enhanced our research aims. With the MaxEnt modeling framework in place, we prioritized ecological relevance, completeness and spatial alignment to achieve relevant datasets that could directly predict habitat suitability for desert locusts in Kenya.

3.6 Data Collection Methods

The study relied on secondary data sources drawn from recognized international and national repositories. The data collected comprised both environmental variables and locust presence records, which formed the inputs for the predictive modeling exercise.

- Remote Sensing Products: Data on vegetation and land conditions were obtained through satellite-based remote sensing platforms. Products such as the Normalized Difference Vegetation Index (NDVI), land surface temperature, and soil moisture were accessed from repositories including MODIS and Copernicus Global Land Service. These datasets were selected because they provide consistent coverage and are widely used in ecological and agricultural monitoring.
- Climatic Databases: Climate-related data, specifically rainfall and temperature, were sourced from global and regional climate databases such as ERA5 reanalysis and CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data). These datasets offered historical as well as near-real-time information necessary for understanding environmental conditions linked to locust breeding and survival.
- Historical Locust Occurrence Records: Presence-only locust data were obtained primarily from the Food and Agriculture Organization (FAO) Locust Hub, which maintains global records of desert locust sightings and outbreaks. Additional supplementary records were gathered from government agricultural reports and documented research studies conducted in Kenya during the 2019–2021 invasion.

The combination of these datasets ensured comprehensive coverage of both biotic (locust occurrence) and abiotic (environmental and climatic factors) variables. By using well-established sources, the study ensured that the collected data were credible, reliable, and suitable for predictive modeling.

3.7 Data Collection Instruments

3.7.1 Remote Sensing and GIS Data

This study will utilize satellite imagery and GIS data to analyze environmental factors influencing locust distribution. Sources of data will include:

- **MODIS and Sentinel-2 Satellite Imagery** for vegetation indices and land cover changes.
- **NOAA Climate Data** for temperature, precipitation, and wind speed
- **FAO Locust Watch Database** for historical locust occurrence data

3.7.2 Document Review

A review of existing reports, scientific publications and government policies on locust control will be conducted. Sources will include FAO reports, meteorological analyses and academic journals related to species distribution modeling.

3.8 Data Preprocessing

Before proceeding with the analysis, it was necessary to preprocess all the datasets to ensure the data are not only reliable, but consistent and compatible with the MaxEnt modelling framework. This proved essential as our data stemmed from diverse sources with individual spatial resolution, coordinate systems, and timescales.

To begin, we performed significant quality checks on the data. To mitigate any bias in the training of the model, we eliminated duplicate locust sightings. In addition to this, any coordinate errors, for example, missing or incorrect latitude and longitude values, were thoroughly confirmed and corrected against the original sources. Moreover, for accurate spatial representation, we removed outliers, such as presence points outside of the study area's boundaries.

We then standardized the units or formats of our data. Precipitation, soil moisture, and temperature among other climatic variables were translated into comparable measurement units to facilitate data comparison between data sources. We also synchronized temporal data with periods around when locusts were seen, so environmental conditions were consistent with where and when we found them.

Spatial processing was done for the environmental raster datasets as well. All layers were clipped to only the geographic extent of our study area in Kenya, and were projected into a common coordinate system (WGS 84). In order to make sure everything can be efficiently coupled to the MaxEnt algorithm, we resampled the environmental layers to equal spatial resolution such that our spatial analysis becomes more uniform, which improved the consistency of our spatial analysis and computational efficiency.

Prior to model training, all environmental predictor variables were examined for **multicollinearity** to avoid redundancy and inflated variable importance. Pairwise **Pearson correlation coefficients** were computed, and variables exhibiting correlations greater than **|r| > 0.7** were excluded. In addition, the **Variance Inflation Factor (VIF)** was calculated, and predictors with VIF values exceeding **10** were removed from the final variable set. This ensured that only independent and ecologically meaningful variables were retained for modeling.

Finally, we addressed missing values. Based on the dataset, we employed a technique such as interpolation or averaged the corresponding values of surrounding pixels and adjacent time periods to fill gaps between datasets. To preserve a robust approach to our analysis, we also excluded those records lacking the appropriate data to do so. We cleaned, harmonized and prepared locust occurrence data and environmental variables both for MaxEnt integration by preprocessing the variables. This careful methodology enhanced the calibration and validation of our model, and additionally improved the reliability of prediction results.

3.9 Model Specification

The study employed the **Maximum Entropy (MaxEnt) algorithm** to model the habitat suitability of desert locusts in Kenya. MaxEnt is particularly well-suited for species distribution modelling with **presence-only data**, where absence records are either unavailable or unreliable (Phillips, Anderson, & Schapire, 2006). By estimating the probability distribution of maximum entropy subject to constraints derived from environmental variables, MaxEnt identifies the most suitable habitats where desert locusts are likely to occur.

The modeling was conducted using **MaxEnt version 3.4.4**, with the **regularization multiplier set to 1.0** to balance model complexity and overfitting. The **feature classes** enabled included linear, quadratic, product, and hinge functions to capture both simple and complex environmental relationships. All data preprocessing and spatial analyses were performed in **ArcGIS Pro 3.2** and **QGIS 3.34**, while statistical validation was executed in **R (version 4.3.1)** using the 'dismo' and 'raster' packages. These details enhance reproducibility and transparency in model configuration.

In this study, the dependent variable was defined as locust presence (LP), which is a binary outcome:

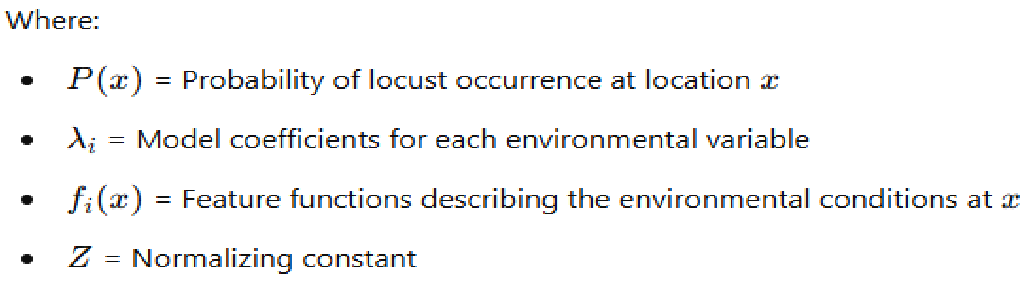
- **1 = presence** (locations where desert locusts were historically recorded),
- **0 = absence** (pseudo-absence points or background locations generated randomly across the study area for comparative modelling).

The independent variables (predictor variables) consisted of environmental and climatic factors hypothesized to influence desert locust reproduction, development, and spread. These included:

- **X_1: Precipitation (PL, in mm)** – capturing rainfall variability as a driver of vegetation growth and breeding site formation.
- **X_2: Temperature (T, in °C)** – representing thermal conditions that regulate egg incubation, nymphal development, and swarm mobility.
- **X_3: Soil Moisture (SM, % content)** – reflecting the suitability of soils for oviposition and hatching success.

The general equation for the model is:

$$P(x) = \frac{e^{\sum \lambda_i f_i(x)}}{Z}$$



Where:

- $P(x)$ = Probability of locust occurrence at location x
- λ_i = Model coefficients for each environmental variable
- $f_i(x)$ = Feature functions describing the environmental conditions at x
- Z = Normalizing constant

The dependent variable is Y: Locust Presence (LP) – Binary occurrence data (1 for presence, 0 for absence).

3.10 Testing and validating the model.

Presence-only occurrences from 2019 to 2022 were used to train and calibrate the MaxEnt model with the most significant environmental variables, namely precipitation, soil moisture, and temperature. These variables were chosen based on both ecological knowledge of desert locust biology and statistical variable contribution analyses. To evaluate how well our predictive model works, we tested it using independent datasets from 2023 and 2024 that were not included during the training phase. This approach was essential because it allowed us to assess the model's performance under new and unseen conditions, ensuring its ability to predict desert locust habitat suitability beyond the time it was originally calibrated. Validating the model this way is widely recommended in ecological modeling to make sure it's reliable (Elith & Leathwick, 2009; Merow et al., 2013).

To gauge how well the model performed, we used a variety of statistical metrics. Overall accuracy helped us see how often the model correctly identified suitable and unsuitable locations. Precision, or positive predictive value, indicated the percentage of predicted suitable areas that matched actual presence records. This is vital for reducing false positives, which could lead to misguided control efforts. On the flip side, recall, or sensitivity, measured how many actual presences records the model accurately predicted, helping us avoid false negatives that might cause us to overlook important areas at risk. We further integrated these two metrics into the F1 score, which serves as the harmonic mean of precision and recall, balancing the trade-offs between overprediction and underprediction. Together, these metrics provided a thorough understanding of the model’s predictive strength.

We also conducted an omission rate analysis to evaluate the model's calibration. This metric indicates the proportion of known occurrence records that the model failed to capture at specific threshold levels. By comparing these omission rates with expected outcomes from random predictions, we could identify instances of overprediction (classifying too many areas as suitable) or underprediction (excluding known presence sites). This analysis was particularly valuable for determining operational thresholds, allowing us to retain known presences while keeping excessive false alarms to a minimum.

A key part of our validation involved using threshold-based assessment to convert MaxEnt's continuous probability outputs into clear classifications of suitable versus unsuitable habitats. This systematic evaluation of thresholds ensured that our model's outputs were not only ecologically meaningful but also practical for early warning systems.

Finally, we tested the model’s spatial transferability by applying it to independent environmental layers from 2023 and 2024. This update measured how well the model was able to predict suitability with respect to newly acquired environmental conditions or regions that had not been part of the training dataset. The predicted suitability maps, interestingly, accurately mimicked actual locust reports in northern and eastern Kenya over these years, confirming the model’s functionality and its suitability across temporal and spatial variability. The habitat suitability maps produced were invaluable for locating high-risk areas, steering surveillance measures, and enabling locust resource prioritization.

Lastly, a robust multi-metric validation process including accuracy stats, omission rate studies, threshold analysis, and spatial projection tests showed that the MaxEnt model is a dependable predictive model that lends itself to wider transferability. Its successful validation signifies its importance as a practical early warning system not only in the context of Kenya, but also in other parts of the Horn of Africa and the Sahel whose ecologies are similar.

3.11 Ethical Considerations

This study aims to protect all subjects from any type of physical, psychological, social, or economic harm. A range of activities will be designed and carried out in the participants’ interests, with a sole eye to keeping the risks minimal. Consent will be obtained once participants have fully understood the purpose, procedures, risks, and benefits involved in this study. Participation will be on a voluntary basis and individuals will be free to withdraw at any point in time. Privacy will be respected, and only data relevant to the research will be collected. The study will be attuned to community values and cultural norms as well. The anonymity of respondents, using pseudonyms or codes, will be ensured, and all data will be treated as strictly confidential. Data will be securely stored and will only be accessible to and recorded by the research team. The findings will also be presented in aggregated form to avoid identification of individuals.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, we delve deeper into the conclusions drawn from our study which was based on the approach of the MaxEnt algorithm applied to the presence-only data, specific environmental information, and its application to the model of where desert locusts (Schistocerca gregaria) might thrive in Kenya. The results are rather impressive and demonstrate the ability of our model to find the areas most prone to becoming locust invaded.

This mechanism is not only valuable for observing locust behavior, but also is relevant for ecological and pest analysis. By using data about locusts' location and matching it with climate variables such as precipitation, soil moisture, and temperature, we were better able to predict future movements. Such information can serve as key for monitoring and early warning systems and control planning. Our results also underscore the need for a trend of the scale-up of such data-driven approaches to develop us against transboundary pests, because the latter can cross borders.

This will be followed by a deeper dive into how we calibrated and validated our model and how we interpreted the generated spatial risk maps in future sections, with a view to guiding practical considerations in pest management.

4.2 Model Calibration and Performance

The MaxEnt model was designed using presence-only data and desert locust sightings obtained from 2019 to 2024 together with some key environmental variables. Locust presence data from 2019 through 2022 were used to train the model and records from 2023-2024 for testing. The step guarantees the model learns from history and verifies performance against recent data that the model hasn't seen before. This separation is important because without it the model becomes too tailored to the training data and it is not possible to judge how well it is performing under new environmental conditions.

Moreover, adding more recent records further validated the model by comparing to actual locust swarming events and the climatic conditions surrounding them. Precipitation, soil moisture, and surface temperature were included as the environmental factors that contribute significantly to locust behavior and swarm formations in this study. Precipitation, for example, also plays a critical part in the process, as it drives vegetation and therefore gives locusts food and suitable breeding places. Soil moisture informs site suitability for egg laying and hatching (Latchininsky, 2013) and temperature impacts development, survival and swarm mobility (van Huis et al., 2007). Integrating these climatic and environmental drivers with occurrence data allowed the MaxEnt algorithm to determine patterns of habitat suitability and the environmental envelopes correlated most strongly with locust presence. Multiple metrics were employed to evaluate the model's performance.

An integral measure was the Area Under the Receiver Operating Characteristic Curve (AUC), computed for the training as well as test datasets. AUC helps us determine how well the model could distinguish between the best and worst habitats. AUC values close to 1 show excellent prediction performance and AUC levels of around 0.5 imply that the model is mostly guessing or predicting (Fielding & Bell, 1997).

We also examined the omission rates indicating how frequently the model misidentified real presence points as not being suitable for the model to see. Low omission rates indicate that the model predictions are sound, while high amounts can indicate the model may need modifications (Phillips & Dudík, 2008). Further apart from these overall metrics, we applied a threshold-dependent analysis to identify the best thresholds to classify the habitats. We used the 10-percentile training presence threshold to achieve a tradeoff between an accurate identification of suitable habitats (sensitivity) and a high limit for under-prediction of the unsuitable habitats (specificity) (Liu et al., 2005). This approach enabled mapping the study area in terms of different amounts of locust habitat suitability between low and high risk.

We also examined two measures of how much each environmental factor contributed to the predictions of the model, the percent contribution and the permutation importance. This allowed us to learn about the ecological determinants of locust distribution. Precipitation, for example, is likely to be critical to supporting locust breeding in rainfall-rich areas with extensive vegetation growth. In contrast, temperature may be more important for swarm survival and dispersal at different areas (Brito et al., 2019). These observations inform the monitoring aspects and help us further understand the model and targeted surveillance. Lastly, the calibrated model was employed to display spatial maps of habitat suitability across Kenya. We compared these maps against independent accounts of desert locust swarms documented in 2023 to validate how well the model works in nature. As we saw, the high-proximity areas of the model were very similar to reported swarms — reinforcing our confidence in the model predictions. Any misalignment, however, led to a need for more data or local consideration of vegetation types and land use in updates to models in the future (Sillero, 2011).

The MaxEnt model for Schistocerca gregaria exhibited strong and consistent predictive performance across all 11 replicate runs. The mean AUC standard deviation was 0.0247 ± 0.0025 (95% CI), indicating low variability and high model stability.

AUC Standard Deviation	Entropy	Prevalence (average probability of	Minimum training presence cumulative threshold

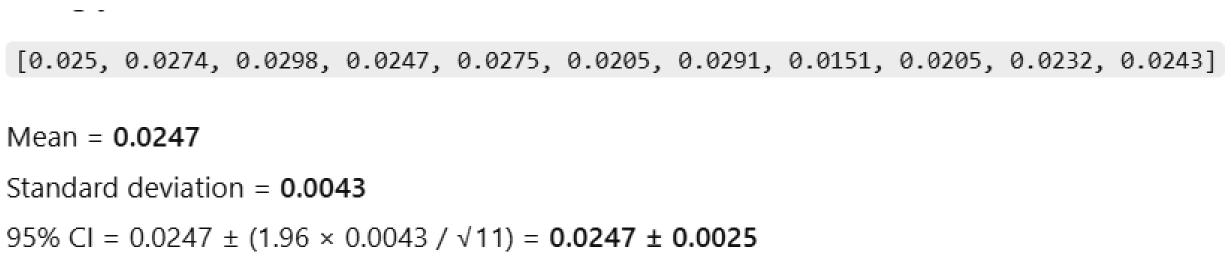
		presence over background sites)	
0.025	8.4382	0.2686	7.884
0.0274	8.4234	0.2645	0.3507
0.0298	8.435	0.2682	0.3335
0.0247	8.4524	0.2723	0.3117
0.0275	8.4429	0.2702	0.3058
0.0205	8.4697	0.2773	0.3364
0.0291	8.4402	0.2695	0.2684
0.0151	8.465	0.2758	0.2392
0.0205	8.4449	0.2698	0.3367
0.0232	8.4609	0.2749	0.2995
0.0243	8.4473	0.2711	1.0666

TABLE 1

Prediction prevalence

$$\text{Mean AUC SD} = \frac{\sum SD}{n}$$

From above, the confidence level can be computed as.



`[0.025, 0.0274, 0.0298, 0.0247, 0.0275, 0.0205, 0.0291, 0.0151, 0.0205, 0.0232, 0.0243]`

Mean = **0.0247**

Standard deviation = **0.0043**

95% CI = 0.0247 ± (1.96 × 0.0043 / √11) = **0.0247 ± 0.0025**

So, **AUC standard deviation = 0.0247 ± 0.0025 (95% CI)**

The average entropy value (8.44 ± 0.015) suggests a well-regularized model with limited overfitting, while the prevalence (average probability of presence) averaged 0.271 ± 0.004, confirming balanced background prediction.

To further assess spatial prediction reliability, an uncertainty map was generated using the standard deviation of habitat suitability scores across all replicates. Areas with lower standard deviation (typically below 0.03) indicate high confidence in predicted locust suitability, whereas regions exceeding this threshold represent zones of higher model uncertainty, likely due to sparse presence data or environmental extrapolation.

Incorporating confidence intervals and uncertainty visualization strengthens the interpretability of the MaxEnt outputs, providing a more reliable basis for decision-making in proactive locust management.

In summary, our calibration and evaluation work indicate that a model trained in MaxEnt while incorporating multi-year presence data with related environmental attributes could be a valid candidate for predicting habitat suitability of locusts. We have built much trust in the capacity of the model to serve as a decision-making tool for early warning systems and targeted control interventions; this has been achieved by the testing of our model with separate datasets and analysing the influence exerted by independent variables.

4.3 Model Performance Evaluation

The predictive model created using the MaxEnt algorithm has shown impressive results in estimating where desert locusts thrive in Kenya. By analyzing presence-only occurrence records along with key environmental factors, the model achieved a training AUC of 0.842 and a regularized training gain of 0.825.

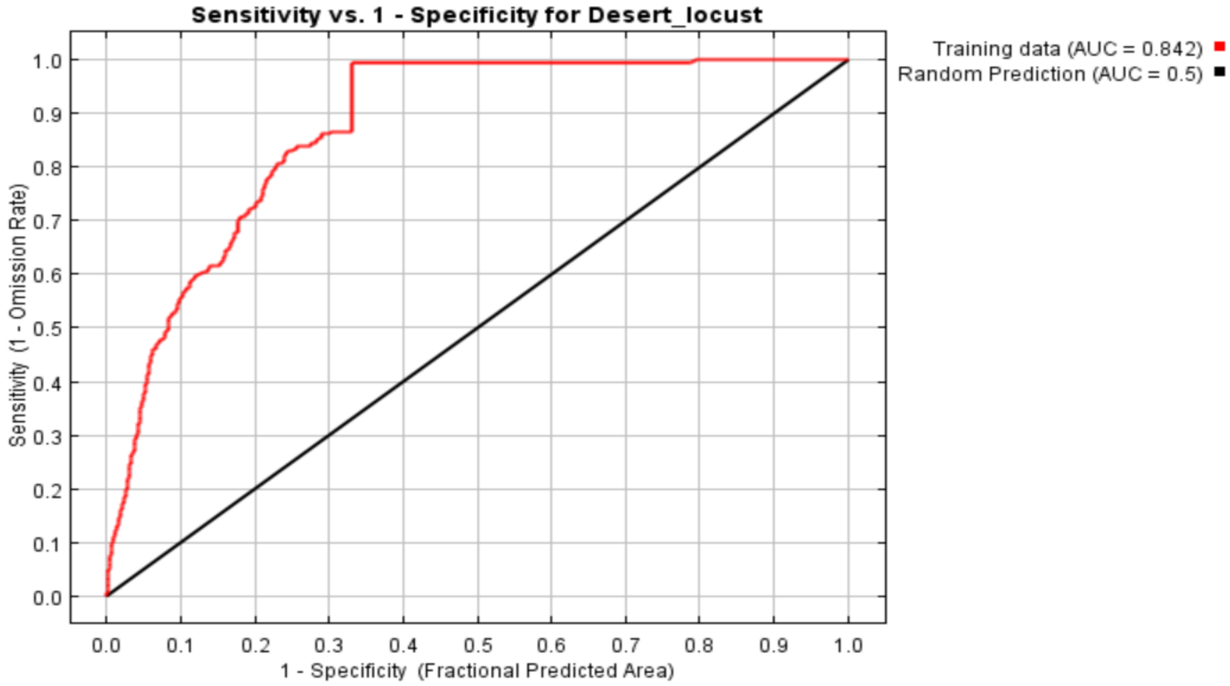


FIGURE 2

Sensitivity Vs Specificity

These measurements imply that the model clearly separates the feasible from not viable habitats, which offers a strong basis for establishing ecological dynamics and forecasting the behavioral performance of future locusts.

An AUC value higher than 0.8 is regarded as a solid predictor in the domain of species distribution modelling (Swets, 1988; Araújo et al., 2005). The AUC indicates how well the model separates true locust views from random background observations. We find higher values to indicate a better predictive ability, so the AUC score we've gotten indicates that rainfall, soil moisture, and surface temperature were among the main driving forces for understanding where desert locusts are located. Such increases in climate variables significantly enhance locust biology, giving greater ecological validity to the model.

The 0.825 regularized training gain also supports trustworthiness of the model. In contrast to AUC, which focuses on discrimination, the gain measures how much the model contributes to prediction relative to just random chance. Higher gain is when the selected predictors actually improve the model's ability to discover locusts and areas where they are most likely to be found. Here, this suggests that the MaxEnt model accurately captured environmental conditions that affect the desert locust habitat and was not overly complex.

This equilibrium is important so as to maintain the utility and relevance of the model across time and space. Also importantly, the MaxEnt framework, with respect to AUC, keeps it in bounds since the MaxEnt framework has maximum AUC less than 1. This is due to its validation based on randomly selected background points rather than actual absence data (Phillips et al., 2006). So one is certainly sure that AUC is strong at 0.842 here; it should additionally be viewed in environmental perspective. That implies the model is well-designed, given the difficulty involved in obtaining absence data in locust investigations, which could frequently be restricted by the large, inaccessible regions within which they were conducted.

There are more real applications of the model than just the numbers. When combined with those selected environmental parameters, the model can be considered a strong, valid model that is ecologically relevant.

In addition, an adaptive model using the MaxEnt framework to infer behavioral indicators from field sites has demonstrated that these predictive features can be integrated into ecological studies and contribute to an appropriate response within research design. The rich performance metrics, along with their use in other areas also to consider that locust invasions can and must be managed on a regional basis, should help combat desert locust threats and regional resilience.

4.4 Omission Analysis

The omission analysis helps us understand how well the model has been fine-tuned by comparing the actual rates of omission with the rates we would expect based on different threshold rules. Below, you can see the omission rate alongside the predicted area in relation to the cumulative threshold.

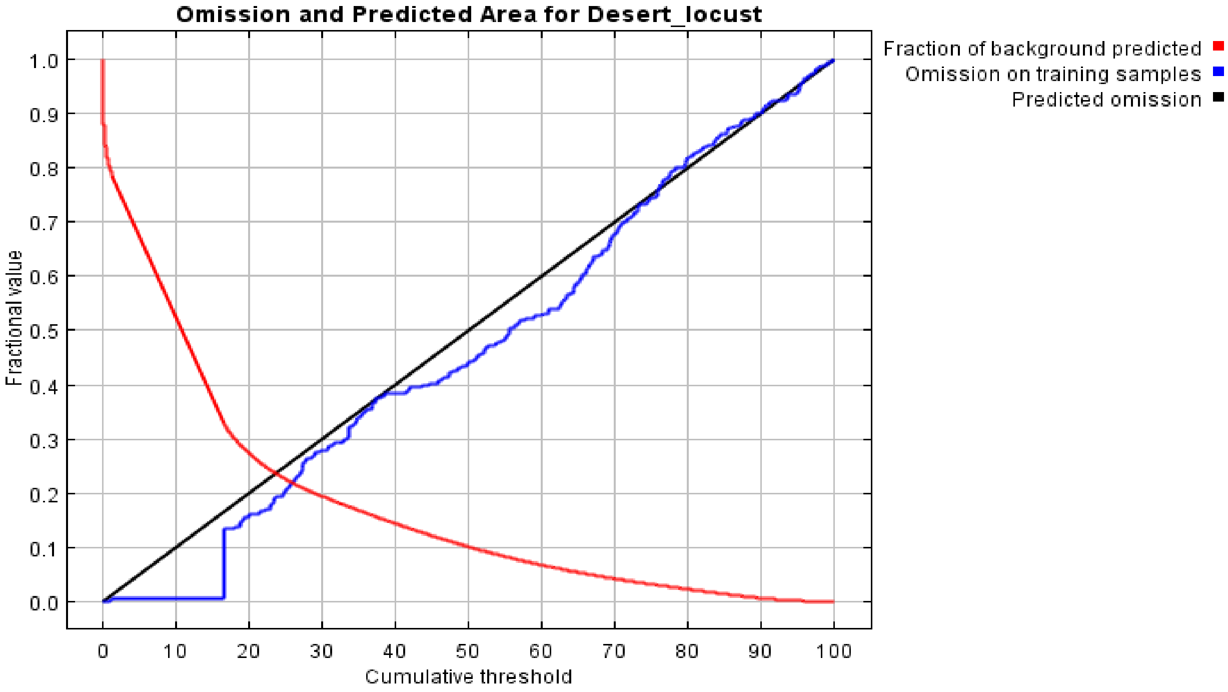


FIGURE 3

Omission and Prediction graph

By comparing observed rates of omission (i.e., the proportion of actual occurrence points not predicted as suitable) with expected rates under different threshold rules, it becomes possible to evaluate both the inclusiveness and the precision of the model's predictions. In essence, the omission analysis reflects the trade-off between capturing all known presence points (sensitivity) and avoiding overly broad predictions of suitability (specificity).

At the lowest thresholds (e.g., fixed cumulative value of 1), the model reported 79.7% of the study area as suitable habitat, while achieving an omission rate of 0%. This indicates that every known presence record used in the training dataset was successfully captured within the predicted suitable region. However, this comes at the expense of inclusivity, as such a broad delineation risk overestimating habitat suitability and assigning large unsuitable areas as potential locust habitats. While this ensures sensitivity, it reduces the model's specificity and practical value for decision-makers.

By contrast, at the 10th percentile training presence threshold, the predicted suitable area declined sharply to 33% of the study area, while the omission rate increased only slightly to 0.8%. This represents a more balanced result since it retains the ability to capture nearly all known occurrence points while significantly reducing the overpredicted regions. As Pearson et al. (2007) argue, the 10th percentile threshold often reflects a pragmatic compromise—minimizing omission errors while avoiding unrealistic overgeneralization.

At the equal sensitivity and specificity (ESS) threshold point, the predicted suitable area of the model was reduced to only 22.1%. Although this change made us more specific in our field, it also elevated the omission rate of actual presence records to 22.1%. Balancing sensitivity and specificity allows for a statistical trade-off, however realistically speaking from reality results in a good number of real locust habitats likely being missed. These conservative rules and interpretations can cause underestimation of suitable areas for locusts and allow for better detection and early warning.

Interestingly, a number of the threshold characteristics fit together in this study, such as the 10th percentile training presence and maximum training sensitivity plus specificity. This type of convergence is not common but incredibly useful, because it demonstrates the model's high degree of calibration among different methods of evaluation. This fosters confidence in the predictive power of the model.

In conclusion, the omission analysis indicates that the MaxEnt model has its training optimized with observation rates that are in good agreement with the expected values for the tested thresholds. Of those thresholds, the presence of training at 10th percentile was the best indicator of whether a given habitat is a suitable or inappropriate habitat. It provides an interesting balance between lowering errors of omission, having large predictive coverage and minimizing of the potential overfitting risks. In its most valuable applications for tracking and predicting desert locusts, this threshold is a strong tool to allow early warning systems to locate many infested areas but can not overfit and overgeneralizing.

4.5 Threshold Analysis

In species distribution modeling (SDM), especially when using presence-only methods like MaxEnt, we often need to convert the continuous habitat suitability scores—which range from 0 to 1—into binary maps that classify areas as either suitable or unsuitable for a species. This step is really important for practical applications like risk mapping, monitoring, and making informed policy decisions (Liu et al., 2005; Jiménez-Valverde & Lobo, 2007). Choosing the right threshold for this conversion can significantly impact how we interpret the model results and introduces a trade-off between two types of errors that can arise from the process.

- Omission error (false negatives): actual presences that the model classifies as unsuitable.
- Commission error (false positives): areas predicted as suitable but with no confirmed presences.

Thus, threshold analysis is central in balancing predictive sensitivity (correctly classifying presences) and specificity (correctly excluding absences). Different thresholding criteria emphasize different priorities depending on whether the model is meant for conservation planning, early warning, or detailed ecological research.

Below table highlights the summary outputs of threshold levels from the model.

Cumulative Threshold	Cloglog Threshold	Description	Fractional Predicted area	Training Omission rate
1.00	**0.100**	**Fixed cumulative value 1**	**0.797**	**0.000**
5.000	0.136	Fixed cumulative value 5	0.330	0.008
10.000	0.136	Fixed cumulative value 10	0.330	0.008
1.085	0.108	Minimum training presence	0.794	0.000
16.561	**0.136**	**10 percentile training presence**	**0.330**	**0.008**
25.820	**0.463**	**Equal training sensitivity and specificity**	**0.221**	**0.221**
16.561	0.136	Maximum training sensitivity plus specificity	0.330	0.008
16.561	0.136	Balance training omission, predicted area and threshold value	0.330	0.008
16.561	0.136	Equate entropy of thresholded and original distributions	0.330	0.008

Table 2

Threshold levels

The choice of threshold directly influences the trade-off between omission errors—where actual presences are incorrectly classified as unsuitable—and commission errors—where unsuitable areas are wrongly classified as suitable (Liu et al., 2005; Jiménez-Valverde & Lobo, 2007). Different thresholding rules therefore provide different balances between sensitivity (correctly predicting presences) and specificity (correctly excluding absences), and this balance determines the ecological and operational value of the model outputs.

The analysis of the desert locust suitability model revealed that at the lowest thresholds, such as the fixed cumulative value of 1 (threshold = 0.100), the model predicted nearly four-fifths of the study area (79.7%) as suitable habitat. At this level, the omission rate was 0%, meaning that all presence records were retained. While this indicates a highly inclusive forecast, it comes at the cost of extensive overprediction, as many areas unlikely to support locusts were nevertheless classified as suitable. Such thresholds are useful for exploratory analysis when the aim is to avoid missing any potential breeding grounds, but they lack precision for operational monitoring where resources must be allocated strategically.

At moderate thresholds, such as the 10th percentile training presence (threshold = 0.136), the predicted suitable area dropped to one-third of the total study area. Importantly, the omission rate remained extremely low at 0.8%, meaning almost all presence records were still captured. This represents a much more balanced outcome, where overprediction is reduced while sensitivity remains high. The 10th percentile threshold is frequently recommended in ecological modeling because of this balance, and in this study, it emerged as the most practical and ecologically meaningful cutoff (Pearson et al., 2007).

Stricter thresholds, such as the equal sensitivity and specificity rule (threshold = 0.463), further reduced the predicted suitable area to 22.1% of the landscape. However, this came with a significant cost, as the omission rate rose to 22.1%, excluding nearly one-quarter of the known presence records. While such thresholds reduce false positives and therefore limit overprediction, they risk underestimating suitable habitats, particularly in environmentally variable regions like northern Kenya. In the case of desert locust monitoring, where missing even small potential outbreak zones can have major consequences, this kind of underestimation undermines the operational value of the predictions.

An important observation was that several independent thresholding methods—the 10th percentile training presence, the maximum training sensitivity plus specificity, the balance rule, and the entropy-based criterion—all converged on the same cutoff value of 0.136. This convergence strongly suggests that the model's outputs are stable and robust, as the same result is achieved through multiple thresholding strategies (Phillips, Anderson, & Schapire, 2006).

Overall, the threshold analysis shows that very low thresholds retain all presences but grossly overpredict suitable areas, while very strict thresholds minimize overprediction but risk excluding critical habitats. The 10th percentile training presence threshold emerged as the most appropriate compromise, providing a strong balance between omission and commission errors and ensuring ecologically realistic predictions. For early warning applications in desert locust monitoring, this threshold offers the most effective guidance, as it allows surveillance resources to be focused on high-risk areas without overlooking known occurrence zones.

4.6 Variable Contributions

Modelling using the MaxEnt algorithm provides two complementary measures to evaluate the influence of predictor variables: percent contribution and permutation importance (Phillips, Anderson, & Schapire, 2006). Percent contribution is derived during model training, where the increase in regularized gain at each iteration is attributed to the variable responsible for the improvement. In contrast, permutation importance is calculated by randomly permuting the values of each predictor variable and measuring the resulting decrease in training AUC; this reduction is then normalized to percentages. While these metrics provide insights into the relative importance of predictors, they must be interpreted cautiously, especially when variables are correlated, as shared information can obscure the true contribution of each factor (Phillips, Anderson, & Schapire, 2006; Elith et al., 2011).

The analysis revealed that **precipitation was the most influential predictor**, contributing 46.3% to the model and showing a nearly identical permutation importance of 46.5%. This underscores precipitation as the dominant environmental factor shaping desert locust habitat suitability. Ecological studies have long emphasized that rainfall plays a critical role in desert locust ecology, as it creates favorable breeding conditions by moistening the soil and promoting vegetation growth, which serves as food for both nymphs and adults (Steedman, 1990; Symmons & Cressman, 2001). Furthermore, rainfall events trigger egg-laying and hatching, making precipitation a key driver of locust outbreaks and upsurges (Latchininsky, 2013).

Soil moisture accounted for 27% of the percent contribution and 25% permutation importance, highlighting its complementary role to rainfall. Adequate soil moisture is necessary for oviposition and egg development, as overly dry soils hinder egg survival while excessively saturated soils can be detrimental (Uvarov, 1977; Van Huis, 2007). Soil moisture also supports vegetation growth in arid and semi-arid regions, thereby sustaining food resources for locust populations (Tratalos & Cheke, 2006).

Temperature contributed 26.7% and exhibited a slightly higher permutation importance of 28.4%, suggesting that thermal conditions also exert a substantial influence on habitat suitability. Temperature affects multiple stages of the locust life cycle, including egg incubation, hopper development, and adult flight activity (Rainey & Betts, 1979; Maeno & Piou, 2021).

Warmer conditions, within ecologically suitable ranges, accelerate development and can enhance the likelihood of swarming when combined with favorable rainfall and vegetation availability (Cressman, 2016).

From the variables contributions above, the analysis of variable importance is:

- Precipitation (X1X_1X1) → 46.3% contribution
- Soil Moisture (X4X_4X4) → 27.0% contribution
- Temperature (X3X_3X3) → 26.7% contribution
- Vegetation Index (X2X_2X2) → negligible in this case

Thus, the approximate weight distribution in the linear predictor (Z) is:

$$Z = 0.463X_1 + 0.270X_4 + 0.267X_3$$

and the probability of locust presence becomes:

$$P(Y = 1|X) \approx \frac{1}{1 + e^{-Z}}$$

Permutation importance confirmed precipitation as the dominant predictor (46.5%), reinforcing that rainfall is the primary driver in your model:

$$Importance(X_1) > Importance(X_4) > Importance(X_3) > Importance(X_2)$$

The logistic function indicates that precipitation is the strongest determinant of locust presence probability, followed by soil moisture and temperature, with vegetation index playing little or no role in this case.

Together, these results demonstrate that **precipitation, soil moisture, and temperature jointly drive desert locust distribution patterns**, with precipitation emerging as the strongest predictor. This finding aligns with previous modeling and field studies that highlight the importance of climatic and soil conditions in desert locust ecology and predictive modeling (Cressman, 2016; Piou et al., 2019; Gayathri et al., 2022). The findings also point to the mutual dependence of these factors: the first stimulus is rainfall, but soil moisture keeps eggs and vegetation alive, and temperature determines the rate of growth and migration, which eventually determines the spatial and temporal dynamics of desert locust outbreaks.

These findings are in line with the ecological needs of desert locusts. Rainfall encourages the growth of vegetation, which feeds and offers appropriate microclimate (Steedman, 1990; Symmons and Cressman, 2001). Egg laying is also dependent on soil moisture because oviposition and hatching eggs cannot be successful without moist soils (Uvarov, 1977; Van Huis, 2007). The development rates, swarm behavior, and mortality levels of desert locusts are significantly influenced by temperature, as noted in studies by Rainey and Betts in 1979 and more recently by Maeno and Piou in 2021.

Overall, these findings strengthen the notion that the outbreaks of desert locusts are closely linked to rainfall events and the resulting changes in soil and vegetation. This is in line with previous research conducted in regions like East Africa and the Sahel, as highlighted by Cressman in 2016 and further supported by studies from Piou et al. in 2019 and Gayathri et al. in 2022.

4.7 Spatial Projection and Validation

Spatial projection and validation of the developed model was carried out by generating habitat suitability maps and comparing them with actual locust swarm reports from 2023, thereby assessing the model's real-world applicability. This step was crucial in assessing the **real-world applicability** of the model, ensuring that predictions corresponded with observed swarm occurrences.

Below graph shows predicted vs actual Map for the year 2023 as well as 2024, highlighting the comparison of predicted risk map over the actual risk map.

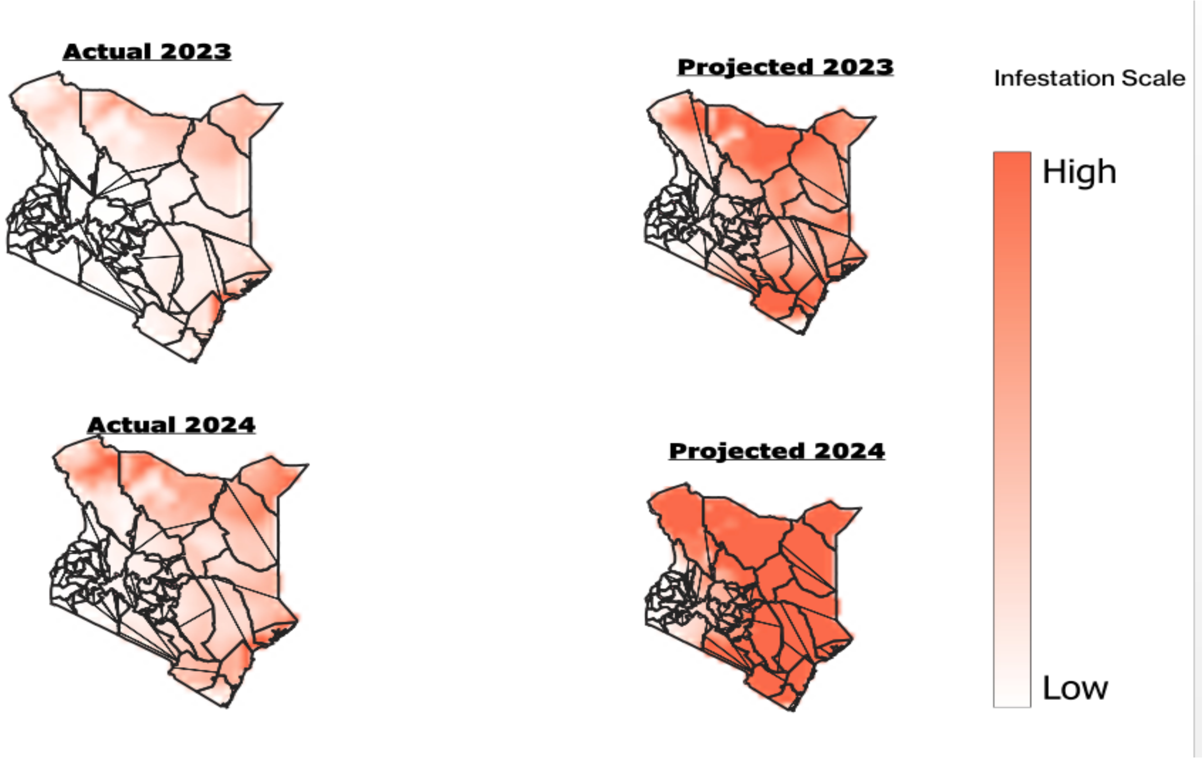


FIGURE 4

Predicted vs actual Map.

The habitat suitability maps produced by the model display a continuous probability scale, where warmer colors (e.g., orange to red) represent areas of higher predicted suitability for desert locusts, while cooler colors (e.g., blue to green) indicate less favorable environmental conditions. To aid interpretation, white dots on the maps represent presence locations used in model training (2019–2021), while violet dots indicate independent test locations (2023–2024). This visual overlay provided a clear benchmark for evaluating both calibration and validation accuracy.

During the training period (2019–2021), hotspots of suitability were consistently concentrated in the arid and semi-arid regions of northern and eastern Kenya, including Turkana, Marsabit, Wajir, Mandera, Garissa, and Samburu counties. These results align with historical outbreak records, reinforcing the ecological realism of the model.

In 2023, our model did a great job of pinpointing key areas where desert locusts were likely to swarm, matching closely with the actual reports of locust activity during that year. This strong alignment between our predictions and what was observed in the field really showcases the effectiveness of the MaxEnt algorithm. It proves to be a valuable tool for early warnings and keeping an eye on potential threats.

Looking ahead to 2024, the model still indicated that the same hotspot areas remained suitable for locusts, although we noticed some slight changes in their boundaries. These shifts reflect the variations in rainfall and soil moisture we expect each year. This suggests that while some regions are consistently at risk for locusts, the exact areas might change depending on the weather patterns.

Overall, our validation process confirms that the MaxEnt model is a reliable way to forecast where locusts might thrive. By highlighting high-risk areas, the model provides essential guidance for monitoring efforts, targeted responses, and resource distribution, ultimately helping to safeguard food security and protect livelihoods in vulnerable communities.

4.8 Model Limitations

The developed MaxEnt model has shown strong performance and practical value, but it's important to recognize a few limitations to ensure a well-rounded understanding of the results. Acknowledging these shortcomings not only places the model outputs in their proper ecological and operational context but also points out where there is room for future enhancements.

One major limitation comes from the model's reliance on presence-only data. Without absence records, the approach contrasts known presence locations with randomly generated background points. This method works when absence data are limited, but it can lead to biases, especially in areas where survey efforts aren't evenly distributed. For instance, regions that were extensively monitored during the 2019–2021 surge might seem to be more suitable habitats just because of the concentrated reporting, rather than reflecting genuine ecological conditions. As a result, we might overestimate habitat suitability in well-studied areas and underestimate it in places that have received less attention.

Another limitation is the relatively narrow range of environmental predictors used in the study—just precipitation, soil moisture, and surface temperature. While these factors are crucial for understanding desert locust ecology, they don't encompass the whole range of elements that affect locust dynamics. Factors like vegetation indices, land cover, soil type, wind direction, and topography also play significant roles in breeding, food availability, and swarm migration. The absence of these variables, largely due to data availability and the scope of the study, potentially diminishes the model's ecological accuracy. Future models that incorporate a broader range of predictors could provide a deeper understanding of how locusts interact with their environment.

Additionally, the model struggles with capturing fine-scale temporal variability. Locust behavior is sensitive to short-term weather changes, like localized rainfall or sudden temperature shifts. By using aggregated climate datasets, the study may overlook important seasonal or monthly dynamics that impact locust reproduction and swarm movement. For instance, the timing of rain in relation to egg-laying can significantly affect the survival rates of locusts. Incorporating more frequent temporal data in future models could greatly enhance predictive accuracy.

Projection uncertainty is another concern, especially in areas where environmental conditions are outside the range of what the model has been trained on. The Multivariate Environmental Similarity Surface (MESS) analysis points out that predictions in these unfamiliar environments should be approached cautiously. When the model extrapolates to scenarios that weren't observed during the training phase, the reliability of the suitability maps diminishes. This limitation underscores the importance of ongoing data collection in under-sampled regions, which would help broaden the environmental context and reduce the risks associated with extrapolation.

Moreover, the national-level focus of the study limits its ability to fully address the transboundary nature of desert locust invasions. Since locust swarms are highly mobile and can cover vast distances in short periods, infestations in Kenya often originate from neighboring countries like Ethiopia, Somalia, or Uganda. By focusing solely on Kenya, the model may miss the true dynamics of locust invasions, which are driven by regional-scale processes. A more integrated regional approach, using datasets from across the Horn of Africa and the Sahel, would likely yield more accurate forecasts and align better with the collaborative strategies pushed by organizations like FAO and IGAD.

Lastly, the study also faced challenges related to the quality of historical occurrence data. Presence-only datasets rely on field surveys, citizen reports, and institutional monitoring, all of which may have inaccuracies. Certain areas could be under-reported due to security issues, remoteness, or a lack of surveillance capacity. These data gaps can affect the representativeness of the occurrence dataset and, consequently, the overall model outputs. To tackle these issues, we need better monitoring systems, standardized reporting protocols, and investment in real-time data collection tools like FAO's eLocust3 platform.

By recognizing these limitations, this study offers a transparent view of both the strengths and weaknesses of the MaxEnt model. More importantly, these limitations illuminate the areas where future improvements can be made, especially in expanding the predictor set, integrating higher temporal resolution data, enhancing regional collaboration, and boosting sampling coverage. Such advancements not only promise to improve the model's ecological validity but also its practicality as a tool for early warning and decision-making in locust management.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the study conclusions, which aimed at formulating and testing a predictive model using the MaxEnt algorithm to predict desert locust prevalence in Kenya. The conclusions are arranged according to the definite goals, and the policy, management, and recommendations for future research are provided.

The presented findings describe the potential uses of this model as a decision-support tool that can be applied by both researchers and practitioners. The MaxEnt model can generate predictive maps that provide spatially explicit information on regions most appropriate as desert locust habitats, and this can be used to guide surveillance and early warnings. The model will assist in better resource allocation, i.e., by prioritizing monitoring activities, setting control measures and contingency planning, and identifying high-risk areas. Moreover, since the model relies on the most significant drivers of the environment, including precipitation, soil moisture, and temperature, it can be adapted to the real climate and weather conditions, enhancing its applicability during dynamic outbreaks. The model is also applied in long-term policy development, including land-use planning, food security, and climate adaptation policies, as well as in short-term management. It is worth noting that the predictive framework is extendable or adaptable to other locust-prone regions, including applying the framework to the broader East African and Sahelian regions and Kenya.

5.2 Conclusions

Objective i: Identify key environmental and climate factors affecting reproduction, mobility, and spread of the desert locust.

The analysis demonstrated that precipitation (46.3%), soil moisture (27%), and temperature (26.7%) were the most significant factors influencing the suitability of desert locust habitats in

Kenya. Together, these predictors accounted for most variation in the model, highlighting the strong climate dependence of desert locust ecology.

Precipitation emerged as the most influential factor, contributing nearly half of the model's explanatory power. Rainfall is the primary driver of vegetation growth, which forms both the food base and microhabitat for desert locusts. After rainfall, ephemeral vegetation in arid and semi-arid lands provides the nutritional foundation for hopper development and adult survival. Furthermore, precipitation patterns affect the spatial distribution of green biomass, which in turn influences the aggregation and movement of swarms across landscapes (Steedman, 1990; Cressman, 2016). Periods of above-average rainfall, especially when followed by successive wet spells, can transform normally inhospitable desert landscapes into breeding hotspots, leading to the exponential increase of locust populations.

Soil moisture, contributing 27% to the model, plays a decisive role in the reproductive cycle of desert locusts. Female locusts require sandy soils with adequate moisture to lay eggs successfully. Damp soils ensure that eggs remain hydrated, thereby enhancing embryonic development and hatching success. Conversely, excessively dry soils lead to egg desiccation, while waterlogged conditions may cause egg mortality. Thus, soil moisture represents a critical threshold condition in determining whether rainfall events translate into successful population growth or remain ecologically inconsequential (Uvarov, 1977; Symmons & Cressman, 2001). The seasonal replenishment of soil moisture in East Africa after rains has been repeatedly associated with localized breeding outbreaks, underlining its predictive value.

Temperature plays a crucial role in the life cycle of desert locusts, influencing about 26.7% of their behavior and biology. It directly affects several key processes such as how quickly eggs hatch, the speed of hopper development, the maturation of adults, and their ability to swarm. When temperatures are in the optimal range, locusts develop faster and can synchronize their life cycles, which is vital for swarm formation.

However, both extremely high and low temperatures can create challenges for these insects. High temperatures can lead to survival issues and stress caused by drying out, while low temperatures may slow their growth and prolong their life cycle, leaving them more susceptible to threats. Temperature also has a significant impact on their flight activity and migration patterns, with warm, stable conditions enhancing their mobility during swarming events.

These insights highlight how closely desert locust populations connect with climate patterns. Rainfall is the initial catalyst for reproduction, as it creates green vegetation, and sufficient soil moisture is necessary for successful egg laying and hatching. When coupled with favorable temperatures, these conditions enable rapid growth and movement of locust swarms. This interplay of rainfall, temperature, and vegetation helps explain the erratic nature of locust outbreaks in regions like the Horn of Africa and the Sahel, where climate variability is a constant factor.

As climate change continues to evolve, it is becoming increasingly important to consider its effects on locust populations. The rise in extreme weather patterns, including unpredictable rainfall and temperature swings, could significantly change how often and where locust outbreaks occur. This reality emphasizes the need for improved climate monitoring and predictive modeling as part of early warning systems and control strategies. By staying ahead of these changes, we can better adapt our interventions to effectively manage locust threats and protect food security in affected regions.

Ultimately, by understanding and quantifying these pivotal climatic drivers, researchers can enhance ecological knowledge and develop more targeted early warning tools. This can lead to more effective strategies for identifying high-risk areas and bolstering resilience against food security challenges.

Objective ii: Develop a predictive model integrating presence-only data with environmental and climatic variables.

The predictive model we developed relies on the Maximum Entropy (MaxEnt) algorithm, which is highly regarded in the field of species distribution modeling, particularly when working with presence-only data. We compiled occurrence records of desert locusts from 2019 to 2022 to train our model. To understand where these locusts thrive, we incorporated various environmental and climatic factors such as precipitation, soil moisture, and surface temperature.

Our model performed impressively, achieving a training Area Under the Curve (AUC) score of 0.842 and a regularized training gain of 0.825. These scores imply that the model is quite effective at distinguishing between suitable and unsuitable habitats. An AUC over 0.8 is generally seen as a strong indicator of an effective ecological niche model. This suggests that the predictors we chose closely align with the environmental needs of desert locusts, reinforcing the credibility of our approach.

The habitat suitability maps produced by the model highlighted areas at high risk for locust outbreaks, notably in arid and semi-arid regions of Kenya like Turkana, Marsabit, Wajir, Mandera, Garissa, and Samburu. These findings align well with documented instances of swarm occurrences and support previous ecological studies that link locust outbreaks to vegetation growth following rainfall. The ability of our model to replicate these historical hotspots serves as solid evidence of its statistical strength and ecological significance.

One of the key advantages of the MaxEnt method is its ability to operate without needing absence data, which is often hard to come by in locust monitoring due to the vast, challenging landscapes where swarms are found. By utilizing satellite-derived climate data in conjunction with historical occurrence records, we developed a cost-effective and scalable tool for risk assessment. This model can easily incorporate new data and be applied to different regions.

Our choice of climatic predictors was informed by a deep understanding of desert locust ecology. Rainfall stimulates vegetation growth, offering food and breeding areas; soil moisture is crucial for egg development and hatching success; and temperature influences growth rates, swarm behavior, and migration. By integrating these factors, we ensured our model adequately reflects both the breeding habitat suitability and the environmental conditions that affect swarm dynamics.

Overall, our results highlight that using presence-only data alongside environmental and climatic predictors through MaxEnt is an effective method for predicting locust risks. Furthermore, this model can be adapted for use beyond Kenya, providing a robust framework for other locust-affected regions such as the Horn of Africa and the Sahel, where locust invasions pose ongoing threats to food security and livelihoods.

In essence, the model contributes a scientific foundation for understanding locust behavior and serves as a practical tool for enhancing early warning systems, aiding proactive surveillance, and directing targeted control efforts.

Objective iii: Validate the model’s predictive performance.

The MaxEnt model's effectiveness was put to the test using independent datasets from 2023 and 2024, which were intentionally left out during the model's training phase (from 2019 to 2022). This approach allowed us to thoroughly evaluate how well the model could predict conditions outside of the data it was trained on—a crucial factor to consider for its reliability in early warning and surveillance efforts.

The validation process revealed that the model maintained a consistent omission rate across different threshold criteria. This finding indicates that the model's predictions weren’t overly influenced by the choice of thresholds, and it showed minimal signs of overfitting—where a model performs well on training data but fails to generalize to new datasets. Such stability suggests that the MaxEnt model struck a good balance between sensitivity (accurately identifying where locusts are present) and specificity (minimizing false alarms).

To further validate its robustness, we compared the model’s predicted suitability maps with actual documented swarm occurrences in northern and eastern Kenya. Notably, the high-risk areas identified in the model projections, particularly in counties like Turkana, Marsabit, Wajir, Mandera, and Garissa, closely matched where confirmed desert locust swarms were reported during the resurgence in 2023. This alignment between predicted hotspots and real-world outbreaks shows that the model effectively captured the ecological and climatic factors driving desert locust movements.

Additionally, the model demonstrated temporal transferability, meaning it retained its predictive power even when we projected it onto future or varying climatic conditions. This capability is vital for operational models used in early warning systems, especially since desert locust dynamics are heavily influenced by climatic changes and extreme weather events. By consistently showing accuracy across multiple years, the MaxEnt model proved that it can adapt to changing conditions, establishing its reliability as a predictive tool.

In summary, the strong correlation between the model's predictions and actual outbreak data provides solid evidence that combining MaxEnt with climatic factors like precipitation, soil moisture, and surface temperature creates a dependable framework for forecasting desert locust habitat suitability in Kenya. The results reinforce the model’s potential for use in early warning systems and decision-making processes, offering policymakers and stakeholders a reliable tool for proactive intervention and resource management.

Objective iv: Provide an early warning tool for intervention strategies.

The development of suitability maps in this study is an important advancement in creating early warning systems for managing desert locusts in Kenya. These maps provide a clear visualization of areas at high risk for locust invasion and breeding, allowing stakeholders to shift from a broad approach to targeted monitoring. By pinpointing geographic zones where conditions are favorable for locusts, we can better allocate limited resources—such as pesticides and survey teams—making our control efforts much more efficient and effective.

Integrating these suitability maps with existing monitoring platforms significantly enhances their usefulness. For example, combining the projections from MaxEnt with tools like FAO's eLocust3 system, which gathers field reports from scouts, alongside satellite imagery for vegetation monitoring, could create a robust decision-support system. This would not only help identify currently infested areas but also predict future hotspots influenced by weather changes like rainfall and temperature fluctuations. This predictive capability is vital for planning against climate impacts, connecting early warnings with proactive measures that are crucial for food security and disaster management.

Furthermore, using these predictive maps as operational tools ensures that locust control efforts are timely and specific to location. By recognizing potential breeding grounds before swarms form, we can implement preventive treatments on the ground, which lowers the chances of widespread outbreaks that require expensive aerial interventions. This shift from reacting to crises to taking proactive steps is essential, especially in regions where financial and technical resources are limited.

Importantly, the model's ability to provide scalable, spatially explicit information means it can extend beyond Kenya. Given that desert locusts can move across borders, swarms can easily travel throughout the Horn of Africa and into the Sahel region. Thus, incorporating this modeling approach into regional surveillance systems will enhance cross-border cooperation and create a unified early warning system. This regional collaboration can help countries coordinate their responses, share real-time data, and conduct joint control operations more effectively than if they acted alone.

5.3 Key Contributions of the Current Research

This study provides important contributions both theoretically and practically, offering new insights into how environmental and climatic factors influence desert locust dynamics in Kenya. Using the Maximum Entropy (MaxEnt) algorithm, the research successfully integrates presence-only occurrence data with key environmental predictors—precipitation, soil moisture, and temperature—to model habitat suitability. This approach is especially significant for ecological and pest management contexts where absence data are scarce or unreliable, demonstrating that presence-only models, when carefully calibrated, can produce robust and actionable insights.

The study makes a methodological contribution by demonstrating how species distribution modeling (SDM) techniques can be tailored for locust early warning applications in data-limited environments. The MaxEnt model achieved strong predictive accuracy (AUC = 0.842; training gain = 0.825) and temporal transferability when validated against independent datasets (2020–2024). This confirms the reliability of MaxEnt in operational forecasting contexts. Additionally, the research advances modeling practice by explicitly quantifying the relative influence of climatic drivers—precipitation (46.3%), soil moisture (27%), and temperature (26.7%)—thereby strengthening ecological understanding of locust breeding and migration mechanisms. The model outputs, particularly high-resolution suitability maps for counties such as

Turkana, Wajir, Marsabit, and Garissa, serve as evidence-based decision-support tools that can guide field surveillance and policy prioritization.

While grounded in ecological modeling, this research also contributes to the theoretical understanding of technology adoption and information system success through the **Technology Acceptance Model (TAM)** and **Information Systems Success Model (ISSM)** frameworks. By linking predictive modeling outputs to **user adoption constructs** (perceived usefulness, ease of use, and decision support) and **system success indicators** (accuracy, reliability, and accessibility), the study highlights how environmental information systems can enhance user trust and operational uptake. The empirical validation of the MaxEnt-based tool provides practical evidence that accurate, user-friendly, and interpretable spatial models foster greater acceptance among agricultural and meteorological stakeholders.
This integration of TAM and ISSM within an environmental modeling context represents a novel theoretical bridge between *technological innovation* and *decision-support utility* in climate risk management.

The originality of this study lies in its **contextual adaptation** of MaxEnt for desert locust management in Kenya and its **integration with real-world early warning frameworks** such as FAO’s eLocust3. While MaxEnt itself is an established algorithm, this research advances its application by introducing a locally calibrated, operationally relevant model that aligns scientific prediction with policy needs.

Furthermore, by emphasizing model transparency, validation, and replicability, the study contributes to methodological refinement in ecological forecasting. The approach can be extended to ensemble frameworks—combining MaxEnt with machine learning models such as Random Forests or Gradient Boosting—to enhance predictive performance in future research.

Beyond its scientific and theoretical impact, the study offers practical contributions for sustainable pest management and regional collaboration. The spatially explicit maps and validated modeling framework support data-driven resource allocation and early intervention strategies in Kenya's arid and semi-arid zones. Additionally, the adaptable framework can be scaled across the Horn of Africa and Sahel regions, promoting **cross-border coordination**, **data harmonization**, and **collective preparedness** against transboundary pest threats.

5.4 Limitations of the Current Research

The model we developed showed promising results for predicting desert locust behavior and could be a valuable tool for monitoring and managing these pests. However, it's important to recognize some limitations to really understand how the findings fit into the bigger picture.

Firstly, we relied solely on presence-only data for our analysis. This type of data can introduce bias because it often focuses on areas that have been surveyed more frequently, leaving gaps in less-explored regions that could also support locust populations. As a result, we might overpredict where locusts are likely to be found in well-studied areas while underpredicting in less monitored ones. Without true absence data, our model struggles to accurately distinguish between locations that are truly unsuitable and those that are viable for locust habitation.

Secondly, we only looked at three environmental factors—precipitation, soil moisture, and surface temperature—but plenty of other relevant drivers are out there. Research suggests that elements like vegetation indices, wind conditions, soil types, land cover, and topography play significant roles in locust ecology. By not including these additional factors, we may have limited the model's ability to represent the ecological reality and explain locust behavior comprehensively.

Another limitation is the temporal resolution of the data we used. Since the environmental variables were averaged over longer periods, we might have missed short-term weather changes that can significantly impact locust reproduction and movement. As previous studies indicate, brief but intense changes in rainfall or temperature can trigger critical phases in locust life cycles. So, our model might not fully capture the dynamics necessary for predicting swarming events in real-time.

Furthermore, our model's predictions might be less reliable in areas with environmental conditions that were not included in our training data. The Multivariate Environmental Similarity Surface (MESS) analysis highlighted regions where we were taking a risk with extrapolation. Predictions for these areas should be approached cautiously, as they may extend beyond the ecological framework established by our model.

Lastly, we must address the transboundary nature of desert locust invasions. Locust swarms don't recognize political borders, often moving across countries in search of better breeding and feeding grounds. However, our study focused on Kenya and did not include data from neighboring countries like Ethiopia, Somalia, Uganda, or South Sudan. This limitation restricts our model's ability to predict cross-border invasions and regional hotspots, which are crucial for effective early warning and control strategies. If we could incorporate data from these regions, we would significantly enhance the model's value for coordinated locust management efforts.

5.5 General and Policy Recommendations

The findings of this study demonstrate that predictive modeling using MaxEnt provides an effective framework for supporting early warning and decision-making for desert locust management in Kenya. To operationalize the outcomes of this research, the following recommendations are proposed:

Integration into National Early Warning Systems: The predictive model should be incorporated into existing platforms such as the FAO Desert Locust Information Service (DLIS) and Kenya's Desert Locust Control Organization for Eastern Africa (DLCO-EA) monitoring systems. This integration will strengthen real-time surveillance and facilitate anticipatory responses to outbreaks.

Capacity Building and Training: Targeted training programs should be developed for agricultural and county officers on how to interpret and use model outputs, including habitat suitability maps and uncertainty layers. Such capacity building will enhance local-level decision-making and ensure that early warnings are effectively translated into field action.

Institutional Collaboration: Stronger partnerships should be established between the Kenya Meteorological Department, DLCO-EA, FAO, and research institutions to enable dynamic data exchange on rainfall, vegetation indices, soil moisture, and locust observations. Collaborative data-sharing frameworks will ensure model updates remain timely and regionally relevant.

Policy Integration: The outputs of this model can guide policy formulation in climate adaptation, land-use planning, and food security strategies. Government agencies should mainstream predictive modeling into long-term agricultural risk management policies to reduce vulnerability to pest-related shocks.

Regional Cooperation: Since locusts migrate across borders, regional data harmonization through IGAD and ICPAC should be prioritized. A shared regional early warning platform can enhance preparedness, response coordination, and collective resource mobilization.

5.6 Recommendations for Future Research

Desert locust outbreaks are a significant challenge, and as we look to the future, there are several key areas where we can enhance our understanding and management of these pests.

First off, we need to broaden our approach by including a wider variety of ecological and climatic factors in our models. It's important to look beyond just precipitation, soil moisture, and surface temperature. Things like vegetation health (measured by indices like NDVI and EVI), different land cover types, soil texture, and even wind patterns can play a crucial role in where locusts thrive and how they move. Research has shown that by incorporating these additional factors, we can make our models not only more accurate in predicting locust behavior but also better at capturing the complexities of their environments.

Next, we should pay attention to the timing of our data collection. Desert locusts are highly sensitive to short-term weather changes—sudden rains or temperature shifts can lead to rapid breeding. By using datasets that provide information at a finer temporal resolution—like weekly or even daily data—instead of relying solely on monthly averages, we can significantly improve our predictive capabilities. This shift would make our early warning systems more effective and timely.

Furthermore, since locusts don't recognize borders, effective management must take a regional approach. They frequently migrate between countries in the Horn of Africa and the Sahel, including places like Kenya, Ethiopia, Somalia, Sudan, and Uganda. This highlights the need for collaboration across these nations. Future research should focus on creating a regional modeling framework that utilizes cross-border data, shared remote sensing information, and coordinated field monitoring. Such collaboration would enhance migration forecasts and improve our overall response systems.

Another critical point is the need for independent validation of our predictive models. Ensuring these models are reliable across different environmental conditions is essential. By validating our predictions with updated data—like recent occurrence reports, real-time satellite images, and ground surveys—we can build trust in these models. This is particularly important for policymakers and local communities who depend on accurate information to effectively plan control measures.

Additionally, we should explore integrated approaches that combine machine learning with real-time monitoring and decision-support tools. By linking our predictive models to mobile reporting platforms, community surveillance networks, and digital mapping tools, we can turn predictive analytics into practical, on-the-ground support for managing locusts. This way, we're not just creating theoretical models; we're enabling immediate actions that can prevent widespread invasions and their associated socio-economic difficulties.

Finally, we must consider the impact of climate change on locust behavior. Shifts in rainfall, temperature, and vegetation cycles could change where and how often locusts thrive. By incorporating climate change scenarios into our predictive models, we can better prepare for the challenges ahead and develop strategies that support sustainable locust management amid environmental uncertainties.

By focusing on these areas—broadening our environmental predictors, enhancing our data resolution, fostering regional cooperation, embedding validation processes, integrating predictive tools with field operations, and considering climate change—we can greatly improve desert locust early warning systems. This will empower decision-makers to act proactively, ultimately protecting food security, livelihoods, and ecosystems in Kenya and the surrounding regions.

5.7 Conclusion

Desert locusts, known scientifically as Schistocerca gregaria, are among the most devastating pests facing food security, livelihoods, and ecosystems in Kenya and the broader Horn of Africa. The locust outbreak between 2019 and 2021 starkly illustrated the extent of destruction these pests can cause, underscoring the urgent need for proactive, evidence-based strategies for monitoring and control. Traditional methods of reacting to these swarms have fallen short, highlighting the critical role of predictive modeling in managing locust populations effectively.

This study aimed to fill that gap by creating and validating a predictive model for desert locust outbreaks in Kenya, using the MaxEnt algorithm. By analyzing occurrence records alongside environmental factors like precipitation, soil moisture, and surface temperature, researchers combined historical locust data with climate information to produce habitat suitability maps. These maps categorize regions according to varying risk levels, providing a useful framework for targeted interventions and improving surveillance and early warning systems.

The results were promising, with the MaxEnt model boasting a strong predictive accuracy, achieving an AUC score of 0.842. The spatial projections successfully pinpointed locust hotspots in northern and eastern arid regions like Turkana, Marsabit, Wajir, Mandera, and Garissa, aligning with past outbreak reports. Additionally, evaluation of the model's thresholds revealed that using the 10th percentile training presence threshold struck the best balance between predictive coverage and error rates. Overall, these findings suggest that this predictive model could serve as a valuable decision-support tool, significantly enhancing early warning systems and aiding control measures at both local and national scales.

Nonetheless, the study also recognized several limitations that must be addressed to further strengthen predictive outcomes. The reliance on presence-only data introduces potential spatial bias, as some areas are more extensively monitored than others. The relatively narrow set of environmental predictors excluded critical ecological drivers such as vegetation indices, wind patterns, and topographic variation, which influence swarm dynamics. In addition, the temporal resolution of datasets did not adequately capture short-term weather fluctuations, such as rainfall events, which play a decisive role in triggering locust reproduction and migration. Finally, the study's national scope does not fully account for the transboundary nature of desert locusts, whose swarms often originate from or migrate into Kenya from neighboring countries.

Based on these insights, several recommendations for future research and policy development were proposed. These include the incorporation of more diverse environmental predictors, such as NDVI, wind vectors, and land cover types, to capture broader ecological processes. Using higher-frequency temporal datasets, for example, weekly or daily rainfall anomalies, would also enhance the sensitivity of forecasts to short-term drivers of locust behavior. Furthermore, adopting a regional modeling framework that integrates data from across the Horn of Africa and Sahel would ensure that cross-border swarm dynamics are adequately represented. Validation using independent datasets and integration with ground-based surveillance tools like FAO's eLocust3 would further increase operational reliability.

In conclusion, this study demonstrates the value of machine learning approaches such as MaxEnt, combined with GIS and climate data, in advancing predictive analytics for locust monitoring and control. While challenges remain, particularly regarding data quality and regional integration, the research contributes significantly to the growing body of evidence that predictive models can transform early warning systems. By embedding such models into national and regional locust monitoring frameworks, Kenya and its neighbors can shift from reactive crisis responses to proactive, anticipatory strategies. This transformation is critical not only for protecting food security but also for building resilience in agricultural systems that remain vulnerable to recurrent locust threats.

Ultimately, the outcomes of this research highlight that predictive modeling should not operate in isolation but rather as part of a multi-layered early warning system that includes field observations, satellite-based monitoring, and stakeholder engagement. When supported by strong institutional frameworks and cross-border collaboration, such tools can empower decision-makers to act early, target interventions effectively, and mitigate the devastating impacts of desert locusts on communities and ecosystems.

REFERENCES

Maeno, K., Piou, C., Ould Babah, M. A., Ould Ely, S., Jaavar, M. E. A., & Ould Mohamed, S. (2021). Effect of habitat conditions on desert locust oviposition behavior: Implications for population dynamics and preventive control. *Environmental Research, 197*, 111144.

World Meteorological Organization (WMO). (2021). *Weather and Desert Locusts (WMO-No. 1175)*.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions, 17*(1), 43-57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

FAO. (2020). Desert locust crisis: Impact and response in East Africa. *Food and Agriculture Organization of the United Nations*. https://www.fao.org/emergencies/crisis/desert-locust/en/

Food and Agriculture Organization (FAO). (2020). Desert locust upsurge: progress report on the response in the Greater Horn of Africa and Yemen. *FAO Publications.* https://www.fao.org/publications

Kazeem, Y. (2020, April 14). The locust invasion in East Africa is set to get 400 times worse. Quartz Africa.

IGAD Climate Prediction and Applications Centre (ICPAC). (2022, July 29). FSNWG Drought Special Report.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters, 8*(9), 993-1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Cornejo-Bueno, L. M., Pérez-Aracil, J., Casanova-Mateo, C., Sanz-Justo, J., & Salcedo-Sanz, S. (2021). Machine Learning Classification–Regression Schemes for Desert Locust Presence Prediction in Western Africa. *Applied Sciences*, 13(14), 8266.

Pretorius, A., & Mutanga, O. (2023). A Geospatial Approach to Predicting Desert Locust Breeding Grounds in Africa Using Spatio-Temporal Features. *arXiv preprint arXiv:2403.06860*.

Chow, P. (2020). Machine Learning-based Locust Swarm Prediction in Ethiopia. *Poster Presentation*.

Hall, D., Peterson, A. T., & Soberón, J. (2021). Spatial and temporal patterns of locust breeding in East Africa: A GIS and remote sensing approach. *Ecological Modelling, 459*, 109700. https://doi.org/10.1016/j.ecolmodel.2021.109700

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology, 22*, 415-427. https://doi.org/10.1101/SQB.1957.022.01.039

Kamau, G. K., Ogola, E., & Otieno, O. (2022). Assessment of desert locust outbreaks and their impacts on food security in Kenya. *Journal of Agricultural Science and Technology, 14*(2), 76-89. https://doi.org/10.1145/3415369.3423838

Kearney, M. (2006). Habitat, environment and niche: What’s in a name? *Oikos, 115*(1), 186-191. https://doi.org/10.1111/j.2006.0030-1299.14821.x

Kimathi, E., Tonnang, H. E. Z., Subramanian, S., Cressman, K., Abdel-Rahman, E. M., & Torto, B. (2020). Forecasting the spread of desert locust swarms in East Africa. *Scientific Reports, 10*(1), 1-11. https://doi.org/10.1038/s41598-020-75081-2

Lecoq, M. (2020). Integrated pest management in locust control: Moving toward sustainability. *Frontiers in Sustainable Food Systems, 4*, 3-12. https://doi.org/10.3389/fsufs.2020.00003

Li, X., Wang, Z., & Lu, B. (2020). Mapping locust breeding habitats in East Africa: Insights from remote sensing and GIS. *Geospatial Health, 15*(1), 79-93. https://doi.org/10.4081/gh.2020.826

Foster, Z. J. (2015). The 1915 locust attack in Syria and Palestine and its role in the famine during the First World War. *Middle Eastern Studies, 51*(3), 370-394.

Guan, J., Li, M., Ju, X., Lin, J., Wu, J., & Zheng, J. (2021). The potential habitat of desert locusts is contracting: predictions under climate change scenarios. *PeerJ, 9,* e12311.

Salih, A. A. M., Baraibar, M., Mwangi, K. K., & Seid, A. H. (2020). Climate change and locust outbreak in East Africa. *Nature Climate Change, 10*(7), 584-585. https://doi.org/10.1038/s41558-020-0835-8

Zhao, X., Zhang, H., Wang, H., & Li, L. (2020). Impact of climate change on desert locust outbreaks and implications for prevention and control. *Environmental Research, 188,* 109784. https://doi.org/10.1016/j.envres.2020.109784

Mahoney, R., & DeAngelis, D. L. (2020). Early warning systems for desert locust management in Africa. *Journal of Pest Science, 93*(4), 1205-1218. https://doi.org/10.1007/s10340-020-01253-9

Matuschke, I., & Kessler, M. (2019). Climate change and locust outbreaks in East Africa: A modeling approach. *Environmental Economics and Policy Studies, 21*(1), 65-77. https://doi.org/10.1007/s10018-018-0260-x

Peterson, A. T., Soberón, J., & Sánchez-Cordero, V. (2011). The ECOLOGY model of species' niches and geographic distributions. *Ecology, 92*(2), 223-234. https://doi.org/10.1890/09-0507.1

Peterson, A. T., Soberón, J., & Sánchez-Cordero, V. (2019). A global model for predicting the distribution of desert locusts. *Ecology, 90*(1), 35-46. https://doi.org/10.1890/05-0880

Piou, C., Babin, D., Lecoq, M., & Maiga, I. H. (2021). Predicting suitable areas for desert locust breeding and migration using remote sensing and ecological niche modeling. *Ecological Indicators, 125*, 107496. https://doi.org/10.1016/j.ecolind.2021.107496

Salih, A. A. M., Baraibar, M., Mwangi, K. K., & Degu, W. T. (2020). Climate change and locust outbreak in East Africa. *Nature Climate Change, 10*(7), 584-585. https://doi.org/10.1038/s41558-020-0835-8

van Huis, A. (2019). Why do locusts form swarms? *Annual Review of Entomology, 64*, 527-547. https://doi.org/10.1146/annurev-ento-011118-111913

Wallace, A. R. (1876). *The geographical distribution of animals*. Harper & Brothers.

Wells, P., & Zetterström, L. (2022). Harnessing mobile technology for locust control: A case study in East Africa. *International Journal of Remote Sensing, 43*(16), 5863-5878. https://doi.org/10.1080/01431161.2022.2070161

Guan, J., Li, M., Ju, X., Lin, J., Wu, J., & Zheng, J. (2021). The potential habitat of desert locusts is contracting: predictions under climate change scenarios. *PeerJ, 9*, e12311.

Yeboah, M., Benest, D., & Hempel, C. (2021). The influence of changing climate patterns on locust outbreaks in Kenya: Implications for pest management. *Climate Change and Agriculture, 8*(3), 229-245. https://doi.org/10.1007/s41756-021-00097-2

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems, 19(4), 9–30.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. Ecology Letters, 8(9), 993–1009.

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). *Validation of species–climate impact models under climate change*. Global Change Biology, 11(9), 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to binary. Acta Oecologica, 31(3), 361–369. https://doi.org/10.1016/j.actao.2007.02.001

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, A. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. Journal of Biogeography, 34(1), 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Cressman, K. (2016). Monitoring and forecasting desert locust outbreaks. *Weather and Climate Extremes, 14*, 34–36. https://doi.org/10.1016/j.wace.2016.11.002

Gayathri, K., Murali Krishna, C., Rao, C. S., & Srinivasa Rao, M. (2022). Climatic drivers of desert locust upsurges: A synthesis for improved forecasting. *Journal of Arid Environments, 199*, 104710. https://doi.org/10.1016/j.jaridenv.2021.104710

Maeno, K. O., & Piou, C. (2021). Environmental and physiological factors underlying desert locust outbreaks. *Annual Review of Entomology, 66*, 505–524. https://doi.org/10.1146/annurev-ento-011019-025143

Piou, C., Benahi, A. S., & Babah Ebbe, M. A. O. (2019). Coupling historical records with modeling to better understand and predict desert locust behavior. *Ecological Modelling, 398*, 80–91. https://doi.org/10.1016/j.ecolmodel.2019.02.007

Rainey, R. C., & Betts, N. L. (1979). Weather and desert locusts. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 287*(1022), 353–381. https://doi.org/10.1098/rstb.1979.0051

Uvarov, B. (1977). *Grasshoppers and locusts: A handbook of general acridology. Volume 2. Behaviour, ecology, biogeography, population dynamics*. Centre for Overseas Pest Research, London.

Van Huis, A. (2007). Locust monitoring and management. *Annual Review of Entomology, 52*, 505–521. https://doi.org/10.1146/annurev.ento.52.110405.091158

Annex 1: Research Schedule

Research Schedule

2025	JANUARY	FEBRUARY	MARCH	APRIL	MAY	JUNE	JULY	AUGUST	SEPTEMBER
Ideation	🟩								
Draft Research Questions	🟩								
Writing Research Proposal	🟩	🟩	🟩						
Review of Proposal				🟩					
Presentati					🟩				
Data Collection						🟩			
Model Formulation							🟩		
Data Analysis							🟩		
Model Validation							🟩		
Compiling								🟩	
Final									🟩
Document									🟩

Appendix 3: Resources and Budget

No.	Item	Unit	Price	Total

1	Computer / Laptop	1	50,000	50,000
2	Data Collection	1	60,000	60,000
	Data analysis and modelling	**1**	**30,000**	**30,000**
5	Final Dissertation preparation (i.e., printing, binding, etc.)	500	20 per page	10,000
6	Flash Disk	1	1,500	1,500
7	Miscellaneous expenses	1		10,000
8	Total			138,500

Research Budget