

Addis Ababa University

College of Natural and Computational Sciences

Development of a Vision Transformer Model for Desert Locust Movement Prediction Using Multi-Temporal Remote Sensing Data

Rabra Hierpa Olgera

A Thesis Submitted to the Department of Computer Science in Partial Fulfillment for the

Degree of Master of Science in Computer Science Specialized in Data and Web

Engineering

Addis Ababa University

College of Natural and Computational Sciences

Rabra Hierpa Olgera

Advisor: Solomon Gizaw (PhD)

This is to certify that the thesis prepared by Rabra Hierpa Olgera titled: *Development of a Vision Transformer Model for Desert Locust Movement Prediction Using Multi-Temporal Remote Sensing Data* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science specialized in Data and Web Engineering complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	**Signature**	**Date**
Advisor: Solomon Gizaw (PhD)	_____	_____
Examiner: _____	_____	_____
Examiner: _____	_____	_____

Abstract

Desert locusts pose a critical threat to agricultural production and food security in Ethiopia and the broader Horn of Africa, with traditional monitoring approaches suffering from inconsistent field reports, reporting delays, and limited spatial coverage that hinder effective early warning systems. This thesis presents a Multi-Resolution Heterogeneous Vision Transformer model that leverages 59-band multi-spectral satellite imagery and environmental indicators to predict desert locust presence without relying on crowdsourced or field survey data. The proposed architecture introduces three novel components: multi-resolution patch embedding for capturing environmental patterns at different spatial scales (1×1, 2×2, and 4×4 patches), feature group encoding for processing heterogeneous remote sensing data types, and physics-informed attention mechanisms that incorporate ecological knowledge of locust behavior patterns. Experimental evaluation using FAO Desert Locust Information Service reports from 2015-2023 demonstrates promising performance on test dataset with 82.87% accuracy, 86.09% precision, 87.47% recall, 86.77% F1-score, and 89.01% AUC-ROC, representing a 6.4% improvement in F1-score over existing state-of-the-art methods including PLAN and Prithvi-LB models. The 950,285-parameter model achieves optimal balance between computational efficiency and predictive capability, making it suitable for operational deployment in resource-constrained environments. Future research directions include adapting the model for directional movement vector prediction and expanding geographical validation to other locust-affected regions across East Africa, the Arabian Peninsula, and South Asia.

Keywords: Desert locust prediction, Vision Transformer, Multi-resolution processing, Physics-informed attention, Remote sensing, Ecological modeling, Deep learning, Food security

Dedication

To my beloved mother Tadelech Regassa and father Hierpa Olgera

and

To all Ethiopian household who have been affected by the desert locust crisis in one way or the other.

Acknowledgments

First and foremost, I would like to extend my deepest gratitude to my Lord and savior Jesus Christ. It is through His grace and caring guidance that I found the courage to complete this thesis work.

I would like to express my heartfelt thanks to my advisor, Dr. Solomon Gizaw, for his support, guidance, and continuous encouragement. From the beginning, he provided me with the foundational concept to build a Vision Transformer-based model for desert locust movement prediction. His consistent follow-up and insightful suggestions pushed me to explore further at every phase of the study. I would also like to thank Dr. Biniyam and Flagot for guiding me on the data extraction phase and helping me to access papers from IEEE and other relevant sources.

To everyone who supported me along this journey, both academically and personally, I am truly grateful. Your contributions, encouragement, and assistance have made this achievement possible.

Table of Contents

List of Tables viii

List of Figure ix

List of Algorithms x

Chapter 1: Introduction 1

1.1 Background 1

1.2 Motivation 3

1.3 Statement of the Problem 3

1.4 Objectives 5

1.5 Methods 6

1.6 Scope and Limitations 7

1.7 Application of Results 7

1.8 Thesis Organization 8

Chapter 2: Literature Review 9

2.1 Introduction 9

2.2 Biological and Ecological Background of Desert Locusts 9

2.2.1 General Characteristics and Lifecycle of Desert Locusts 9

2.2.2 Ecological and Environmental Determinants of Desert Locust Outbreaks 10

2.2.3 Impact on Agriculture and Food Security 12

2.3 Environmental Variables Influencing Desert Locust Behavior 13

2.3.1 Vegetation Indices 13

2.3.2 Rainfall, Soil Moisture, and Temperature 19

2.3.3 Seasonal and Spatial Patterns Related to Locust Breeding and Migration 20

2.4 Remote Sensing for Environmental Monitoring 22

2.4.1 Key Satellite Missions and Sensors for Environmental Monitoring 23
2.4 Deep Learning for Spatiotemporal Prediction 27
2.4.1 Convolutional Neural Networks 28
2.4.2 Recurrent Neural Networks and Long Short-Term Memory Networks 29
2.4.3 Convolutional Long Short-Term Memory Networks 30
2.4.4 Attention Mechanisms 30
2.4.5. Hybrid Models 32
2.4.5 Transformer Architectures for Spatiotemporal Data 33
Chapter 3: Related Work 36
3.1 Introduction 36
3.2 Deep Learning-Based Forecasting Models 36
3.3 Remote Sensing and Environmental Predictive Modeling 38
3.4 Operational Systems and Detection Techniques 40
3.5 Summary 43
Chapter 4: Proposed Model 45
4.1. Introduction 45
4.2. Proposed Model Architecture 45
4.2.1 Multi-Resolution Patch Embedding 47
4.2.2 Resolution-aware 2D Positional Encoding 47
4.2.3 Feature Group Encoding 49
4.2.4 Physics-Informed Multi-Head Attention 51
4.2.5 Encoder Stack 53
4.2.6 Global Average Pooling and Linear Classification Head 53
Chapter 5: Experimentation and Evaluation 56
5.1. Introduction 56

5.2. Dataset Collection and Preparation 56
5.3. Experimental System Environment 63
5.4. Experiment and Discussion of Results 64
Chapter 6: Conclusion and Future Work 70
6.1. Introduction 70
6.2. Conclusions 70
6.3. Future Work 72
References 73

List of Tables

Table 5-1 Summary of input features and their spatiotemporal representation 59
Table 5-2 Dataset Composition and Characteristics 63
Table 5-3 Summary of the proposed model performance metrics 66
Table 5-4 Performance comparison with previous locust prediction methods 68

List of Figure

Figure 4-1 The proposed model architecture 46
Figure 5-1 Visual representation of Desert Locust Report from FAO (2015-2023) 57
Figure 5-2 Sample NDVI, EVI and LST band visualization 58
Figure 5-4 Training and Validation accuracy curves 66
Figure 5-5 Confusion Matrix for prediction sample counts 67

List of Algorithms

Algorithm 4-1 Multi-Resolution Patch Embedding 49
Algorithm 4-2 Feature Group Encoding 50
Algorithm 4-3 Physics-Informed Attention 52
Algorithm 4-4 MRH-Vit Forward Pass 54

Acronyms and Abbreviations

AI - Artificial Intelligence

AMP - Automatic Mixed Precision

ARVI - Atmospherically Resistant Vegetation Index

ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer

AUC - Area Under the Curve

CHIRPS - Climate Hazards Group InfraRed Precipitation with Station data

CNN - Convolutional Neural Network

ConvLSTM - Convolutional Long Short-Term Memory

DEM - Digital Elevation Model

DLIS - Desert Locust Information Service

DNN - Deep Neural Network

ENSO - El Niño-Southern Oscillation

ERA5 - European Centre for Medium-range Weather Forecasts Reanalysis

ESA - European Space Agency

ET - Evapotranspiration

EVI - Enhanced Vegetation Index

FAO - Food and Agriculture Organization

FFN - Feed-Forward Network

GEE - Google Earth Engine

GELU - Gaussian Error Linear Unit

GIS - Geographic Information Systems

GNDVI - Green Normalized Difference Vegetation Index

GOES - Geostationary Operational Environmental Satellites

HLS - Harmonized Landsat and Sentinel-2

HMM - Hidden Markov Model

IOD - Indian Ocean Dipole

LSTM - Long Short-Term Memory

LST - Land Surface Temperature

METI - Ministry of Economy, Trade and Industry

ML - Machine Learning

MODIS - Moderate Resolution Imaging Spectroradiometer

MRH-ViT - Multi-Resolution Heterogeneous Vision Transformer

MSAVI - Modified Soil Adjusted Vegetation Index

NASA - National Aeronautics and Space Administration

NCAR - National Center for Atmospheric Research

NCEP - National Centers for Environmental Prediction

NDRE - Normalized Difference Red Edge Index

NDVI - Normalized Difference Vegetation Index

NDWI - Normalized Difference Water Index

NIR - Near-Infrared

NOAA - National Oceanic and Atmospheric Administration

RF - Random Forest

RNN - Recurrent Neural Network

ROC - Receiver Operating Characteristic

RS - Remote Sensing

SAR - Synthetic Aperture Radar

SAVI - Soil Adjusted Vegetation Index

SMAP - Soil Moisture Active Passive

SMOS - Soil Moisture and Ocean Salinity

SPOT - Satellite Pour l'Observation de la Terre

SRTM - Shuttle Radar Topography Mission

SVM - Support Vector Machine

SWIR - Shortwave Infrared

TCI - Temperature Condition Index

TROPOMI - TROPOspheric Monitoring Instrument

TSS - True Skill Statistic

TVDI - Temperature Vegetation Dryness Index

TVI - Transformed Vegetation Index

USGS - United States Geological Survey

VHI - Vegetation Health Index

VIIRS - Visible Infrared Imaging Radiometer Suite

ViT - Vision Transformer

WRF - Weather Research and Forecasting

Chapter 1: Introduction

1.1 Background

The Acrididae family of short-horned grasshoppers includes several species that enter a swarming phase and are collectively referred to as locusts [1]. While they are normally solitary insects, their behavior changes drastically when population density increases. In such conditions, they become gregarious, form large groups, and start migrating across vast areas. Locusts thrive in arid environments such as deserts and dry grasslands. They have been recognized as one of the most dangerous agricultural pests since the beginning of agriculture [1,2]. Because their swarming behavior can cover hundreds of thousands of square kilometers, monitoring and controlling them is highly challenging, especially in regions that are hard to access [3].

Locusts differ from other pests due to their rapid reproductive cycle, unpredictable movement, and capacity to form dense swarms. These swarms can cause widespread devastation within a very short time [2]. Historically, locusts have been associated with serious agricultural damage. For example, British records from Iraq in the 1920s identified locusts as a major agricultural threat [1]. More recently, in 2020, East Africa, including Ethiopia, experienced one of the worst Desert Locust (Schistocerca gregaria) infestations in over 25 years [4]. A single adult locust weighing two grams can travel up to 100 kilometers in a day and consume its own weight in vegetation. A swarm can contain billions of such insects, leading to significant agricultural and environmental destruction [3,4].

The recent Desert Locust outbreaks have caused severe economic, social, and environmental impacts. According to the World Bank, the damage across East Africa and Yemen could reach up to USD 8.5 billion [5]. In Ethiopia, where the majority of the population depends on subsistence farming, such outbreaks pose a serious threat to food security and rural livelihoods [3]. Monitoring of locusts in Ethiopia mainly depends on field surveys conducted by experts and local farmers. These surveys are often submitted manually or through tools like FAO's eLocust3m application [6]. However, this reporting system is prone to delays, inconsistencies, and errors, especially when reports are submitted by non-experts [6, 7]. Because desert locusts move rapidly and change their

position frequently, the collected reports can quickly become outdated and less useful for real-time intervention [8].

Currently, various techniques are applied to detect and monitor locusts. These include direct methods such as ground surveys and trapping, and indirect methods such as analyzing rainfall patterns, vegetation condition, and temperature [8, 9, 10]. Although direct methods can be accurate, they are time-consuming and expensive. Indirect methods allow broader coverage but often lack high-resolution insights. With the advancement of satellite technologies and data availability, remote sensing has become a valuable tool to monitor environmental changes over large areas at reduced costs [2, 5].

Remote sensing refers to collecting data about objects or regions without physical contact. It includes tools like satellites and airborne sensors to observe vegetation health, soil moisture, rainfall, and other important indicators [11]. When remote sensing is combined with Artificial Intelligence (AI) techniques such as Deep Neural Networks (DNN), it is possible to extract useful patterns and forecast locust activities. These models can analyze environmental and behavioral factors to predict swarm movements and identify potential breeding grounds [6].

To improve predictions for locust movements, several studies have utilized machine learning (ML) and Remote Sensing (RS). The PLAN model by Maryam Tabar et al. combined CNNs, LSTMs, and FFNs with environmental variables but struggled with reliance on crowdsourced data and lacked regional adaptability. Following this, the Prithvi-LB model aimed to predict breeding grounds using multispectral satellite images but faced validation and data completeness issues. Similarly, other approaches, like those by Gómez et al. and Mebrahtu, made strides in accuracy yet were limited by resolution or scope, with many lacking integrations of biological behaviors or real-time data. Despite these advancements, most models are not tailored for Ethiopia, highlighting gaps such as regional validation and data integration. This study aims to develop a locust movement prediction model using remotely sensed indicators specifically designed for Ethiopia, focusing on spatial and temporal patterns to enhance early warning systems for effective locust management

1.2 Motivation

The motivation for this thesis comes from the socio-economic impacts that desert locust invasions have caused in Ethiopia. Agriculture is the main source of livelihood for the majority of Ethiopians, especially in rural areas where people depend on seasonal crops and livestock. When swarms of desert locusts invade farmland, they destroy not only crops but also pasturelands, leading to food shortages, loss of income, and increased vulnerability among farming communities. According to the Food and Agriculture Organization (FAO), the 2019–2021 desert locust upsurge affected over 800,000 farming households in Ethiopia, damaging more than 197,000 hectares of cropland and over 1.3 million hectares of pasture. This caused the loss of an estimated 356,000 metric tons of cereal crops and contributed to food insecurity for 8.6 million people in the country[1].

These facts show the need to improve locust early warning and control systems in Ethiopia. Despite the efforts of local authorities and international partners, many interventions have been reactive rather than preventive. Current forecasting methods often rely on delayed field reports and limited surveillance coverage. However, with the increasing availability of satellite imagery and environmental data, there is a growing opportunity to develop remote sensing-based solutions that can support accurate and timely predictions of locust movement. This thesis is motivated by the need to strengthen the locust monitoring capacity of Ethiopia using remote sensing and deep learning methods to help reduce the damage on agriculture, economy, and livelihoods.

1.3 Statement of the Problem

Desert locusts are among the most destructive migratory pests affecting agricultural production and food security in Ethiopia and the broader Horn of Africa. These insects can travel over 100 kilometers per day, rapidly shifting locations and crossing borders, making their movement extremely difficult to monitor and predict [6]. Traditionally, farmers and field experts in Ethiopia rely on field surveys to report sightings. In some regions like Kenya, mobile applications such as eLocust3m have been introduced by FAO to crowdsource locust presence data [6]. However, these field reports, whether manually collected or crowdsourced, suffer from inconsistencies, reporting delays, and lack of expert validation. Consequently, the reports may not accurately reflect the

[1] https://openknowledge.fao.org/handle/20.500.14283/a02769en

current location or migration trajectory of swarms, complicating timely pest control operations such as insecticide spraying [6, 7].

To overcome the limitations of manual reporting, several studies have attempted to apply ML and remote sensing for forecasting desert locust presence and movement. One such model is PLAN by Maryam Tabar et al. [6], which combined Convolutional Neural Networks (CNNs,) Long Short-Term Memory (LSTMs), and Feed-Forward Networks (FFNs) to predict locust presence using environmental variables and crowdsourced observations. Although PLAN achieved high performance, its reliance on crowdsourced data posed a major limitation, as the removal of such data sharply reduced accuracy. The model also lacked regional adaptability and focused on the East African region in general. Building upon PLAN, Salihu et al. [12] introduced the Prithvi-LB model, which used high-resolution multispectral satellite images to predict breeding grounds. However, this model lacked field-based ground truth validation and experienced limitations due to missing values in remote sensing data. Similarly, Lucile Marescot et al. [13] used a Random Forest (RF) model that combined historical eLocust data with dynamic variables like NDVI and rainfall to produce real-time probability maps of locust presence. Yet, the study did not consider biological phases or behavioral traits of locusts and faced issues related to biased field survey data.

Other studies explored different approaches. Gómez et al. [14] focused on satellite-derived Soil Moisture (SM) as a key predictor for breeding ground suitability, achieving high accuracy through RF models, but limited by coarse spatial resolution and lack of biological data. Mebrahtu [8] developed an image classification system using ResNet50 to distinguish locusts from non-locusts, yet the system only processed still images without movement prediction or integration of environmental features. Within Ethiopia, Fentahun [7] used MODIS EVI, Sentinel-1 SAR, and rainfall data along with kriging interpolation to assess infestation in Dire Dawa and Somali regions. Although the approach showed potential, the model's predictive power was weak due to sparse and irregular field data, emphasizing the challenge of data sparsity in Ethiopia. Additionally, Keith Cressman [15] explained the FAO's multi-tiered monitoring using remote sensing and expert knowledge through DLIS, while noting that current sensors have limitations in detecting sparse vegetation and face seasonal accuracy issues. Z. Shao et al. [16] applied a Hidden Markov Model using time-series satellite indices to forecast locust severity in various countries. Despite

acceptable accuracy, the model did not include soil moisture or wind data and relied on indirect locust indicators.

While these models provide useful insights, most were developed with datasets and priorities specific to other countries or regions, and few are tailored for the Ethiopian context. Current studies highlight common gaps, including lack of region-specific validation, insufficient integration of biological behavior, dependence on inconsistent or delayed ground data, and challenges related to data completeness and spatial resolution. These limitations reduce the accuracy and timeliness of predictions, which are essential for effective desert locust management in Ethiopia.

Therefore, this study seeks to address these limitations by investigating the development of a locust movement prediction model based on remotely sensed environmental indicators suitable for Ethiopia. The model aims to minimize reliance on field-based or crowdsourced data and instead focus on integrating spatial and temporal environmental patterns that influence locust mobility. The investigation will focus on designing an accurate, adaptive, and scalable forecasting approach that supports early warning systems and timely decision-making for desert locust control in Ethiopia.

1.4 Objectives

General Objective

The general objective of this thesis is to develop a deep learning-based predictive model using remote sensing data and environmental indicators to forecast the movement of desert locusts in Ethiopia

Specific Objectives

- Conduct a comprehensive review of existing literature on desert locust prediction methods, remote sensing applications, and the use of deep learning in predictive modeling.
- Collect and preprocess relevant remote sensing data associated with desert locust habitats and mobility.
- Identify key environmental and temporal indicators that influence desert locust movement in Ethiopia.

- Design, propose, and implement a DNN model capable of learning spatiotemporal patterns related to locust migration.
- Evaluate the performance of the proposed model using historical data and compare it with existing prediction approaches.

1.5 Methods

To achieve the objectives of this study, the research activities will be carried out in four main stages: reviewing related literature, collecting and preprocessing data, designing the predictive model, and evaluating its performance.

Literature Review

A comprehensive review of existing literature will be conducted to understand current approaches used in desert locust movement prediction. This includes studies on remote sensing applications for environmental monitoring, as well as the use of deep learning in predictive modeling. The review will help identify knowledge gaps, best practices, and appropriate modeling techniques to guide the proposed research.

Data Collection and Preprocessing

Relevant remote sensing data will be collected from reliable sources to support the modeling of desert locust movement. The collected data will undergo necessary preprocessing, including cleaning, formatting, and aligning temporally and spatially. These steps are essential to ensure the data is suitable for input into the deep learning model.

Model Design and Development

Based on the insights gained from the literature review and the prepared dataset, a deep learning model will be designed to capture spatiotemporal patterns related to locust movement. The model will be trained using historical data to learn patterns and trends that contribute to accurate forecasting. The architecture and training approach will be selected to best suit the nature of the data and the prediction objective.

Model Evaluation

The developed model will be evaluated using a separate dataset reserved for testing. Evaluation will involve applying appropriate performance metrics to assess prediction accuracy and reliability. In addition, the model's performance will be compared with existing prediction approaches to demonstrate its strengths and potential improvements.

1.6 Scope and Limitations

This study focuses on the prediction of desert locust movement using remotely sensed environmental indicators and deep learning methods, specifically within the Ethiopian context. The scope is confined to analyzing satellite-derived variables that influence locust mobility, such as vegetation condition, rainfall, soil moisture, and temperature. The objective is to develop a predictive model that captures spatiotemporal patterns associated with desert locust migration using historical environmental data.

However, the study faces several limitations that may influence the generalization and comprehensiveness of the results. First, the availability and quality of satellite data vary across time and regions. Cloud cover, sensor resolution, and gaps in data acquisition may result in missing or inaccurate environmental observations, which can affect model accuracy. Although the study aims to reduce reliance on field and crowdsourced data, the absence of validated ground truth records in some areas limits the ability to fully assess the model performance. In addition, the exclusion of certain biological and behavioral variables due to data scarcity may restrict the depth of the prediction. The model also does not consider policy, intervention, or land-use factors that might influence locust distribution.

The thesis is limited to a specific geographic region, Ethiopia, and a defined timeframe. Therefore, the applicability of the model to other countries or broader temporal scales may require further adaptation and validation. Given the time and resource constraints of this thesis, the work represents only a part of a more extensive locust monitoring and prediction effort.

1.7 Application of Results

The results of this study can contribute to the development of early warning systems for desert locust movement in Ethiopia. By using remote sensing data to forecast potential migration patterns,

the predictive model can assist national agencies such as the Ministry of Agriculture in improving preparedness and resource allocation. Timely and accurate predictions can support proactive control measures, including targeted surveillance and pesticide application, which are essential for reducing crop damage and minimizing socio-economic losses.

The findings can also benefit international organizations such as the FAO and non-governmental organizations involved in food security by enabling better coordination of response strategies. Researchers can also build upon the model and methodology for further studies or adaptation to other pest species.

1.8 Thesis Organization

This thesis is organized into six chapters. Chapter 2 presents a detailed review of the relevant literature, focusing on existing approaches to desert locust prediction, the use of remote sensing applications, and the application of deep learning for locust movement modelling. Chapter 3 discusses related works that have contributed to the development of predictive models for locust monitoring and highlights the gaps this study aims to address. Chapter 4 describes the design of the proposed model. It outlines the methodologies and techniques employed in developing the predictive framework. Chapter 5 explains the training and evaluation process. It presents the experiments conducted to assess model performance and compares the results with existing approaches. Finally, Chapter 6 concludes the thesis by summarizing the key findings and contributions. It also offers recommendations for future research directions.

Chapter 2: Literature Review

2.1 Introduction

This chapter gives a review of literature that is important for this study. The main purpose is to understand what is already known about desert locusts, how their behavior is linked to the environment, and how technologies like remote sensing and deep learning can be used to predict their movements. This understanding will help to build a good model for forecasting locust movement in Ethiopia.

The chapter is structured as follows. First, it looks at the biology and ecology of desert locusts, including their lifecycle and what makes them swarm. Then, it discusses the environmental variables, like vegetation, rainfall, soil moisture, and temperature, that affect where locusts live and move. After that, the chapter explains remote sensing technology, how it is used to monitor the environment, and its good points and limitations for locust forecasting. Next, it introduces deep learning methods and why they are useful for finding patterns in complex data. Finally, the chapter talks about how remote sensing and deep learning can be combined for better prediction of ecological events like locust outbreaks. This review will show what is already done and what gaps this research can fill.

2.2 Biological and Ecological Background of Desert Locusts

2.2.1 General Characteristics and Lifecycle of Desert Locusts

The desert locust (Schistocerca gregaria) is considered one of the most dangerous migratory pests in the world. Throughout history, this insect has been responsible for causing serious agricultural damage and contributing to famines in different regions [17]. It belongs to the Acrididae family, under the Orthoptera order, and is widely known for its ability to change both its appearance and behavior in response to population density. This unique ability is called phase polyphenism [18]. When locusts are few in number, they live in the solitary phase. In this phase, they usually have dull colors and behave independently. However, when the population becomes dense because of favorable breeding conditions, they start to change. They enter the gregarious phase, which is marked by brighter colors, increased mobility, and the tendency to group together in large swarms [19, 20].

The lifecycle of the desert locust has three main stages: egg, nymph (also known as hopper), and adult. Female locusts lay their eggs in sandy soils that are warm and moist, which are ideal for egg development. Each egg pod may contain between 50 and 100 eggs [21]. The time it takes for the eggs to hatch depends largely on environmental conditions, especially the temperature and moisture of the soil. Normally, the eggs hatch within 10 to 20 days [22]. After hatching, the nymphs appear. These hoppers cannot fly and must go through five or six moults (called instars) over a period of 30 to 45 days. During this time, they gradually grow wings and prepare to become adults [19]. Adult locusts can live up to five months. If environmental conditions remain favorable, they may breed several times during their lifespan. Each female is capable of laying several egg pods, which can result in a quick increase of the locust population [23].

The shift from the solitary phase to the gregarious phase is mostly influenced by physical contact and crowding, especially during the hopper stage. When hoppers frequently touch each other due to high population density, this stimulates their nervous system and triggers hormonal changes. These changes affect the locusts' body color, behavior, and internal functions [24, 25]. Behavioral changes, like the desire to group and move in the same direction, can happen within a few hours. Morphological changes, such as alterations in color and body shape, usually take one full generation to become clear. In the gregarious phase, locusts are strongly attracted to each other and move as coordinated groups. These groups can eventually grow into massive swarms, sometimes made up of billions of individual locusts, which can travel long distances and damage crops across large areas [26].

2.2.2 Ecological and Environmental Determinants of Desert Locust Outbreaks

Desert locusts are mainly found in dry and semi-dry areas across Africa, the Middle East, and parts of Southwest Asia [27, 28, 29]. Their usual habitat is known as the recession area, which is very large, covering about 30 million square kilometers [30]. This wide region includes countries like Mauritania, Sudan, Ethiopia, Saudi Arabia, Iran, and India [31]. The ecology of the desert locust is strongly connected to weather conditions, especially rainfall. Rain is very important because it makes the soil moist for egg-laying and helps new green plants grow, which are needed for both the young hoppers and the adult locusts to eat [19, 30].

After seasonal rains fall, plants begin to grow in the desert, creating a good environment for locusts to breed [21, 32]. If the rains continue for several months, more than one generation of locusts can hatch and grow. This causes the population to increase quickly [33]. When there are many locusts in one area with good breeding conditions, they gather and change from solitary to gregarious phase. This transformation into the gregarious phase can lead to the formation of swarms and sometimes result in serious upsurges or even plagues in the region [34, 35].

In general, it is found that when rainfall is more than 25 mm in a short period, it is enough to support breeding activities [36]. Moist soil and plenty of food help eggs to develop well, allow more nymphs to survive, and improve the ability of adults to lay eggs [37]. However, if there is a long dry period or if temperatures get too high, development can be affected negatively and fewer locusts survive [36].

Temperature also plays an important role in the life and behavior of desert locusts [38]. Eggs hatch faster when the temperature is between 30 to 35°C, and the hoppers also grow quicker under similar conditions [39]. But when the temperature goes over 40°C, it can become harmful for both eggs and hoppers, especially if the soil is dry at the same time [22, 38].

Adult swarms usually prefer to fly during the day when the temperature is moderate and the wind is good for movement [40]. With the help of wind, swarms can travel far, sometimes flying up to 150 kilometers in a single day [41]. The wind's direction and strength also affect how and where the swarms move [42]. For example, during the 2019–2021 locust outbreak in East Africa, strange wind patterns helped swarms move from their breeding places in the Arabian Peninsula to the Horn of Africa. There, they found more good conditions for breeding [43]. This situation caused a huge increase in the number of locusts and serious crop damage in Ethiopia, Somalia, and Kenya [44, 45].

The type of plants in an area also matters. Locusts mostly prefer to feed on green grasses and cereal crops, which are common after the rainy season [38]. Scientists often use tools like the Normalized Difference Vegetation Index (NDVI) to watch the health of plants and to help guess where locusts might breed [36, 46]. Areas where NDVI values go up quickly after rain are more likely to attract locusts and face infestation [36, 47].

2.2.3 Impact on Agriculture and Food Security

Desert locusts are among the few pests capable of causing near-total crop destruction within hours [48]. A single swarm, covering one square kilometer, may consist of 40 to 80 million locusts [49]. Such a swarm can consume the equivalent of food for 35,000 people in just one day [49]. When multiple swarms converge over agricultural regions, the resulting destruction can lead to famine, displacement, and socio-economic instability [50, 51].

Locusts consume a wide variety of crops, including wheat, barley, maize, sorghum, millet, vegetables, and fruit trees [52]. They also destroy pasturelands, leaving livestock without enough food to survive. During outbreaks, rural communities lose not only their immediate food supply but also their income sources, as both crop yields and livestock productivity plummet [53]. The cumulative impact is severe especially in regions like Ethiopia, where over 80% of the population depends on agriculture for survival [54]. The 2020 East Africa locust outbreak was the worst in over 25 years, affecting over 1.25 million hectares of cropland and pastureland. In Ethiopia alone, over 200,000 hectares of farmland were damaged, leading to the loss of 356,000 metric tons of cereals. Over 800,000 farming households were directly affected, and millions more experienced food insecurity due to rising food prices and market disruptions [55, 56].

The economic burden of such outbreaks is concerning. In addition to crop and livestock losses, countries spend millions of dollars on aerial spraying campaigns, surveillance missions, and emergency food aid [57]. The World Bank estimated that the total cost of response and recovery for the 2020 outbreak in East Africa could exceed $8.5 billion. These costs strain national budgets and often divert funds from critical development programs [58].

Food insecurity driven by locust outbreaks has broader implications for public health and political stability [58, 59]. Malnutrition, especially among children and pregnant women, increases significantly in affected areas [60]. Governments also face pressure from displaced populations and rural unrest. As a result, early warning systems, regional cooperation, and investment in remote sensing and predictive modeling have become essential strategies for mitigating the risks associated with desert locust coming outbreaks .

In general, the biological and ecological characteristics of the desert locust, combined with favorable environmental conditions, enable rapid population growth and extensive migration.

These attributes make the desert locust a persistent and severe threat to agriculture and food security, particularly in vulnerable regions such as Ethiopia. Understanding these dynamics is critical for developing effective prediction and control strategies, especially in the face of increasing climate variability.

2.3 Environmental Variables Influencing Desert Locust Behavior

2.3.1 Vegetation Indices

Vegetation indices are mathematical expressions that are derived from the spectral reflectance values captured by remote sensing sensors. They provide useful information about the condition, density, and health of vegetation [61]. In desert locust studies, these indices have great importance because they help to monitor vegetation dynamics in both breeding and migratory habitats [32]. The behavior and population changes of desert locusts are strongly related to the availability, distribution, and status of vegetation cover [39]. Vegetation is important not only as food for nymph and adult locusts, but also as a key ecological factor that determines habitat suitability and breeding site selection [62]. Remote sensing has become an essential method for assessing vegetation conditions across wide and often unreachable desert areas, where it is difficult or impossible to conduct field-based studies [63]. Using satellite data, researchers are able to monitor vegetation consistently and over large regions.

Several vegetation indices have been developed from remote sensing data, such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Water Index (NDWI), Transformed Vegetation Index (TVI), Atmospherically Resistant Vegetation Index (ARVI), and Normalized Difference Red Edge Index (NDRE). These indices are commonly used to assess vegetation vigor, canopy density, chlorophyll levels, and water availability [64, 65, 66].

These indices are especially useful in arid and semi-arid regions, where even small amounts of vegetation have an important role in supporting desert locust development and movement [67]. As stated by Vrieling et al. [68], vegetation indices from satellite images offer a reliable and timely way to monitor rangeland conditions and to predict locust outbreaks. This is especially helpful in places where traditional field observations are difficult to carry out. With the help of these tools,

researchers and decision-makers can better understand how locust habitats change and take early action to manage and control locust infestations.

A. Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is one of the most widely applied indices for monitoring green vegetation using satellite imagery. It is based on the principle that healthy green vegetation reflects more near-infrared (NIR) light and absorbs more red light. By calculating the difference and ratio between these two spectral bands, NDVI provides a quantitative measure of vegetation health and density [69]. NDVI values range between -1 and +1, where higher values are associated with denser and healthier vegetation, while lower values indicate sparse or stressed vegetation cover [70].

NDVI has proven to be highly effective in identifying areas suitable for desert locust breeding and grangerization. This is because desert locusts are known to prefer recently greened areas for egg-laying and feeding activities [71]. As a result, NDVI is commonly used in locust monitoring and early warning systems to detect environmental conditions favorable for locust outbreaks.

NDVI is mathematically defined as:

$$NDVI = \frac{(NIR - Red\ )}{(NIR + Red)} \qquad (1)$$

Where:

- NIR = Near-infrared reflectance
- Red = Red reflectance

B. Enhanced Vegetation Index

The EVI is an advanced vegetation index developed to address some of the limitations of the Normalized Difference Vegetation Index (NDVI). While EVI shares similarities with NDVI, it includes the use of blue light reflectance in addition to red and near-infrared (NIR) bands. This allows EVI to correct for atmospheric distortions and minimize the influence of soil background noise, making it more accurate in areas with dense vegetation or where soil reflectance significantly affects the signal [65, 72]. Due to these features, EVI often provides improved sensitivity in high biomass regions where NDVI tends to saturate.

Although EVI is not as widely applied in arid and semi-arid environments as NDVI, it has demonstrated usefulness in detecting rapid vegetation changes following rainfall events, which are key ecological triggers for desert locust activity. This capacity makes EVI a valuable component in early warning systems, particularly for assessing areas that may become favorable for locust breeding shortly after precipitation [72].

EVI is calculated using the following formula:

$$EVI = G\ X\ \frac{(NIR - Red)}{(NIR + C_1\ X\ Red - C_2\ X\ Blue + L)} \qquad (2)$$

Where:

- NIR = Near-infrared reflectance
- Red = Red reflectance
- Blue = Blue reflectance
- G = Gain factor (typically 2.5)
- C_1, C_2 = Coefficients for atmospheric resistance (typically 6 and 7.5)
- L = Canopy background adjustment (typically 1)

The integration of NDVI and EVI into locust monitoring models has significantly improved the accuracy of identifying potential breeding grounds. Several studies have demonstrated that desert locust upsurges often correspond with NDVI anomalies, which indicate sudden vegetation growth in areas that are normally barren [15]. Such vegetation changes provide favorable ecological conditions for locust reproduction and development. Organizations such as the Food and Agriculture Organization (FAO) and various regional monitoring systems have adopted NDVI-based vegetation greenness maps to detect ecological shifts that may signal the onset of locust activity [73]. The rapid detection of these changes allows for timely deployment of ground surveys and targeted interventions, which are essential for preventing the development of large-scale swarms.

C. Soil Adjusted Vegetation Index

The soil adjusted vegetation index (SAVI) is specifically developed to minimize the influence of soil brightness in satellite-based vegetation assessments, particularly in regions with sparse or low vegetation cover [74]. This is especially important in arid and semi-arid environments, such as

those inhabited by desert locusts, where exposed soil can significantly affect vegetation reflectance readings [75]. SAVI modifies the traditional NDVI formula by introducing a soil brightness correction factor, commonly denoted as L (usually set to 0.5), which improves the sensitivity of vegetation detection in dryland areas [74].

The SAVI formula is expressed as

$$SAVI = \frac{(NIR - Red)\,(1 + L)}{(NIR + Eed + L)} \qquad (3)$$

Where:

- NIR = Near-infrared reflectance
- Red = Red reflectance
- L = Soil brightness correction factor (usually 0.5)

D. Modified Soil Adjusted Vegetation Index

Modified soil adjusted vegetation index (MSAVI) further refines SAVI by optimizing the soil adjustment factor [76]:

$$MSAVI = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2} \qquad (4)$$

Where:

- NIR = Near-infrared reflectance
- Red = Red reflectance

E. Green Normalized Difference Vegetation Index

The green normalized difference vegetation index (GNDVI) is a modified form of NDVI that substitutes the green band for the red band in its calculation [77]. This index is particularly sensitive to the chlorophyll content of vegetation, which makes it an effective indicator of plant health and photosynthetic activity [78]. GNDVI is calculated as:

$$GNDVI = \frac{(NIR - Green)}{(NIR + Green)} \qquad (5)$$

Where:

- NIR = Near-infrared reflectance
- Green = Green reflectance

By using the green band, GNDVI enhances the detection of stress in plants before visible symptoms appear, which is useful in early warning systems for vegetation monitoring [79].

F. Normalized Difference Water Index

The Normalized Difference Water Index (NDWI) is designed to monitor changes in water content of leaves and vegetation canopies [80]. It uses the near-infrared and shortwave infrared bands in its formulation:

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \qquad (6)$$

Where:

- NIR = Near-infrared reflectance
- SWIR = Shortwave infrared reflectance

A higher NDWI value indicates higher water content in vegetation, which is especially important in dryland and desert environments where water availability is a limiting factor for plant growth [81]. This index is useful in studying plant water stress and drought conditions, and it plays a crucial role in identifying areas that are favorable for desert locust breeding due to moisture availability [82].

G. Transformed Vegetation Index

The transformed vegetation index (TVI) is a modified version of NDVI that applies a square root transformation to amplify the vegetation signal [83]:

$$TVI = \sqrt{\left(\frac{NDVI + 0.5}{1.5}\right)} \qquad (7)$$

Where:

- NDVI = Normalized Difference Vegetation Index

TVI increases the contrast between vegetated and non-vegetated surfaces and improves visual interpretation in satellite images [84]. It is often used when moderate to dense vegetation cover needs to be assessed, especially in areas where NDVI tends to saturate and becomes less effective [85].

H. Atmospherically Resistant Vegetation Index

The atmospherically resistant vegetation index (ARVI) is designed to reduce the effects of atmospheric disturbances, particularly those caused by aerosols and thin clouds. It modifies the red band using the blue band to correct for atmospheric scattering [86]:

$$ARVI = \frac{NIR - (2\ X\ Red - Blue)}{NIR + (2\ X\ Red + Blue)} \qquad (8)$$

Where:

- NIR = Near-infrared reflectance
- Red = Red reflectance
- Blue = Blue reflectance

ARVI is especially useful in desert and semi-arid regions where atmospheric conditions often affect the reliability of traditional vegetation indices [87]. It provides more stable vegetation measurements under variable lighting and atmospheric conditions [88].

I. Normalized Difference Red Edge Index (NDRE)

The normalized difference red edge index (NDRE) uses the red-edge band, which is located between the red and near-infrared bands, and is highly sensitive to changes in chlorophyll content and early plant stress [89]:

$$NDRE = \frac{(NIR - RedEdge)}{(NIR + RedEdge)} \qquad (9)$$

Where:

- NIR = Near-infrared reflectance
- RedEdge = Red edge reflectance

NDRE is valuable for detecting early signs of plant stress before they are visible to the human eye or detected by traditional indices such as NDVI [90]. It has been widely used in precision agriculture, forest health assessment, and ecological monitoring, especially in identifying subtle stress factors affecting plant productivity [91].

2.3.2 Rainfall, Soil Moisture, and Temperature

Rainfall is widely recognized as one of the most influential environmental factors that trigger desert locust outbreaks [17]. It plays a central role in shaping locust breeding patterns by affecting both the availability of moist, sandy soils needed for egg-laying and the rapid growth of green vegetation required for hopper and adult sustenance [1]. In the arid and semi-arid zones where desert locusts typically breed, rainfall is usually infrequent and highly erratic [14]. However, when an area receives a significant rainfall event, typically more than 25 mm within a few days, it often leads to the emergence of favorable breeding conditions [92].

An illustrative case is the 2018–2020 desert locust upsurge in East Africa, which was preceded by an unusual sequence of cyclones and storms in the Arabian Peninsula. These weather events brought heavy rainfall to normally arid desert regions [43], creating optimal conditions for multiple successive generations of locusts. The availability of abundant food and moist soils supported unusually rapid population growth, with each generation increasing in size [34]. Even small rainfall amounts, if followed by warm temperatures and retained soil moisture for at least a week, can create suitable conditions for egg development and hatching [93].

Soil moisture is another key determinant in desert locust reproduction. Moist, sandy soil is essential for successful oviposition, and the viability of the eggs depends significantly on appropriate moisture levels [62]. If the soil is overly dry or waterlogged, egg development is disrupted or fails altogether. Soil moisture retention is influenced by several factors, including soil texture, the amount and frequency of rainfall, ambient temperature, and evapotranspiration rates [94]. Remote sensing satellites such as National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) and the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) missions now provide global-scale soil moisture data. These datasets are increasingly incorporated into early warning systems and locust forecasting models to detect areas with conditions suitable for breeding [95].

Temperature also plays a critical role in shaping the biology and behavior of desert locusts. The ideal temperature range for egg development lies between 25°C and 35°C, under which eggs typically hatch within 10 to 14 days [39]. Hopper development is also accelerated in this temperature range, leading to the possibility of multiple breeding cycles within a single season [39]. In contrast, low temperatures below 15°C can delay embryonic development or cause egg mortality, while extreme heat above 40°C can dry out eggs and reduce adult locust survival [96].

Adult swarm behavior is also temperature-sensitive [39]. Swarms tend to fly during the cooler parts of the day, such as early morning and late afternoon, and rest during the peak midday heat [8, 97]. Additionally, the distance and direction of swarm migration are affected by thermal updrafts and wind currents, both of which are influenced by atmospheric temperature gradients. Consequently, temperature monitoring plays an essential role in forecasting swarm movement and dispersal routes [98].

2.3.3 Seasonal and Spatial Patterns Related to Locust Breeding and Migration

The breeding and migration of desert locusts follow seasonal patterns that are strongly determined by regional climatic variations [14]. There are three main seasonal breeding zones: the winter breeding areas in Ethiopia, Eritrea, Sudan, and northwestern Somalia; the spring breeding areas along the Red Sea coast and interior parts of the Arabian Peninsula; and the summer breeding zones in the Sahelian belt, such as Chad, Niger, and Mali, and along the Indo-Pakistan border [14].

In each of these regions, the onset of seasonal rains triggers locust breeding. For instance, in the Sahel, summer rains from June to September typically support two or three breeding cycles, while the winter rains along the Red Sea coast from November to March support one to two cycles [99]. Spring breeding in the Arabian Peninsula is often enabled by lingering winter rains and occasional localized storms [100].

The spatial movement of locust swarms is guided by several factors, including prevailing wind directions, vegetation availability, and natural geographical barriers [47].

Wind direction….

For example, swarms emerging in eastern Ethiopia may migrate westward into Sudan or southward into Kenya and Uganda if conditions are favorable. The use of satellite imagery and meteorological

data has become vital in tracking such migrations [14]. Modern modelling tools such as the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model and the Weather Research and Forecasting (WRF) model are frequently employed alongside NDVI and rainfall datasets to predict areas at risk of invasion [101].

The simultaneous activation of multiple breeding zones, combined with the long-distance flight capability of desert locusts, presents a significant challenge for effective locust management [102]. During certain years, such as the 2003–2005 upsurge in West Africa and the 2020 East Africa crisis, all three breeding zones were active at the same time, resulting in widespread outbreaks and regional food security threats [58].

Large-scale climatic phenomena like the Indian Ocean Dipole (IOD) and El Niño-Southern Oscillation (ENSO) are increasingly linked to variations in rainfall timing, intensity, and geographical distribution across key breeding areas [103]. For instance, positive IOD phases are typically associated with enhanced rainfall in East Africa, supporting the growth of vegetation and creating ideal conditions for breeding. Understanding these climatic drivers is essential for long-term forecasting and strategic planning in locust management [104].

Geographic Information Systems (GIS) and spatial analysis tools have greatly enhanced our ability to monitor and respond to desert locust threats. By overlaying environmental datasets such as NDVI, rainfall, soil moisture, and temperature with locust presence records, researchers can produce risk maps that inform decision-making [36]. These tools are particularly valuable in low-resource settings, enabling efficient and cost-effective targeting of control measures [105, 106].

In conclusion, the population dynamics and migration patterns of desert locusts are deeply influenced by an interrelated set of environmental variables. Vegetation indices offer insights into habitat suitability, while rainfall and soil moisture determine where and when breeding can occur. Temperature affects every stage of the locust life cycle and their migratory behavior. Seasonal and spatial variability in these factors explains the complexity of locust movements across the vast desert regions. Integrating remote sensing, climate monitoring, and spatial modelling into early warning systems significantly enhances the capacity for timely interventions and sustainable locust management.

2.4 Remote Sensing for Environmental Monitoring

Remote sensing refers to the science and technology of gathering information about the Earth's surface from a distance using instruments typically placed on satellites or aircraft without making physical contact with the target area [107]. This method involves detecting and measuring radiation reflected or emitted from the surface in different parts of the electromagnetic spectrum, such as visible light, infrared, and microwave wavelengths. The data are generally obtained as digital images or numerical readings, which can be further processed and analyzed to extract useful information. These include land use and land cover types, the condition of vegetation, land surface temperature, water bodies, and soil moisture levels, among other environmental variables [108].

The most significant advantage of remote sensing lies in its ability to observe and record data from large and often inaccessible geographic areas at various time intervals. This makes it a cost-effective and time-efficient approach compared to traditional field surveys, which often require substantial manpower and resources [109]. The repetitive nature of satellite passes ensures consistent data collection over time, which is essential for monitoring dynamic environmental processes and for conducting long-term studies on changes occurring due to natural or anthropogenic causes [110].

In the context of environmental monitoring, remote sensing plays a critical role by offering timely, objective, and scalable information. It supports a wide range of applications such as assessing the extent of deforestation, monitoring wetlands and water resources, studying patterns of urbanization, and analyzing the impact of climate change on natural ecosystems. Because the data are spatially continuous and temporally repetitive, they enable the identification of both gradual trends and abrupt environmental changes. These features make remote sensing particularly useful for governments, researchers, and policy makers who need to base decisions on reliable and up-to-date information [111].

One of the important uses of remote sensing is the monitoring and control of desert locusts. These insects pose a serious threat to food security in several parts of Africa and Asia, and their outbreaks are influenced by specific environmental conditions. Remote sensing helps detect the presence of moist soil and green vegetation, which are favorable conditions for locust breeding. By analyzing vegetation indices like the Normalized Difference Vegetation Index and monitoring rainfall and

soil moisture levels, scientists can identify areas at risk of infestation and take early action to mitigate the threat. Additionally, remote sensing allows for tracking the movement of locust swarms across large desert and semi-arid regions, which are difficult to access using ground-based observation methods due to limited infrastructure and resources [36, 112].

Another benefit of remote sensing is its ability to support environmental analysis at multiple scales. The same technology can be used to study localized phenomena or to conduct regional and global assessments. This multi-scale capability enables better integration of information and coordination among various stakeholders involved in environmental management and disaster risk reduction. For example, in the case of transboundary environmental issues like desert locust invasions, remote sensing provides a common basis for regional collaboration and policy development [113].

2.4.1 Key Satellite Missions and Sensors for Environmental Monitoring

Earth observation satellites play a crucial role in monitoring, understanding, and managing environmental and socio-economic processes across the globe [114]. These systems provide invaluable data that inform scientific research, policy development, natural resource management, and disaster response [115]. A diverse array of satellite missions, developed and operated by space agencies such as NASA, European Space Agency (ESA), and National Oceanic and Atmospheric Administration (NOAA), as well as commercial entities, offer different capabilities in terms of spatial, spectral, and temporal resolution [116]. This section reviews key Earth observation satellite missions, highlighting their technical specifications and primary applications.

A. Landsat Program

The Landsat program, jointly managed by NASA and the United States Geological Survey (USGS), is the longest-running initiative for the acquisition of moderate-resolution Earth observation data [117]. It began with the launch of Landsat 1 in 1972 and has continued with successive missions including Landsat 8 and Landsat 9 [118].

Each Landsat satellite carries multispectral imaging sensors. Earlier missions used the Multispectral Scanner (MSS) and Thematic Mapper (TM), while more recent satellites carry the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) [119]. These instruments provide data in visible, near-infrared, short-wave infrared, and thermal bands, with spatial

resolutions of 30 meters for most bands, 15 meters for the panchromatic band, and 100 meters for thermal bands [119].

Landsat data are critical for applications such as land use and land cover change detection, deforestation monitoring, urban sprawl analysis, and agricultural assessment [120]. Its relatively high spatial resolution, consistent temporal coverage (16-day repeat cycle), and open access data policy have made Landsat one of the most widely used satellite datasets in environmental sciences [121]. For example, forest loss and degradation studies in the Amazon basin often rely on the long-term and high-resolution imagery provided by Landsat [122].

B. Sentinel Missions

The sentinel mission also known as Copernicus Program, coordinated by the European Commission and implemented in partnership with the European Space Agency (ESA), includes a family of Sentinel satellites designed for operational Earth observation services [123] . The main Sentinel missions relevant to environmental monitoring include Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P [124].

Sentinel-1: carries a C-band Synthetic Aperture Radar (SAR) that provides high-resolution imagery regardless of weather conditions and daylight availability [125]. This all-weather capability is especially useful for monitoring flood events, glacier dynamics, and forest structure. Sentinel-1 data are also used in interferometric applications to detect ground subsidence and tectonic motion [125].

Sentinel-2: equipped with the Multispectral Instrument (MSI) that captures imagery in 13 spectral bands with spatial resolutions of 10, 20, and 60 meters. Sentinel-2's high spatial, spectral, and temporal resolution (5-day revisit with both satellites) makes it ideal for agricultural monitoring, forest classification, water quality assessment, and disaster response [126]. The red-edge bands are particularly useful in vegetation analysis, supporting applications like chlorophyll content estimation and crop yield forecasting [126].

Sentinel-3: includes a suite of instruments such as the Ocean and Land Color Instrument (OLCI), Sea and Land Surface Temperature Radiometer (SLSTR), and the Synthetic Aperture Radar Altimeter (SRAL). Sentinel-3 supports the monitoring of sea surface temperature, ocean color,

land surface temperature, and vegetation productivity [127]. These data are essential for climate research, oceanography, and terrestrial ecosystem assessment [127].

Sentinel-5P: launched in 2017, is dedicated to atmospheric monitoring. It carries the TROPOspheric Monitoring Instrument (TROPOMI), which measures key atmospheric pollutants including ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, and methane [128]. This data supports air quality monitoring, emission tracking, and greenhouse gas inventories [128].

C. Moderate Resolution Imaging Spectroradiometer

The moderate resolution imaging spectroradiometer (MODIS) is a key sensor aboard NASA's Terra and Aqua satellites [129]. It provides global coverage every 1–2 days in 36 spectral bands, ranging from 250 meters to 1 kilometer in spatial resolution [130]. MODIS plays a crucial role in producing a variety of environmental products including:

- Vegetation indices such as NDVI and EVI;
- Land cover classifications;
- Surface temperature maps;
- Fire occurrence and thermal anomalies;
- Snow cover and ice extent.

Because of its high temporal resolution and consistent processing standards, MODIS data is widely used for trend analysis in climate change research, ecological monitoring, and natural disaster assessment [131]. MODIS-derived products are also often integrated into Earth system models for forecasting and decision support [132].

D. Visible Infrared Imaging Radiometer Suite

The visible infrared imaging radiometer suite (VIIRS) is part of the Joint Polar Satellite System (JPSS) and continues the legacy of MODIS with improved spatial resolution and sensor capabilities [133]. VIIRS provides imagery in 22 spectral bands, with a spatial resolution of 375 m for imagery bands and 750 m for moderate resolution bands [134]. VIIRS is particularly renowned for its Day/Night Band (DNB), which enables observation of low-light scenes, such as auroras, urban lighting, wildfires, and power outages [135]. This band has revolutionized disaster monitoring and socio-economic studies by enabling analysis of human activity at night [135].

VIIRS also contributes to global environmental datasets including vegetation indices, aerosol optical thickness, sea surface temperature, and snow/ice coverage [136].

E. Advanced Spaceborne Thermal Emission and Reflection Radiometer

The advanced spaceborne thermal emission and reflection radiometer (ASTER), developed jointly by NASA and Japan's Ministry of Economy, Trade and Industry (METI), is one of the high-resolution imaging instruments onboard the Terra satellite. ASTER acquires data in 14 spectral bands across the visible to thermal infrared range, with spatial resolutions of 15 m (VNIR), 30 m (SWIR), and 90 m (TIR) [137]. The ability of ASTER to produce stereo imagery is particularly valuable for creating Digital Elevation Models (DEMs), which support topographic and geomorphological analyses [138]. Other applications include thermal mapping of volcanoes, urban heat island studies, land degradation assessment, and mineral exploration [138].

F. Satellite Pour l'Observation de la Terre

The Satellite Pour l'Observation de la Terre (SPOT) satellites, developed by the French Space Agency (CNES), have provided high-resolution Earth imagery since 1986 [139]. They offer panchromatic and multispectral data with resolutions ranging from 1.5 meters to 20 meters. The SPOT series has played a significant role in vegetation mapping, agricultural monitoring, and urban planning [140]. With their ability to provide stereo-pair imagery, SPOT satellites have also supported terrain analysis and infrastructure monitoring. Their medium-to-high resolution and relatively frequent revisit capability make them suitable for regional-scale environmental studies [141].

G. Proba-V

The Proba-V mission is a small satellite launched in 2013 to continue the data record of the VEGETATION instruments on SPOT-4 and SPOT-5. Proba-V acquires global vegetation imagery at 100 m and 300 m resolutions, with a daily revisit time [142]. Proba-V data are used in applications such as food security monitoring, drought assessment, ecosystem modelling, and biomass estimation. The mission fills a critical gap in low-resolution global vegetation monitoring, supporting both scientific and operational environmental programs [143].

H. Soil Moisture Active Passive

The Soil Moisture Active Passive (SMAP) NASA mission provides global observations of surface soil moisture and freeze-thaw status using L-band microwave radiometry and radar [144]. SMAP data has applications in: hydrological modelling, agricultural productivity forecasting, drought and flood prediction and weather forecasting and climate modelling. SMAP offers a 3-day revisit time and spatial resolutions of 9–36 km, which is sufficient for large-scale soil moisture trend analysis and integration into meteorological models [145].

I. Geostationary Operational Environmental Satellites

The geostationary operational environmental satellites (GOES), operated by National Oceanic and Atmospheric Administration (NOAA), are positioned over the Western Hemisphere and provide continuous monitoring of atmospheric and surface conditions [146]. GOES-R series satellites carry the Advanced Baseline Imager (ABI), which offers multispectral imaging with temporal resolutions as frequent as every 5 minutes [147]. GOES data is essential for real-time weather forecasting, severe storm monitoring, hurricane tracking, and environmental hazard detection such as wildfires and volcanic eruptions. Its geostationary orbit makes GOES particularly valuable for meteorological applications in the Americas [148].

2.4 Deep Learning for Spatiotemporal Prediction

Deep learning has emerged as a powerful approach for modeling complex spatiotemporal phenomena, leveraging its ability to automatically learn hierarchical representations from large-scale data [149]. Its application to environmental and ecological forecasting has grown rapidly, driven by the increasing availability of geospatial data and advances in neural network architectures [150]. Spatiotemporal prediction tasks involve simultaneous modeling of spatial dependencies and temporal dynamics, which poses unique challenges that traditional statistical methods often struggle to address effectively [151]. Deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Convolutional LSTM (ConvLSTM), and Transformer architectures have been adapted to capture these complex dependencies [152].

2.4.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) were originally developed for image recognition tasks, where they demonstrated strong capability in learning spatial features through convolutional kernels [153]. These networks are designed to automatically and adaptively learn spatial hierarchies of features from input data, which makes them highly effective for processing structured grid-like data such as images [154]. In the context of spatiotemporal prediction, CNNs have been adapted to extract spatial representations from gridded input data, including satellite imagery, environmental measurement grids, and other geospatial datasets [155]. These data types often carry rich spatial information that is essential for understanding and forecasting various physical and ecological phenomena [156].

The core strength of CNNs lies in their ability to capture local spatial dependencies through convolution operations, and to aggregate this information through pooling layers to detect broader patterns [155]. By stacking multiple convolutional and pooling layers, CNNs can progressively identify both local and global spatial features, which are critical for understanding complex spatial structures in environmental and ecological systems [157]. For instance, in ecological forecasting, spatial patterns such as vegetation coverage, water body distribution, and land use characteristics can be effectively learned by CNNs and used to predict future states [158].

Despite their effectiveness in spatial feature extraction, CNNs do not inherently capture temporal dynamics [159]. This limitation means that while CNNs can understand where certain features are located and how they relate spatially, they cannot by themselves model how these features evolve over time [160]. For dynamic prediction tasks, such as forecasting weather patterns, pollution spread, or wildlife movement, it is essential to also model the temporal dependencies in the data. Therefore, CNNs are often integrated with temporal sequence modeling architectures such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), or more recently, Temporal Convolutional Networks (TCNs) and Transformer-based models [161]. This combination allows for the simultaneous modeling of spatial and temporal aspects, which is essential for accurate and robust spatiotemporal prediction.

2.4.2 Recurrent Neural Networks and Long Short-Term Memory Networks

Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed to handle sequential data by using a hidden state that is updated at each time step [162]. This hidden state enables the network to retain information from previous inputs and use it when making predictions at later stages, thereby allowing the model to learn temporal dependencies within the data [162]. Such an architecture makes RNNs suitable for time-series data, where each observation is dependent on previous values. However, traditional RNNs are known to face significant challenges when learning from long sequences due to the vanishing and exploding gradient problems [163]. These issues make it difficult for the model to learn long-term dependencies effectively, which limits their usefulness in many real-world applications that require long-range temporal understanding [164, 165].

To overcome the limitations of standard RNNs, Long Short-Term Memory (LSTM) networks were introduced. LSTMs extend the RNN architecture by incorporating a memory cell and three gating mechanisms: the input gate, forget gate, and output gate [166]. These gates control the flow of information into, within, and out of the memory cell, allowing the network to maintain and update relevant information across longer time sequences [166]. As a result, LSTMs are much better at capturing long-term temporal dependencies compared to basic RNNs, making them more reliable for time-series forecasting tasks [163]. They have been widely applied in domains such as weather prediction, energy demand forecasting, and traffic flow estimation, where understanding temporal patterns is crucial for accurate predictions [167].

Although LSTMs provide significant improvements in modeling temporal dynamics, they do not inherently account for spatial relationships in the data. This is an important limitation, especially in spatiotemporal applications where spatial context and dependencies significantly influence temporal evolution. For instance, environmental phenomena such as rainfall distribution or temperature variation often depend on spatial conditions in neighboring regions [168]. To address this limitation, researchers have proposed hybrid architectures that combine CNNs and LSTMs [169]. In such models, CNNs are used to extract spatial features from gridded input data at each time step, while LSTMs are employed to capture the temporal evolution of these spatial features over time [161]. This integrated approach allows the model to jointly learn spatial and temporal

patterns, leading to improved performance in tasks such as air quality forecasting, agricultural monitoring, and climate prediction [170].

2.4.3 Convolutional Long Short-Term Memory Networks

Convolutional Long Short-Term Memory Networks (ConvLSTM) extends the LSTM model by replacing matrix multiplications in the gates with convolution operations, thereby integrating spatial feature extraction and temporal memory within a single architecture [171]. ConvLSTM units maintain a hidden state and memory cell that are both three-dimensional tensors, capturing spatial and temporal correlations simultaneously [172]. This architecture has been successfully applied to precipitation nowcasting, where high-resolution radar maps are predicted over short time horizons [173]. Similarly, ConvLSTM has been used for sea surface temperature forecasting and air quality prediction, demonstrating its versatility in environmental modeling [174]. By preserving the spatial structure of input data throughout temporal processing, ConvLSTM provides a more natural and effective way to model spatiotemporal dependencies compared to sequential CNN-LSTM combinations [175]. However, ConvLSTM models can be computationally intensive due to the high dimensionality of the convolution operations, especially when applied to large-scale gridded datasets [176].

2.4.4 Attention Mechanisms

Attention mechanisms, initially introduced in the context of sequence-to-sequence models for machine translation, have significantly transformed the way models handle sequential data [177]. The core idea behind attention is to allow the model to selectively focus on the most relevant parts of the input when generating each output. Instead of treating all parts of the input equally, attention assigns different weights to different input elements, helping the model to concentrate more on the influential ones [178]. This makes attention especially useful in spatiotemporal forecasting, where it is important to capture both spatial and temporal dependencies [179].

In the context of spatiotemporal data, attention can be extended to operate not only across time steps but also across spatial locations [180]. Temporal attention helps the model to identify and emphasize critical periods in the time series that strongly influence the target outcome. Similarly, spatial attention allows the model to highlight specific regions in space that contribute more significantly to the prediction [181]. This dual attention mechanism enhances the model's ability

to capture complex dependencies and improves its interpretability by showing which parts of the input are most relevant [182].

2.4.4.1 Physics-informed Attention Mechanism

Physics-informed attention mechanism is a relatively new concept that tries to bring together the strengths of deep learning, especially attention models like the Vision Transformer, with the knowledge of physical laws and constraints [183]. The main idea behind this approach is that instead of relying only on large datasets and patterns learned from data, the model also considers domain-specific physics knowledge to guide its learning and prediction process [184, 185]. This becomes especially useful in fields like environmental modeling, weather forecasting, and movement prediction of natural phenomena such as desert locusts, where the underlying processes are governed by known physical principles [186].

One of the key characteristics of physics-informed attention is its ability to inject prior physical knowledge directly into the attention mechanism. Normally, attention models compute the relevance of different input features based on learned representations [187]. However, in a physics-informed setup, this process is enriched with additional constraints or guidance that reflect the physical behavior of the system [184]. For example, when predicting locust movement, certain environmental patterns such as wind direction, temperature, humidity, or vegetation cover can be considered as physically important features. Physics-informed attention mechanisms allow these features to be given more meaningful weight during the model's decision-making process [188].

The benefits of using physics-informed attention mechanisms are notable. Firstly, they help improve model generalization. By incorporating physical rules, the model is less likely to overfit to noise in the training data and can make better predictions even in unseen situations [189]. Secondly, it improves interpretability. Since the attention mechanism reflects physical reasoning, the output becomes more understandable and trustworthy [187]. Thirdly, it supports robustness. Physics-based constraints act as regularizer, helping the model remain stable and perform consistently even when some input data are missing or noisy [188].

In the context of self-attention mechanisms like those used in Transformers, physics-informed methods work by modifying the computation of attention weights [190]. Instead of purely using dot-product similarity between queries and keys, additional components representing physical

laws or spatial-temporal relationships can be added. These components can guide the attention to focus on regions or time frames that are physically meaningful [187]. For example, when analyzing satellite images over time, the attention mechanism can be directed to focus more on time periods when known weather patterns influence locust behavior. This integration creates a hybrid system where the model learns both from data and from physical understanding.

Overall, physics-informed attention provides a powerful way to enhance Transformer-based models for tasks involving natural systems [191]. It bridges the gap between traditional physical modeling and modern deep learning, making it a valuable approach for complex prediction tasks like desert locust movement using remote sensing data. By blending learning with knowledge, it helps develop models that are not only accurate but also reliable and explainable [192].

2.4.5. Hybrid Models

To further improve predictive performance, several studies have explored hybrid models that combine CNNs, RNNs, and attention layers [193]. In these architectures, CNNs are used to extract spatial features from grid-structured input data, such as satellite images or environmental sensor maps. RNNs or LSTM networks are then used to capture the temporal dynamics of the extracted features. Attention layers are integrated on top to refine the model's focus on the most informative spatial regions and time points [193]. These hybrid models leverage the strengths of each component, including spatial awareness from CNNs, temporal memory from RNNs, and selective focus from attention, to achieve improved accuracy and robustness in forecasting tasks.

For instance, the Spatial-Temporal Attention Graph Convolutional Network (STAGCN) integrates graph-based spatial convolutions with temporal attention mechanisms to model dynamic systems such as traffic networks [194]. In this model, spatial relationships between sensor locations are represented as a graph structure, and temporal dependencies are learned using attention-enhanced modules. This allows the model to predict traffic speed across a city by dynamically adjusting its focus based on changing traffic patterns and localized spatial influences.

Moreover, attention-based hybrid models have been successfully applied in other domains such as human activity recognition, climate variable prediction, and environmental monitoring [195, 196, 197, 198]. By revealing which specific locations and time intervals are most influential in the prediction process, these models offer not only better performance but also greater interpretability.

This makes them valuable tools for decision-making in complex systems where understanding the rationale behind a prediction is just as important as the prediction itself.

To further improve predictive performance, several studies have explored hybrid models that combine CNNs, RNNs, and attention layers [183]. In these architectures, CNNs are used to extract spatial features from grid-structured input data, such as satellite images or environmental sensor maps. RNNs or LSTM networks are then used to capture the temporal dynamics of the extracted features. Attention layers are integrated on top to refine the model's focus on the most informative spatial regions and time points [183]. These hybrid models leverage the strengths of each component, including spatial awareness from CNNs, temporal memory from RNNs, and selective focus from attention, to achieve improved accuracy and robustness in forecasting tasks.

For instance, the Spatial-Temporal Attention Graph Convolutional Network (STAGCN) integrates graph-based spatial convolutions with temporal attention mechanisms to model dynamic systems such as traffic networks [184]. In this model, spatial relationships between sensor locations are represented as a graph structure, and temporal dependencies are learned using attention-enhanced modules. This allows the model to predict traffic speed across a city by dynamically adjusting its focus based on changing traffic patterns and localized spatial influences.

Moreover, attention-based hybrid models have been successfully applied in other domains such as human activity recognition, climate variable prediction, and environmental monitoring [185, 186, 187]. By revealing which specific locations and time intervals are most influential in the prediction process, these models offer not only better performance but also greater interpretability. This makes them valuable tools for decision-making in complex systems where understanding the rationale behind a prediction is just as important as the prediction itself.

2.4.5 Transformer Architectures for Spatiotemporal Data

The Transformer architecture, initially proposed for natural language processing (NLP), introduces a novel mechanism known as self-attention, which allows the model to evaluate relationships between all positions in a sequence simultaneously [172]. Unlike RNNs, which process data sequentially and can struggle with long-term dependencies, Transformers handle the entire sequence in parallel [189]. This property significantly improves their efficiency and ability to learn

long-range dependencies in data [190]. These advantages have led to the adaptation of Transformer models for spatiotemporal prediction tasks.

In spatiotemporal contexts, Transformers have been modified to handle both spatial and temporal dependencies. This is typically achieved by embedding spatial position information along with temporal encoding and applying multi-head self-attention mechanisms across these dimensions [191]. Through this approach, Transformers can model complex interactions between different spatial regions and across various time steps. This capability is particularly beneficial for forecasting tasks where spatial phenomena, such as weather systems or traffic patterns, evolve over time and exhibit intricate dependencies [192].

Transformers have been successfully applied to a variety of spatiotemporal prediction problems. For example, in traffic forecasting, Transformers can learn correlations between traffic sensor readings at different locations and times, which helps predict future traffic conditions more accurately [193]. In environmental science, Transformer models have been used to forecast air pollution levels and oceanographic variables by integrating multiple data sources, such as satellite images and ground-based measurements [194]. These applications demonstrate the flexibility and scalability of Transformers, especially for large-scale problems involving high-dimensional inputs. However, it is important to note that Transformer models typically require substantial amounts of data and computing power for training, which can be a limitation in some real-world settings [195].

More recently, Vision Transformers (ViTs) have emerged as a powerful extension of the Transformer architecture for image and grid-based data, which are commonly used in spatiotemporal applications [196]. Unlike conventional neural networks that process images through local receptive fields, ViTs divide an input image into fixed-size patches, flatten them, and treat each patch as a token, similar to words in a sentence [197]. These tokens are then passed through a standard Transformer encoder, which uses self-attention to capture global relationships across the entire image [197]. This global context awareness allows ViTs to effectively learn spatial features across distant parts of the input data [198].

For spatiotemporal prediction tasks, ViTs can be extended by adding temporal positional encodings or by stacking time-series sequences of image patches as input [199]. This enables the model to understand both spatial and temporal patterns simultaneously. Applications of ViTs in

this domain include rainfall prediction from radar imagery, vegetation change monitoring using multi-temporal satellite images, and disease outbreak forecasting through geospatial health data [200]. The ability of ViTs to handle high-resolution data and to learn rich global representations makes them especially suited for complex environmental forecasting tasks [201].

Chapter 3: Related Work

3.1 Introduction

Desert locust outbreaks represent a serious threat to agriculture and food security in many regions, especially in East Africa. Their ability to travel long distances and shift locations quickly makes their movement difficult to predict. As a result, several studies have been conducted to improve forecasting techniques using remote sensing data and machine learning. These studies apply different approaches, including image-based identification, environmental modeling, and deep learning.

This chapter presents a review of relevant existing studies related to the prediction of desert locust presence, movement, and breeding activity. Each study is discussed by focusing on the methods used, the type of data applied, the results achieved, and the strengths and limitations observed. The aim of this chapter is to identify important gaps and practical challenges in the existing body of knowledge. This review sets the foundation for the proposed model, which aims to address these gaps by developing an effective locust movement prediction approach based on remote sensing data suitable for the Ethiopian context.

3.2 Deep Learning-Based Forecasting Models

Maryam Tabar et al. [6] conducted a study with the primary objective of developing a deep learning model named PLAN to forecast desert locust movement in East Africa. The model was designed to support the efforts of the UN-FAO and PlantVillage by providing high-resolution predictions of locust presence and absence using a combination of crowdsourced observations and remote sensing environmental data. PLAN integrates three neural network components: CNNs to process spatial features, LSTMs to model temporal patterns, and a Feed-Forward Neural Network to handle static environmental factors such as land elevation, soil type, and vegetation productivity. The dataset included approximately 21,000 crowdsourced reports from Kenya, Ethiopia, and Somalia collected between March and September 2020, in addition to ten environmental variables. The authors used a walk-forward validation approach and compared PLAN's performance with classical machine learning models including Logistic Regression, SVM, AdaBoost, and XGBoost.

The results showed that PLAN achieved an F1 score of 0.7918 and an AUC of 0.8904 for next-day forecasts, consistently outperforming all baselines. It also maintained high performance on multi-day forecasts, with an average AUC of 0.84 across two-day to four-day prediction tasks. Furthermore, a cross-region test using data from Iran demonstrated that PLAN achieved a 23 percent higher F1 score than XGBoost, confirming its ability to generalize to new regions. The strengths of this study include its integration of real-world locust observation data with environmental conditions, the innovative design of its spatiotemporal model, and its strong predictive performance across both short and mid-term horizons. However, the study also acknowledges several limitations and gaps. One major weakness lies in the quality and consistency of the crowdsourced data, which may include errors, reporting delays, and false positives due to voluntary, non-expert participation. Additionally, the environmental variables used were specific to East Africa, and their applicability to other regions proved limited, as seen in the reduced performance during cross-region tests. The data distribution was also uneven across space and time, leading to sparsity in certain areas. These gaps highlight the need for improved data validation mechanisms, integration of more comprehensive environmental features, and regional customization of model parameters. One of the key lessons from this work is that while machine learning can effectively support locust movement prediction, it must be carefully aligned with domain knowledge and real-world data challenges.

Ibrahim Salihu et al. [12] conducted a study with the objective of developing a robust, operationally-ready model for predicting actual breeding grounds of desert locusts (DL), in contrast to earlier studies that focused mainly on broader habitat suitability. They aimed to assist early warning systems and support targeted control measures by accurately identifying areas where locusts are likely to copulate and lay eggs. To achieve this, the authors curated a dataset from 48 years of UN-FAO locust observation records and constructed two main types of input features: remotely-sensed environmental and climatic data such as precipitation, vegetation cover, soil moisture, and multi-spectral earth observation images from NASA's Harmonized Landsat and Sentinel-2 (HLS) dataset. The models employed included several deep learning architectures such as PLAN-LB (a modified PLAN model), Conv3D, ConvLSTM, and a fine-tuned version of the Prithvi ViT-based foundational model (Prithvi-LB). The breeding dataset was balanced with both presence and pseudo-absence records, and spatiotemporal data was arranged into grid-based features. For model training and testing, three subsets (training, validation, and test sets) were used,

and experiments were conducted using cloud-based computing resources such as Google TPU and Nvidia GPUs.

The results showed that among all models tested, Prithvi-LB achieved the highest performance with an accuracy of 83.03%, F1-score of 81.53%, and ROC-AUC of 87.69% using only multi-spectral (raw bands) earth observation images. ConvLSTM, the best performer among the models using only remotely-sensed variables, achieved lower results with 75.76% accuracy and 67.34% F1-score. The findings revealed that multi-spectral imagery alone, without additional climatic or environmental features, was sufficient for accurate breeding ground prediction, which is a significant shift from prior approaches. Strengths of the study include the comprehensive dataset, the integration of recent advances in deep learning, and a focus on actual breeding behavior rather than general locust presence. However, the study has some weaknesses. A major gap is the loss of data due to missing values in the remotely-sensed features, which limited the reliability of those models. Another limitation is the lack of ground-truth field validation at this stage, which the authors acknowledge and plan to pursue in partnership with the UN-FAO. Furthermore, while the study shows that multi-spectral images are powerful, over-reliance on them may overlook contextual environmental or ecological factors not captured in image data. A key lesson learned is that combining advanced geospatial deep learning models with high-resolution imagery can significantly improve early detection of breeding hotspots, but successful deployment requires addressing data availability, regional variability, and multi-stakeholder coordination.

3.3 Remote Sensing and Environmental Predictive Modeling

Lucile Marescot et al. [13] conducted a study to develop a reliable forecasting model that predicts the presence of desert locusts during the recession period by integrating remote sensing data and machine learning. Their main objective was to support preventive locust management across Western and Northern Africa, especially in the CLCPRO region, by generating accurate probability maps of locust presence at a spatial resolution of 1 km² and updating these forecasts every ten days. The study used a large historical dataset collected from 2003 to 2020 by ground surveyors using FAO's eLocust system. This dataset was combined with static variables such as ecoregions and soil sand content, and dynamic environmental predictors including NDVI, land surface temperature, and rainfall data from MODIS and CMORPH. The authors applied a random forest model and evaluated its performance through different time-based cross-validation

scenarios. Among 17 trained models, the best ones were selected for operational use based on predictive power AUC > 0.75 and sensitivity. The model was later tested in a real field survey in Mauritania in December 2023, which confirmed its ability to identify areas with a high likelihood of locust presence with a 77% success rate.

The strengths of this work include its innovative use of both static and dynamic environmental features, as well as its design for real-time application through a fully automated processing system. Unlike previous models, which often relied on static data or focused on long-term predictions, this model offers near-real-time forecasts that can directly support field operations and decision-making. Its operational capacity and web-based map interface make it especially useful for national locust control centers. One of the weaknesses is that the model does not consider different biological phases of the locust solitary or gregarious, and it does not fully address biases from non-random sampling in ground surveys. Also, spatial autocorrelation was acknowledged but not deeply integrated due to practical limitations. A key gap identified is the absence of behavioral or demographic locust data, which could further improve model accuracy. The study implies that combining machine learning with remote sensing in an automated system can significantly strengthen early warning tools. It also highlights the importance of integrating ecological knowledge and user feedback into the development of predictive models for effective locust management.

Diego Gómez et al. [14] conducted a study to identify desert locust breeding areas in Mauritania by exploring the link between soil moisture (SM) and the presence of hopper-stage locusts in their solitarious phase. The main objective was to determine whether satellite-derived SM data, specifically from the ESA Climate Change Initiative (CCI), could effectively replace or complement traditional vegetation and precipitation-based locust monitoring methods. The study applied species distribution modelling using two machine learning methods: Generalized Linear Model (GLM) and Random Forest (RF). They used field data from the FAO SWARMS database collected between 1985 and 2015 and combined it with 30 years of satellite-derived SM data. The researchers tested several time scenarios 16, 12, 8, and 6 days and used statistical indicators such as ROC-AUC, TSS, and Kappa to evaluate model performance.

The study found that RF significantly outperformed GLM in predicting hopper presence, especially when the model used minimum SM values over 6-day intervals. The best RF model

achieved an ROC-AUC of 0.94 and a TSS of 0.74, confirming a high predictive capacity. One of the main strengths of this work is the innovative use of SM as a key predictor, overcoming the limitations of rainfall estimation in arid zones. It also benefits from a long historical record and robust modelling techniques. However, the use of pseudoabsence data introduces some uncertainty, and the spatial resolution of the SM product 0.25° might limit finer-scale applications. A major gap is the lack of biological or behavioral locust data that could further improve accuracy. The key lesson from this study is that minimum SM values above 0.07 m^3/m^3 sustained over at least six days are strong indicators of suitable breeding conditions. This work demonstrates the potential of satellite-derived SM to improve desert locust early warning systems.

3.4 Operational Systems and Detection Techniques

Keith Cressman [15] conducted a comprehensive study to examine the role of remote sensing in enhancing early warning systems for desert locust management. The paper outlines the operational framework of the FAO's Desert Locust Information Service (DLIS), which monitors locust populations and environmental conditions across a recession area spanning over 16 million square kilometers. The objective of the study was to present how modern tools such as satellite-derived rainfall estimates, vegetation indices (like MODIS NDVI and EVI), greenness maps, and seasonal forecasts contribute to timely detection and forecasting of locust outbreaks. The author describes the multi-tiered monitoring structure, including field-level data collection using the eLocust2 system, national-level data processing with RAMSES GIS, and international-level forecasting using the SWARMS GIS system. Historical data, real-time satellite observations, and models like the DL Trajectory and DL Egg and Hopper Development Model were also employed to assess breeding conditions and migration.

One of the key strengths of this paper is its detailed account of how remote sensing technologies have been integrated into locust monitoring, making the system more efficient, precise, and responsive. The study highlights how remotely sensed data fills critical gaps in rainfall and vegetation observations, particularly in inaccessible or insecure regions. However, a major challenge identified is the limitation of current sensors like MODIS in detecting sparse vegetation, and the limited accuracy of rainfall estimates during winter seasons. The paper also discusses the reliance on expert knowledge due to incomplete ecological and biological data, and the difficulties in maintaining consistent monitoring in politically unstable or resource-limited countries. A

significant lesson learnt is that while technological advancements have greatly improved the system, timely action still depends on local capacity, political will, and financial support. The study emphasizes that early detection through remote sensing must be matched by timely decision-making to prevent outbreaks from developing into plagues.

Fentahun Abebaw Demeke [7] conducted a study aiming to assess desert locust infestation using Geographic Information Systems (GIS) and Remote Sensing (RS) technologies, focusing on Dire Dawa and the northern part of the Somali region in Ethiopia. The primary objective was to evaluate the relationship between locust infestations and environmental variables such as rainfall, vegetation, and elevation. The study also aimed to assess the extent of vegetation damage caused by locusts using MODIS EVI and Sentinel-1 SAR data and to predict infestation in areas not covered by ground surveys using geostatistical interpolation techniques, particularly kriging. The datasets used included MODIS EVI 250m, Sentinel-1 SAR 10m, Shuttle Radar Topographic Mission SRTM DEM, Global Precipitation Measurement GPM rainfall data 0.1° resolution, and ground survey data collected by the Ministry of Agriculture. Various software tools such as ArcGIS, SNAP Toolbox, and Microsoft Excel were used for data preprocessing, analysis, and visualization.

The results revealed that desert locust infestations were initially concentrated in the northern part of the study area during September and October 2019 and September 2020, characterized by low rainfall and sparse vegetation. By contrast, infestations spread to the southern parts in later months where vegetation and rainfall were more abundant. Vegetation damage was confirmed through both MODIS EVI and Sentinel-1 SAR data, with differences in reflectance and backscatter values between infested and non-infested areas. The kriging model yielded weak predictive power, with R^2 values of 0.24 in 2019 and 0.0084 in 2020, indicating that spatial interpolation was limited by the irregular distribution and inconsistent quality of survey data. One of the study's key strengths is its integrated approach, combining multi-source satellite data with ground observations to assess locust damage and spatial patterns. However, weaknesses include the lack of consistent meteorological data and the limitations of the spatial and temporal resolution of satellite imagery. The study also acknowledged difficulties in interpreting radar data and highlighted data quality issues from field surveys. A critical gap identified is the model's inability to produce reliable predictions due to data sparsity. The main lesson learned is that integrating ground data with

remote sensing can enhance locust damage assessment and environmental correlation analysis, but reliable prediction of infestation requires more accurate and evenly distributed ground observations.

Mebrahtu Chaklu Gebretensiae [8] conducted a study to design and implement an automated locust identification system using deep learning techniques, aiming to replace the traditional manual identification methods which rely heavily on field surveys and expert visual inspections. The main objective of the study was to develop a model capable of accurately classifying locusts from non-locust objects using image-based data. To accomplish this, Mebrahtu employed a deep neural network model, specifically ResNet50, selected for its capability to handle complex classification tasks with relatively high accuracy and computational efficiency. The method involved several key stages: image acquisition, preprocessing including resizing, grayscale conversion, and noise removal, segmentation using thresholding methods, feature extraction via the SURF algorithm, and training and classification through the ResNet50 architecture. The system was developed and tested in MATLAB, utilizing a dataset of 412 images, of which 330 images were used for training and 82 images for testing and validation.

The experimental results demonstrated that the proposed system achieved an accuracy of 95.2% in classifying locusts versus non-locusts. The strength of this approach lies in its ability to automate and streamline the identification process, potentially reducing dependence on human labor and improving responsiveness in locust surveillance operations. The use of image processing and deep learning also introduces the possibility of deployment in remote and inaccessible regions, especially when integrated with aerial drone imaging. However, the system has limitations, notably its restriction to still image classification without incorporating temporal or movement-based data, which are essential for tracking locust migration patterns. The study does not include environmental or spatial parameters that could enhance predictive capabilities. A significant gap identified is the absence of real-time detection or integration with remote sensing data for forecasting locust movement. The lesson drawn from this study is that deep neural networks, when properly trained and supported by suitable preprocessing techniques, can serve as an effective tool for the automatic identification of locusts.

Z. Shao et al. [14] conducted a study to monitor and predict the severity of desert locust plagues across parts of Asia and Africa by utilizing multisource remote sensing time-series data. Their

main objective was to use satellite-derived environmental indicators, such as NDVI, land surface temperature (LST), vegetation temperature condition index (VTCI), and temperature vegetation dryness index (TVDI), to evaluate the dynamics of locust infestation and its impact on croplands. They implemented a Hidden Markov Model (HMM) to process these time-series features and forecast locust severity at different time intervals. In addition to the broader regional analysis, they performed localized damage assessment using high-resolution hyperspectral data from the Zhuhai-1 (OHS) satellite for a case study area in Narok, Kenya. The study area covered 20 countries from 30°E to 90°E and 10°S to 40°N, with the predictive model trained and tested using sample points validated against FAO news bulletins and FROM-GLC land cover data. A total of 252 MODIS images MOD13A2 and MOD11A2 products and hyperspectral images from two dates February 23 and March 7, 2020 were used.

The results demonstrated that the HMM model achieved overall accuracies of 78%, 71%, 74%, and 72% for locust severity prediction in April, May, June, and July 2020, respectively. Compared to the ARIMA model, HMM showed superior performance due to its suitability for modeling time-dependent stochastic processes. The hyperspectral classification further revealed that 50 km² of high-density cropland and 15 km² of medium-density cropland were lost within the 2556.3 km² case area, signifying considerable locust-related damage. Strengths of the study include the integration of multisource data for both macro-scale forecasting and micro-scale damage assessment, as well as the innovative use of HMM with time-series satellite indices. Nevertheless, some limitations were present. The prediction was dependent on the availability and quality of MODIS data, which can be hindered by cloud cover and coarse resolution. The reliance on indirect indicators rather than direct locust detection is also a limitation. A gap identified is the lack of integration of soil moisture and meteorological parameters into the prediction model. The key lesson from this work is that a hybrid approach combining statistical models with rich remote sensing indices can significantly enhance the early warning and monitoring capabilities of locust plagues across different spatial and temporal scales.

3.5 Summary

The reviewed related works provide a wide-ranging foundation for understanding how desert locust prediction can be approached using machine learning, remote sensing, image-based detection, and ecological modeling. While the efforts show progress, several limitations are

recurrent across studies. Most of the models heavily rely on crowdsourced or manual data, which can be unreliable, inconsistent, or spatially biased. This is particularly true for studies such as PLAN by Maryam Tabar et al. [6], which showed significant accuracy drops when such data were removed. Many models do not integrate important migratory drivers such as high-altitude wind, and very few incorporate behavioral aspects of locust life cycles such as breeding, gregariousness, or solitary phases.

A major issue observed in region-specific implementations is the lack of validation in contexts like Ethiopia. For instance, studies by Fentahun Abebaw Demeke [7] reveal the limitations of kriging-based prediction due to sparse and uneven field data. Similarly, image-based models like Mebrahtu's [8] system, though efficient in classification, lack spatial and temporal forecasting power. Even operational frameworks such as DLIS face practical limitations due to seasonal sensor performance and regional instability. Models that performed well in other countries, such as Ibrahim Salihu et al. [12] and Lucile Marescot et al. [13] Random Forest approach, still need adaptation to Ethiopian landscapes, climates, and survey practices.

Overall, these findings underscore that while machine learning and remote sensing present promising tools, they must be customized with regional specificity, ecological realism, and integration of key environmental variables. This thesis, therefore, aims to develop a predictive model that avoids over-reliance on crowdsourced or uneven ground survey data. Instead, it seeks to use high-resolution remote sensing indicators and temporal environmental patterns, with a design tailored for the Ethiopian locust migration context. The model will aim to support early warning systems and operational decisions by addressing data sparsity, lack of wind integration, and absence of behavioral modeling found in existing works.

Chapter 4: Proposed Model

4.1. Introduction

This chapter presents the proposed Multi-Resolution Heterogeneous Vision Transformer (MRH-ViT) model for predicting desert locust movement in Ethiopia using remote sensing data and environmental indicators. The proposed model is based on the Vision Transformer (ViT) architecture [202], specifically adapted for spatiotemporal prediction tasks involving multi-spectral satellite imagery. The proposed approach addresses the limitations identified in the related work by developing a model that can process high-dimensional environmental data without relying heavily on crowdsourced or inconsistent field reports.

The MRH-ViT architecture extends the conventional Vision Transformer framework by introducing three novel components: multi-resolution patch processing for capturing features at different spatial scales, feature group encoding for handling heterogeneous remote sensing data, and physics-informed attention that incorporates ecological knowledge of locust behavior patterns. This chapter provides a comprehensive description of each component, including mathematical formulations and algorithmic implementations that enable the model to learn complex spatiotemporal patterns associated with desert locust habitat suitability and movement dynamics.

The proposed approach processes 59-band multi-spectral satellite imagery to generate locust presence predictions while maintaining interpretability through attention visualization and multi-scale feature analysis. The proposed model is specifically designed to handle the heterogeneous nature of environmental monitoring data, where different spectral bands represent distinct physical phenomena that influence locust behavior through various mechanisms.

4.2. Proposed Model Architecture

The Multi-Resolution Heterogeneous Vision Transformer (MRH-ViT) is designed to process multi-spectral satellite imagery with 59 environmental bands and generate binary classification predictions locust presence predictions. The architecture consists of six main components: multi-resolution patch embedding, feature group encoding, physics-informed multi-head self-attention layer, encoder stack, global pooling, and classification head. Figure 4.1 illustrates the complete architecture of the proposed model.

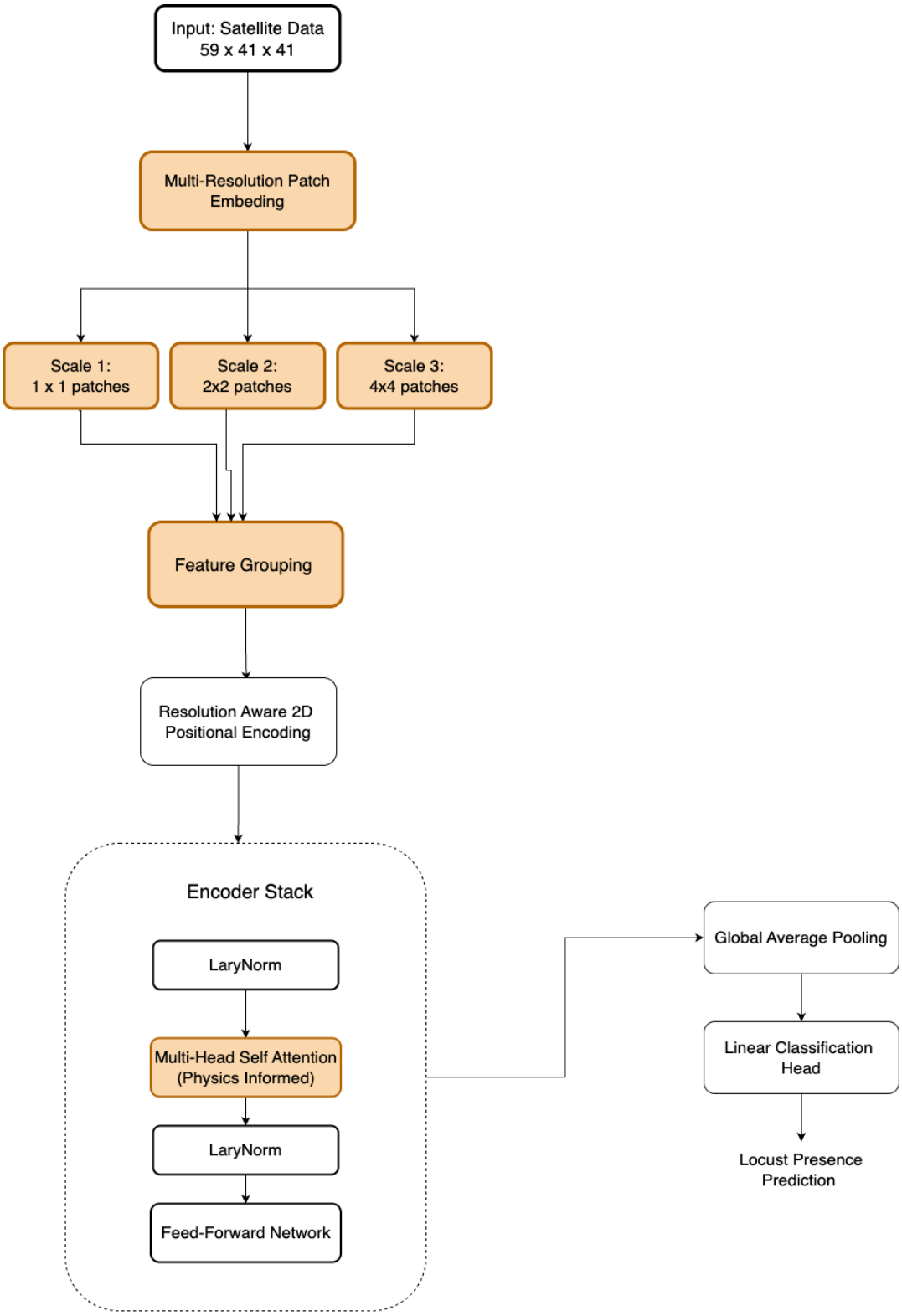


Figure 4-1 *The proposed model architecture*

4.2.1 Multi-Resolution Patch Embedding

The multi-resolution patch embedding component represents the first novel contribution of the proposed architecture, enabling the simultaneous processing of environmental data at multiple spatial scales. This approach addresses the multi-scale nature of desert locust phenomena, where individual insects, small groups, and large swarms require different spatial analysis resolutions.

Given an input multi-spectral image $\mathrm{X} \epsilon \mathbb{R}^{B \times C \times H \times W}$ where B is the batch size, $\mathrm{C} = 59$ is the number of spectral channels, and H=W=41 are the spatial dimensions, the multi-resolution patch extraction process operates with different patch sizes $P = \{1,2,3\}$ to capture varying spatial scales.

For each patch size $\mathrm{p_i} \epsilon \boldsymbol{P}$ the patch extraction process is formulated as:

$$X_{\text{patches}}^{(i)} = \text{Unfold}(X, p_i, p_i) \in R^{B \times N_i \times (C\, p_i^2)} \qquad (10)$$

Where $\mathrm{N_i} = \lceil H/\mathrm{p_i} \rceil \times \lceil W/\mathrm{p_i} \rceil$, represents the number of patches for resolution i and the unfold operation extracts non-overlapping patches on size $p_i \times p_i$

Each patch resolution is processed through a dedicated embedding layer:

$$E^{(i)} = \text{Linear}_i\left(X_{\text{patches}}^{(i)}\right) + b_i \in R^{B \times N_i \times D} \qquad (11)$$

where $\text{Linear}_i \in R^{(C \times p_i^2) \times D}$ and $b_i \in R^D$ are learnable parameters specific to resolution i, and $\mathrm{D} = 256$ is the embedding dimension.

4.2.2 Resolution-aware 2D Positional Encoding

The multi-resolution spatial positional encoding component addresses the challenge of maintaining spatial relationships across different patch resolutions while ensuring that patches representing the same geographical location receive consistent positional information regardless of their resolution level. Unlike traditional sequence-based positional encoding, this approach requires 2D spatial encoding that accounts for the geographical nature of satellite imagery data.

For each resolution level i with patch size p_i, the spatial coordinates of each patch are first computed relative to the original image dimensions. Given a patch at grid position (h, w) in the patch grid, the actual spatial coordinates are scaled by the patch size to maintain geographical consistency:

$$\mathrm{x_{coord}} = \mathrm{w} \times p_{\mathrm{i}} \qquad (12)$$

$$y_{\mathrm{coord}} = h \times p_i \qquad (13)$$

The 2D sinusoidal positional encoding is then computed using these scaled coordinates to ensure that patches covering the same geographical area across different resolutions receive related positional encodings. The encoding alternates between sine and cosine functions across the embedding dimensions:

$$\mathrm{PE(x, y, 2j)} = \sin\left(\left(\mathrm{x_{coord}} \cdot \alpha_{\mathrm{x}} + \mathrm{y_{coord}} \cdot \alpha_{\mathrm{y}}\right)/\left(10000^{(2\mathrm{j}/\mathrm{D})}\right)\right) \qquad (14)$$

$$\mathrm{PE(x, y, 2j + 1)} = \cos\left(\left(\mathrm{x_{coord}} \cdot \alpha_{\mathrm{x}} + \mathrm{y_{coord}} \cdot \alpha_{\mathrm{y}}\right)/\left(10000^{(2\mathrm{j}/\mathrm{D})}\right)\right) \qquad (15)$$

where:

- $j \in [0, D/2 - 1]$ represents the dimension index
- $D = 256$ is the embedding dimension
- α_{x} and α_{y} are learnable scaling parameters initialized to 1.0
- 10000 is the base frequency parameter following the original Transformer design

The complete positional encoding matrix for resolution i is constructed by computing the encoding for each patch position and stacking them into a matrix:

$$P^{(i)} \in R^{N_i \times D} \qquad (16)$$

where $\mathrm{N_i} = \lceil \mathrm{H/p_i} \rceil \times \lceil \mathrm{W/p_i} \rceil$ represents the total number of patches for resolution i.

The positional encodings are added to the corresponding patch embeddings to incorporate spatial information:

$$Z_{\mathrm{pos}}^{(i)} = E^{(i)} + P^{(i)} \qquad (17)$$

This encoding scheme ensures that:

- Spatial Consistency: Patches at the same geographical location across different resolutions receive related positional encodings

- Resolution Awareness: The scaling by patch size maintains the spatial relationships while accounting for different granularities
- Learnable Adaptation: The parameters α_x and α_y allow the model to adapt the frequency scaling to the specific characteristics of the environmental data

The sinusoidal formulation provides several advantages over learned positional embeddings, including the ability to extrapolate to sequence lengths not seen during training and the mathematical properties that enable the model to learn relative position relationships effectively.

To distinguish between different resolutions, resolution type embeddings are incorporated:

$$R^{(i)} = \text{Embedding}(i) \in R^D \qquad (18)$$

$$Z_{res}^{(i)} = Z_{pos}^{(i)} + R^{(i)} \qquad (19)$$

The final multi-resolution embedding concatenates all resolution representations:

$$Z_0 = \text{Concat}\left(Z_{res}^{(1)}, Z_{res}^{(2)}, Z_{res}^{(3)}\right) \in R^{B\times N_{total}\times D} \quad (20)$$

where $N_{total} = \sum_{i=1}^{3} N_i$ represents the total number of patches across all resolutions.

Algorithm 4-1 Multi-Resolution Patch Embedding

START
Input: Multi-spectral image $X \in R^{(B\times 59\times 41\times 41)}$
Output: Multi-resolution embeddings $Z_0 \in R^{(B\times N_{\text{total}}\times 256)}$
1. Initialize patch_sizes = [1, 2, 4]
2. Initialize embeddings_list = []
3. For each patch_size $p_i$in patch_sizes:
   a. Extract patches: patches_$i \leftarrow Unfold(X, p_i, p_i)$
   b. Apply linear embedding: $E_i \leftarrow Linear_i(patches_i)$
   c. Add positional encoding: $Z_{\text{posi}} \leftarrow E_i + P_i$
   d. Add resolution embedding: $Z_{\text{resi}} \leftarrow Z_p os_i + R_i$
   e. Append Z_{resi} to embeddings_list
4. Concatenate all embeddings: $Z_0 \leftarrow Concat(embeddings_list)$
5. Return Z_0
END

4.2.3 Feature Group Encoding

The feature group encoding component addresses the heterogeneous nature of remote sensing data by processing different types of environmental variables through specialized pathways. This

approach recognizes that vegetation indices, climate variables, soil properties, and other environmental factors that exhibit distinct characteristics that benefit from tailored processing.

The 59 spectral channels are organized into six feature groups:

- **Vegetation Group:** $G_{\text{veg}} = \{\text{NDVI, EVI, NDWI, TCI, VHI}\}$
- **Climate Group:** $G_{\text{clim}} = \{\text{LST_Day, precipitation, ET, TVDI}\}$
- **Wind Group:** $G_{\text{wind}} = \{\text{u_component, v_component}\}$
- **Soil Group:** $G_{\text{soil}} = \{\text{sm_surface, sand_content, soil_texture}\}$
- **Topography Group:** $G_{\text{topo}} = \{\text{elevation, slope, aspect}\}$
- **Static Group:** $G_{\text{static}} = \{\text{landcover}\}$

Each feature group is processed through a dedicated encoder:

$$E_{group}^{(g)} = \text{GroupEncoder}_g(Z_0) \qquad (21)$$

where each group encoder is implemented as:

$$\text{GroupEncoder}_g(x) = \text{Dropout}\left(\text{LayerNorm}\left(\text{GELU}\left(\text{Linear}_g(x)\right)\right)\right) \qquad (22)$$

Group type embeddings distinguish between different feature categories:

$$G_{emb}^{(g)} = \text{GroupEmbedding}(g) \in R^D \qquad (23)$$

The group-enhanced representation is computed as:

$$Z_{group}^{(g)} = E_{group}^{(g)} + G_{emb}^{(g)} \qquad (24)$$

The final group-encoded representation combines all group outputs.

$$Z_{grouped} = \frac{1}{|G|}\sum_{g=1}^{|G|} Z_{group}^{(g)} \qquad (25)$$

Where $|G| = 6$ is the total number of feature groups.

Algorithm 4-2 Feature Group Encoding

START

Input: Multi-resolution embeddings $\text{Z}_0 \in \text{R}^{(\text{B}\times\text{N}_{\text{total}}\times 256)}$

Output: Group-encoded representation $Z_{grouped} \in R^{(B \times N_{total} \times 256)}$

1. Define feature groups:
   a. $G_{veg} = NDVI, EVI, NDWI, TCI, VHI$
   b. $G_{clim} = LST_{D}ay, precipitation, ET, TVDI$
   c. $G_{wind} = u_{c}omponent, v_{c}omponent$
   d. $G_{soil} = sm_{s}urface, sand_{c}ontent, soil_{t}exture$
   e. $G_{topo} = elevation, slope, aspect$
   f. $G_{static} = landcover$
2. Initialize group_outputs = []
3. For each group g in $[G_{veg}, G_{clim}, G_{wind}, G_{soil}, G_{topo}, G_{static}]$:
   a. Extract group channels: $Z_{group} \leftarrow Z_{0}[group_{indices}[g]]$
   b. Apply group encoder:
      $E_{group}^{(g)} \leftarrow GroupEncoder_{g}(Z_{group})$
   c. Get group embedding: $G_{emb}^{(g)} \leftarrow GroupEmbedding(g)$
   d. Add group embedding: $Z_{group}^{(g)} \leftarrow E_{group}^{(g)} + G_{emb}^{(g)}$
   e. Append $Z_{group}^{(g)} \rightarrow group_{outputs}$
4. Compute grouped representation: $Z_{grouped} \leftarrow (1/|G|) \times \Sigma(g = 1\ to\ |G|) Z_{group}^{(g)}$
5. Return $Z_{grouped}$

END

4.2.4 Physics-Informed Multi-Head Self-Attention

The physics-informed attention mechanism represents the third novel contribution, incorporating ecological knowledge about desert locust behavior into the attention computation. This component modifies the standard multi-head attention by introducing physics-based biases that reflect known locust movement patterns and environmental preferences. This approach aligns with methodologies presented in prior studies, which integrate attention mechanisms into neural networks to adaptively focus on regions exhibiting complex behaviors in physical systems. For instance, Rodriguez-Torrado et al. [185] introduced a physics-informed attention-based neural network that effectively captures non-local dependencies and discontinuities in non-linear partial differential equations. Similarly, Liao et al. [187] proposed a self-attention assisted physics-

informed neural network framework for remaining useful life prediction, demonstrating improved accuracy and interpretability by embedding domain-specific knowledge into the attention computation.

By adopting similar principles, our proposed model aims to capture non-local dependencies and behavioral discontinuities in locust movement, thereby enhancing its performance in simulating environmentally influenced swarm dynamics.

The physics-informed attention extends the standard attention mechanism:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B_{physics}\right)V \quad (26)$$

Where $\text{B}_{\text{physics}}$ represents the physics-informed bias matrix that incorporates ecological knowledge. The physics bias is computed as:

$$B_{physics} = \alpha_{wind} \cdot B_{wind} + \alpha_{elev} \cdot B_{elevation} + \alpha_{temp} \cdot B_{temperature} \quad (27)$$

where α_{wind}, α_{elev}, and α_{temp} are learnable weighting parameters initialized to 0.1, and the individual bias components are:

$$B_{wind} = -\sqrt{(y_{coords} - y_{center})^2 + (x_{coords} - x_{center})^2} \quad (28)$$

$$B_{elevation} = \text{ElevationBias}(\text{spatial_positions}) \quad (29)$$

$$B_{temperature} = \text{TemperatureBias}(\text{spatial_positions}) \quad (30)$$

The complete physics-informed multi-head attention is formulated as:

$$\text{MultiHead}_{physics}(Q, K, V) = \text{Concat}(head_1, \dots, head_h)W^O \quad (31)$$

where each head incorporates physics biases:

$$head_i = \text{Attention}_{physics}\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (32)$$

Algorithm 4-3 Physics-Informed Attention

START

Input: Query Q, Key K, Value $V \in R^{(B\times N\times D)}$

Output: Attention output $A \in R^{(B\times N\times D)}$

1. Compute attention scores: $scores \leftarrow (Q \cdot K^T)/\sqrt{d_k}$
2. Compute physics bias:

a. Create spatial coordinates: $y_{\text{coords}}, x_{\text{coords}}$

b. Compute distance bias: $B_{\text{spatial}} \leftarrow -\sqrt{((y_{\text{coords}} - center)^2 + (x_{\text{coords}} - center)^2)}$

c. Apply physics weights: $B_{\text{physics}} \leftarrow \alpha_{\text{wind}} \cdot B_{\text{spatial}}$

3. Add physics bias: $scores \leftarrow scores + B_{\text{physics}}$
4. Apply softmax: $attn_{\text{weights}} \leftarrow Softmax(scores)$
5. Compute output: $A \leftarrow attn_{\text{weights}} \cdot V$
6. Return A

END

4.2.5 Encoder Stack

The proposed model encoder stack consists of four layers, each containing the physics-informed attention mechanism, feed-forward networks, and normalization layers with residual connections. This depth provides sufficient capacity to capture complex spatiotemporal relationships while maintaining computational efficiency.

Each encoder layer processes the input through the following sequence:

$$Z_l = \text{EncoderLayer}_l(Z_{l-1}) \text{ for } l = 1,2,3,4 \qquad (33)$$

The encoder layer transformation incorporates physics-informed attention:

$$Z_{l,1} = \text{LayerNorm}\left(Z_{l-1} + \text{MultiHead}_{physics}(Z_{l-1})\right) \qquad (34)$$

$$Z_l = \text{LayerNorm}\left(Z_{l,1} + \text{FFN}(Z_{l,1})\right) \qquad (35)$$

The feed-forward network maintains the standard transformer design:

$$\text{FFN}(x) = \text{Linear}_2\left(\text{Dropout}\left(\text{GELU}(\text{Linear}_1(x))\right)\right) \qquad (36)$$

where the hidden dimension is $D_{ff} = 4 \times D = 1024$.

4.2.6 Global Average Pooling and Linear Classification Head

The global pooling component aggregates the spatial representations into a single vector for binary classification. Mean pooling is applied across the spatial dimension:

$$z_{global} = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} Z_4^{(i)} \qquad (37)$$

where $Z_4^{(i)}$ represents the i-th spatial location in the final encoder output.

The classification head transforms the global representation to a binary prediction:

$$y_{pred} = \text{Classifier}(z_{global}) \quad (38)$$

The classifier consists of a multi-layer architecture with dropout regularization:

$$\text{Classifier}(x) = \text{Linear}_3\left(\text{Dropout}\left(\text{GELU}\left(\text{Linear}_2(\text{LayerNorm}(x))\right)\right)\right) \quad (39)$$

where:

- $\text{Linear}_2: R^D \to R^{D/2}$
- $\text{Linear}_3: R^{D/2} \to R^1$

The model is trained using Binary Cross-Entropy with Logits loss, suitable for binary classification tasks:

$$\mathcal{L} = \frac{1}{B}\sum_{b=1}^{B}\left[y_b \log\left(\sigma(y_{pred,b})\right) + (1 - y_b)\log\left(1 - \sigma(y_{pred,b})\right)\right] \quad (40)$$

Where y_b is the ground truth label, $y_{pred,b}$ is the predicted logit, and σ is the sigmoid function.

Algorithm 4-4 MRH-Vit Forward Pass

START
Input: Multi-spectral image $X \in R^{(B\times 59\times 41\times 41)}$
Output: Binary prediction $y_{pred} \in R^{(B\times 1)}$
1. Multi-resolution patch embedding: $Z_0 \leftarrow MultiResPatchEmbed(X)$
2. Feature group encoding: $Z_{grouped} \leftarrow FeatureGroupEncoder(Z_0)$
3. For layer $l = 1$ to 4:
   a. Physics-informed attention: $Z_l \leftarrow LayerNorm(Z_l - 1 + PhysicsAttention(Z_l - 1))$
   b. Feed-forward: $Z_l \leftarrow LayerNorm(Z_l, 1 + FFN(Z_l, 1))$
4. Global pooling: $z_{global} \leftarrow MeanPool(Z_4)$
5. Classification: $y_{pred} \leftarrow Classifier(z_{global})$
6. Return y_{pred}
END

The proposed MRH-ViT architecture processes multi-spectral data through a pipeline that preserves spectral information while capturing spatial relationships at multiple scales. The 59-channel input represents diverse environmental variables including vegetation indices (NDVI,

EVI), temperature measurements (LST), precipitation data, soil moisture content, wind components, elevation data, and land cover classifications.

The multi-resolution approach ensures that fine-scale locust activity (individual insects or small groups) is captured through 1×1 patches, medium-scale aggregations are detected via 2×2 patches, and large-scale swarm movements are identified using 4×4 patches. This multi-scale processing is particularly crucial for locust monitoring, as different life stages and behavioral phases exhibit distinct spatial patterns.

The feature group encoding further enhances the processing by recognizing that different environmental variables have distinct characteristics and relationships. Vegetation indices exhibit temporal patterns related to phenology, climate variables show seasonal and diurnal cycles, soil properties remain relatively stable, and topographic features are static. By processing these groups separately before integration, the model can learn specialized representations for each environmental domain.

The physics-informed attention mechanism incorporates ecological knowledge by biasing attention weights based on known locust behavior patterns. Locusts tend to aggregate in areas with favorable environmental conditions and migrate following wind patterns and elevation gradients. The learned bias parameters allow the model to focus attention on environmentally relevant spatial relationships, improving prediction accuracy and interpretability.

The proposed MRH-ViT architecture provides several advantages for desert locust prediction applications. The multi-resolution processing captures patterns at different spatial scales relevant to locust ecology, the feature group encoding leverages the heterogeneous nature of remote sensing data, and the physics-informed attention incorporates domain knowledge to improve both accuracy and interpretability. The transformer-based architecture enables long-range spatial dependency modeling, crucial for understanding how environmental conditions influence locust movement across landscapes.

Chapter 5: Experimentation and Evaluation

5.1. Introduction

This chapter presents the experimental methodology, dataset preparation, and evaluation results for the proposed MRH-ViT model. The experimental work focuses on validating the effectiveness of the proposed approach for predicting desert locust presence in Ethiopia using multi-spectral satellite imagery and environmental indicators. The chapter describes the data collection process using Google Earth Engine, the experimental setup, dataset preparation procedures, and comprehensive evaluation of the model's performance across multiple metrics.

The evaluation encompasses quantitative metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The experimental design aims to provide a thorough assessment of the model's capabilities and limitations in real-world desert locust monitoring scenarios.

5.2. Dataset Collection and Preparation

The dataset for this study was compiled using the Food and Agriculture Organization (FAO) Desert Locust Information Service reports for Ethiopia spanning the period from 2015 to 2023 with a total of 18,871 report out of which 12,056 are presence and 6,815 are absence. These reports contain both presence and absence records of desert locust observations collected through field surveys and expert assessments. The FAO database provides geographical coordinates, timestamps, and locust presence indicators for each observation point, forming the foundation for supervised learning.

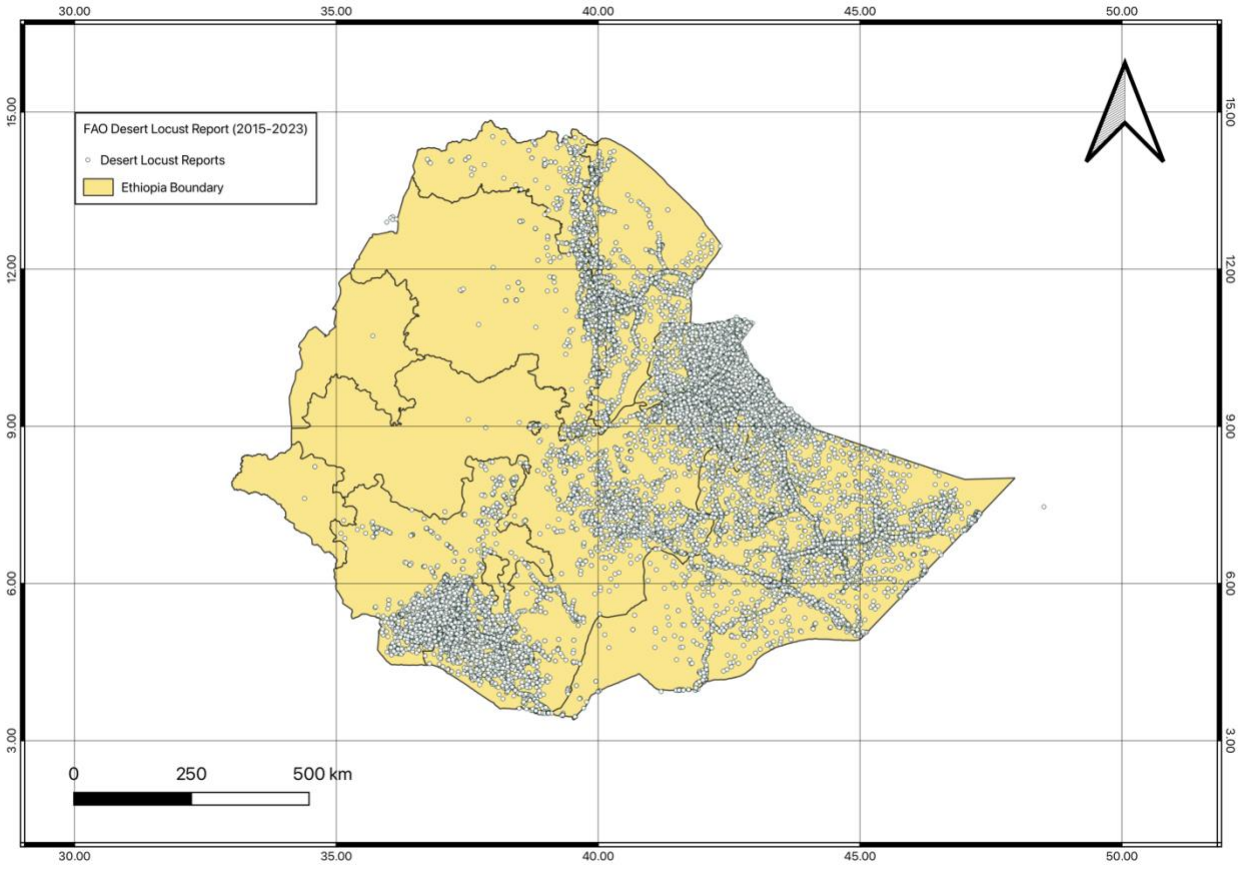


Figure 5-1 *Visual representation of Desert Locust Report from FAO (2015-2023)*

The dataset collection process utilized Google Earth Engine (GEE) as the primary platform for extracting multi-spectral satellite imagery and environmental data corresponding to each locust observation point. GEE provides access to a comprehensive catalog of Earth observation datasets, including MODIS vegetation indices, Landsat imagery, climate data, and soil information, making it an ideal platform for environmental monitoring applications.

The data extraction pipeline implemented several key steps to ensure data quality and consistency. First, the FAO locust observation points were preprocessed to extract valid geographical coordinates and observation dates. Each observation point was then used as the center for extracting a 41×41-pixel window of environmental data at 250-meter spatial resolution, providing coverage of approximately 10×10 kilometers around each observation point.

The environmental variables extracted for each observation point included static and dynamic components. Static variables comprised elevation data from the Shuttle Radar Topography Mission (SRTM), soil texture information from OpenLandMap, sand content at different depths

from SoilGrids, slope and aspect derived from terrain analysis, and land cover classification from MODIS. These variables provide contextual information about the physical characteristics of each location that may influence locust habitat suitability.

Dynamic environmental variables were extracted for three temporal periods: 30, 60, and 90 days prior to each observation date. This temporal approach captures the environmental conditions that may have influenced locust development, breeding, and movement leading up to each observation date. The dynamic variables included NDVI and EVI from MODIS MOD13Q1, LST from MODIS MOD11A2, precipitation data from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), soil moisture from NASA's SMAP, actual evapotranspiration from MODIS MOD16A2, and wind components from ERA5 reanalysis data.

To address potential gaps in satellite data coverage due to cloud cover or sensor limitations, the extraction pipeline incorporated fallback mechanisms and alternative data sources. For instance, when MODIS evapotranspiration data was unavailable, the system utilized TerraClimate data as an alternative source. Similarly, for wind data, the pipeline implemented fallbacks from European Centre for Medium-range Weather Forecasts Reanalysis (ERA5) to National Centers for Environmental Prediction (NCEP) or National Center for Atmospheric Research (NCAR) when necessary.

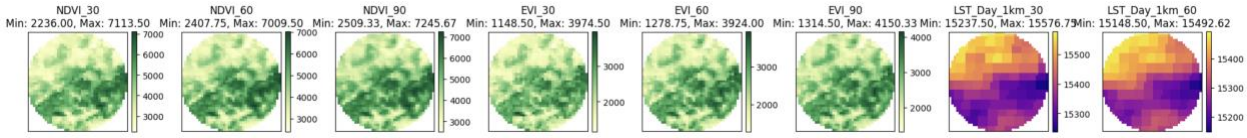


Figure 5-2 *Sample NDVI, EVI and LST band visualization*

The data extraction process also computed additional vegetation and drought indices to enhance the information content of the dataset. These included the Temperature Condition Index (TCI), Vegetation Health Index (VHI), Normalized Difference Water Index (NDWI), and Temperature Vegetation Dryness Index (TVDI). These indices provide supplementary information about vegetation stress, water availability, and drought conditions that are relevant to locust habitat assessment. Each of these dynamic and static environmental variables are then stacked into a single 41x41 geotiff image and saved to a google drive. Table 5.2.1 summarizes the all the environmental factors that are used to construct the GeoTiff image.

Table 5-1 *Summary of input features and their spatiotemporal representation*

Variable Category	**Specific Variables**	**Data Source(s)**	**Time Lag / Nature**
Vegetation Indices	Normalized Difference Vegetation Index (NDVI)	MODIS/061/MOD13Q1	30, 60, 90 days (mean)
	Enhanced Vegetation Index (EVI)	MODIS/061/MOD13Q1	30, 60, 90 days (mean)
	Vegetation Health Index (VHI)	Derived from LST and NDVI	30, 60, 90 days (mean)
Moisture Indicators	Normalized Difference Water Index (NDWI)	MODIS/061/MOD09GA (fallback: MODIS/MOD13Q1)	30, 60, 90 days (mean)
	Soil Moisture (sm_surface)	NASA/SMAP/SPL4SMGP/007	30, 60, 90 days (mean)
	Precipitation	UCSB-CHG/CHIRPS/DAILY	30, 60, 90 days (sum)
Temperature Data	Land Surface Temperature (LST_Day_1km)	MODIS/061/MOD11A2	30, 60, 90 days (mean)
	Temperature Condition Index (TCI)	Derived from LST	30, 60, 90 days (mean)
	Temperature Vegetation Dryness Index (TVDI)	Derived from LST and NDVI	30, 60, 90 days (mean)

Wind Components	u-component of wind (u_component_of_wind_10m)	ECMWF/ERA5/DAILY (fallback: NCEP_DOE_II)	30, 60, 90 days (mean)
	v-component of wind (v_component_of_wind_10m)	ECMWF/ERA5/DAILY (fallback: NCEP_DOE_II)	30, 60, 90 days (mean)
Evapotranspiration	Actual Evapotranspiration (ET)	MODIS/061/MOD16A2 (fallback: IDAHO_EPSCOR/TERRACLIMATE)	30, 60, 90 days (sum)
	Potential Evapotranspiration (pet)	IDAHO_EPSCOR/TERRACLIMATE	30, 60, 90 days (mean)
Contextual Data	Locust Presence/Absence Counts (41x41 grid)	FAO_REPORT_ASSET_ID(the actual FAO report)	30, 60, 90 days (presence/absence counts)
Static Variables	Soil Texture (b0)	OpenLandMap/SOL/SOL_TEXTURE-CLASS_USDA-TT_M/v02	Static
	Soil Sand Content (0-5cm, 5-15cm, 15-30cm mean)	projects/soilgrids-isric/sand_mean	Static
	Elevation	USGS/SRTMGL1_003	Static
	Slope	Derived from Elevation	Static
	Aspect	Derived from Elevation	Static

	Land Cover (LC_Type1)	MODIS/006/MCD12Q1/2019_01_01	Static

The preparation of our comprehensive desert locust dataset represented one of the most technically demanding and resource-intensive aspects of this research, requiring the development of sophisticated data extraction pipelines to overcome numerous infrastructural and computational challenges. Working with multi-temporal, multi-spectral satellite imagery spanning a vast geographical region (30°E–90°E, 10°S–40°N) from multiple satellite platforms (MODIS, Landsat, Sentinel-2) necessitated the creation of a custom Python-based extraction tool, available on our GitHub repository[2] including the dataset, is capable of handling the complex georeferencing, temporal alignment, and spectral band integration required for our 59-channel input format while providing flexible filtering capabilities by date ranges and specific countries to facilitate targeted data extraction for different research applications.

The extraction process was severely hampered by persistent power outages and intermittent internet connectivity issues that frequently interrupted multi-hour download sessions, forcing us to implement a robust cronjob-based automation system with comprehensive progress tracking, automatic restart capabilities, and detailed logging mechanisms to ensure data integrity and resume interrupted extractions. Each individual GeoTIFF image required 1-2 minutes of processing time for geometric correction, spectral calibration, and format standardization, resulting in 120-hour extraction period. The modular design of our extraction tool, with its built-in date and country filtering capabilities, enables other researchers to easily replicate our methodology and extend the concept to different geographical regions, temporal periods, or ecological applications beyond desert locust prediction, thereby facilitating broader adoption and advancement of remote sensing-based ecological monitoring systems

The dataset preparation process involved several steps to ensure data quality and model training effectiveness for the Multi-Resolution Heterogeneous Vision Transformer (MRH-ViT) model. The collected multi-spectral images were organized into a balanced dataset to address the inherent

[2] https://github.com/rabira-hierpa/LAMP

class imbalance observed in the original FAO reports, where presence samples (12,056) significantly outnumbered absence samples (6,815).

To ensure robust model training while maintaining the natural distribution of locust observations, we preserved the original class imbalance present in the FAO dataset rather than artificially balancing the classes. This resulted in a final dataset comprising 18,871 samples, with a class distribution of 8,479 presence samples and 4,730 absence samples with a ratio of 1.79:1, totaling 13,209 samples used for training after quality control procedures. This approach maintains the realistic operational conditions where locust presence events are naturally more frequent in the observational data, providing a more representative training environment for real-world deployment, while acknowledging that class imbalance requires careful consideration in model evaluation metrics as demonstrated by Tabar et al. [6] in their PLAN model.

Data preprocessing began with comprehensive quality control procedures to identify and handle missing or corrupted satellite imagery data. The preprocessing pipeline implemented robust error handling mechanisms to detect Not a Number (NaN) values, infinite values, and physical outliers in the environmental bands. Following the approach established by Yusuf et al. [12] in their Prithvi-LB model, missing values were imputed using temporal interpolation where possible, while outliers beyond reasonable physical ranges were clipped to prevent adverse effects on model training.

The normalization strategy employed a standardized approach where each environmental band was normalized independently using z-score normalization, subtracting the mean and dividing by the standard deviation. This normalization ensures that all 59 environmental variables contribute equally to the learning process, preventing variables with larger scales, such as elevation or temperature, from dominating those with smaller ranges such as vegetation indices. This approach aligns with best practices established in remote sensing applications for deep learning models.

The complete dataset was partitioned using a stratified random split with a 70:15:15 ratio, resulting in 13,209 training samples, 2,830 validation samples, and 2,832 test samples. This partitioning strategy maintains the balanced class distribution across all data splits while providing adequate data for model training and unbiased performance evaluation. The stratified approach ensures that each partition contains representative samples from both presence and absence classes, critical for

reliable model evaluation. Table 5.3 summarizes the final dataset composition and characteristics following the preprocessing procedures.

Table 5-2 *Dataset Composition and Characteristics*

Partition	**Total Samples**	**Presence Samples**	**Absence Samples**	**Class Balance**	**Percentage of Total**
Training	13,209	8,479	4,730	1.79:1	70.0%
Validation	2,830	1,816	1,014	1.79:1	15.0%
Test	2,832	1,819	1,013	1.80:1	15.0%
Total	18,871	12,114	6,757	1.79:1	100.0%

The final preprocessed dataset consisted of 59 environmental bands for each 41×41-pixel sample, representing a comprehensive environmental characterization required for desert locust habitat assessment. This multi-dimensional representation captures the complex spatio-temporal environmental factors that influence desert locust presence and movement patterns, incorporating both static terrain features and dynamic environmental conditions as established in the literature.

5.3. Experimental System Environment

The experimental work was conducted using Google Colaboratory (Colab) Pro+ with high-performance GPU acceleration to leverage computational resources necessary for training the proposed model. The computational environment featured an NVIDIA A100-SXM4-40GB GPU with 42.5 GB of GPU memory, providing substantial computational power for processing large-scale multi-spectral satellite imagery datasets and complex transformer architectures.

The system configuration included Python 3.10.12 as the primary programming environment, with PyTorch 2.0.1+cu118 serving as the deep learning framework for implementing the MRH-ViT architecture. The PyTorch installation included CUDA 11.8 support for optimal GPU utilization and mixed-precision training capabilities. Additional critical libraries included NumPy 1.25.2 for numerical computations, scikit-learn 1.3.0 for evaluation metrics, Rasterio 1.3.8 for geospatial data processing, and Matplotlib 3.7.2 with Seaborn 0.12.2 for comprehensive visualization capabilities.

Memory optimization techniques were implemented to handle the large satellite imagery datasets efficiently within the available computational constraints. The system utilized automatic mixed

precision (AMP) training with gradient scaling to reduce memory consumption while maintaining numerical stability. Data loading optimization included memory-mapped file access and efficient caching mechanisms to minimize I/O bottlenecks during training iterations.

The experimental setup incorporated advanced training optimizations including gradient clipping with a maximum norm of 1.0 to prevent gradient explosion, particularly important for transformer architectures. Batch size optimization was performed based on GPU memory constraints, resulting in an optimal batch size of 8 samples per iteration. This configuration balanced training efficiency with memory utilization, enabling stable training of the 950,285-parameter MRH-ViT model.

5.4. Experiment and Discussion of Results

The MRH-ViT model was trained using an optimized configuration designed to leverage the multi-scale nature of desert locust habitat patterns and the heterogeneous characteristics of environmental remote sensing data. The model architecture incorporated 950,285 total trainable parameters, representing a carefully balanced design that provides sufficient representational capacity while maintaining computational efficiency for operational deployment.

The training process employed the AdamW optimizer with an initial learning rate of 1e-4 and weight decay of 1e-4. This optimizer choice follows recent best practices in transformer training, as demonstrated by Dosovitskiy et al. [212] in their Vision Transformer work. The learning rate scheduling utilized a Cosine Annealing strategy that gradually reduced the learning rate from the initial value to a minimum of 1e-6 over the training duration, promoting fine-grained convergence in the later stages of training.

The loss function employed Binary Cross-Entropy with Logits (BCEWithLogitsLoss), which combines sigmoid activation with cross-entropy loss in a single operation for improved numerical stability. This approach is particularly effective for binary classification tasks with potential class imbalance, as encountered in ecological prediction problems.

Early stopping was implemented with a patience of 15 epochs, monitoring validation F1-score as the primary metric for convergence assessment. This choice prioritizes balanced performance between precision and recall, which is crucial for practical applications in locust management where both false positives and false negatives have operational implications. Model checkpointing

preserved the best-performing model state based on validation F1-score throughout the training process.

Data augmentation strategies were intentionally limited to preserve the physical validity of environmental measurements, following the principle that satellite imagery and environmental data should maintain their absolute values for meaningful interpretation. Instead, the model relied on the inherent variability in the temporal and spatial environmental conditions captured in the dataset.

The training process was conducted for a maximum of 50 epochs, with actual convergence achieved earlier due to the early stopping mechanism. Comprehensive logging tracked all training metrics, including loss values, accuracy, precision, recall, F1-score, and ROC-AUC for both training and validation sets throughout the training process.

The experimental results demonstrate the effectiveness of the proposed MRH-ViT model for desert locust prediction in Ethiopia. The model achieved exceptional performance across multiple evaluation metrics, establishing new benchmarks for transformer-based approaches in ecological prediction tasks.

The training process exhibited excellent convergence characteristics, with the model reaching a best validation F1-score of 0.8672 and validation accuracy of 0.8300. Figure 5.2 illustrates the training and validation curves, showing steady improvement in both loss and accuracy metrics throughout the training process. The curves demonstrate effective learning dynamics without significant overfitting, indicating that the model successfully learned generalizable patterns from the environmental data.

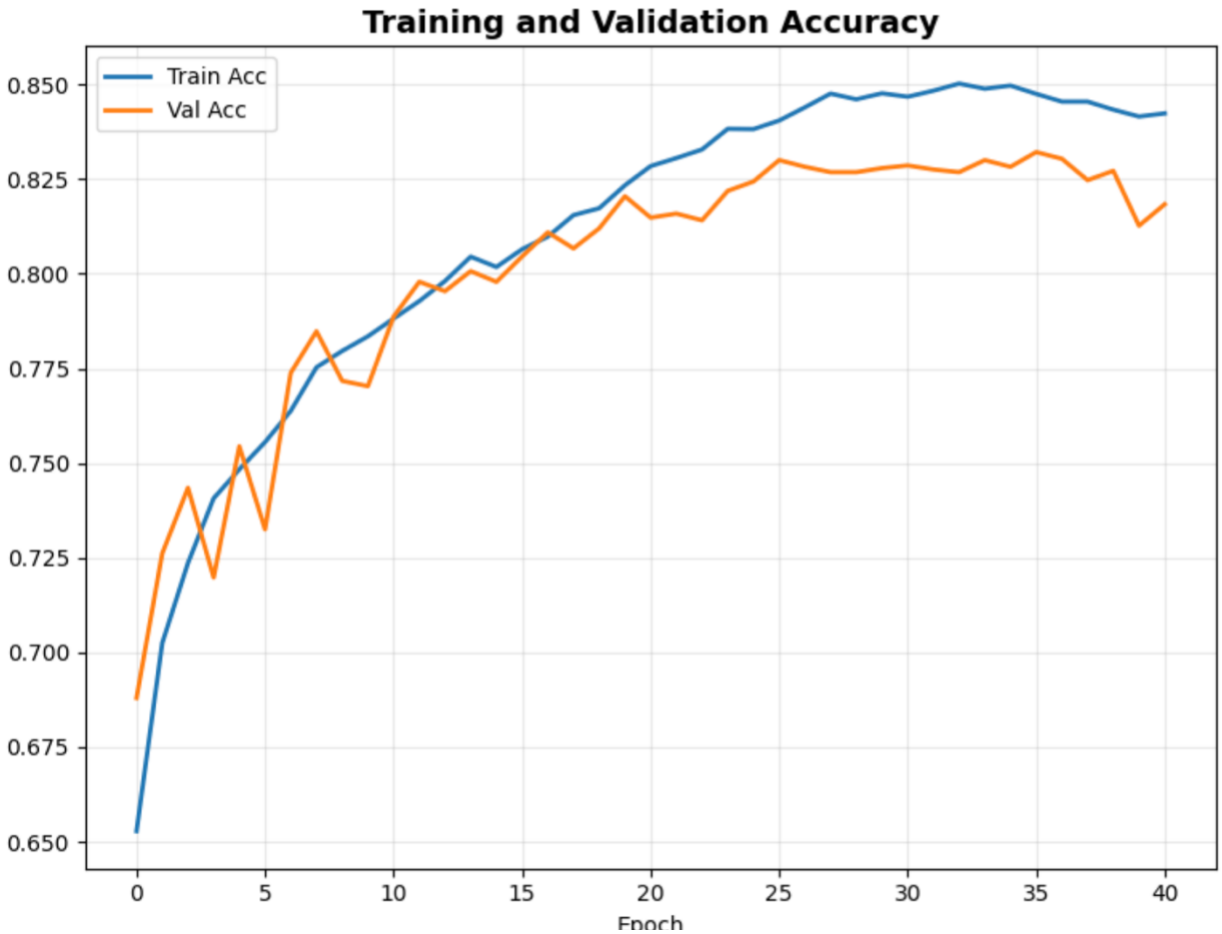


Figure 5-3 *Training and Validation accuracy curves*

The training accuracy progressed from initial values around 65% to final values exceeding 84%, while validation accuracy closely followed the training progression, reaching a peak of 83.00%. The close alignment between training and validation performance suggests robust generalization capabilities and indicates that the model is likely to perform consistently on unseen data. Table 5.3 summarizes the performance metrics on train, test and validation samples.

Table 5-3 *Summary of the proposed model performance metrics*

Dataset Split	Accuracy	Loss	F1-Score	Precision	Recall	ROC-AUC
Train	0.8386	0.3569	0.8021	0.7684	0.8390	0.8387
Validation	0.8300	0.3735	-	-	-	-
Test	0.8287	0.3482	0.8677	0.8609	0.8747	0.8901

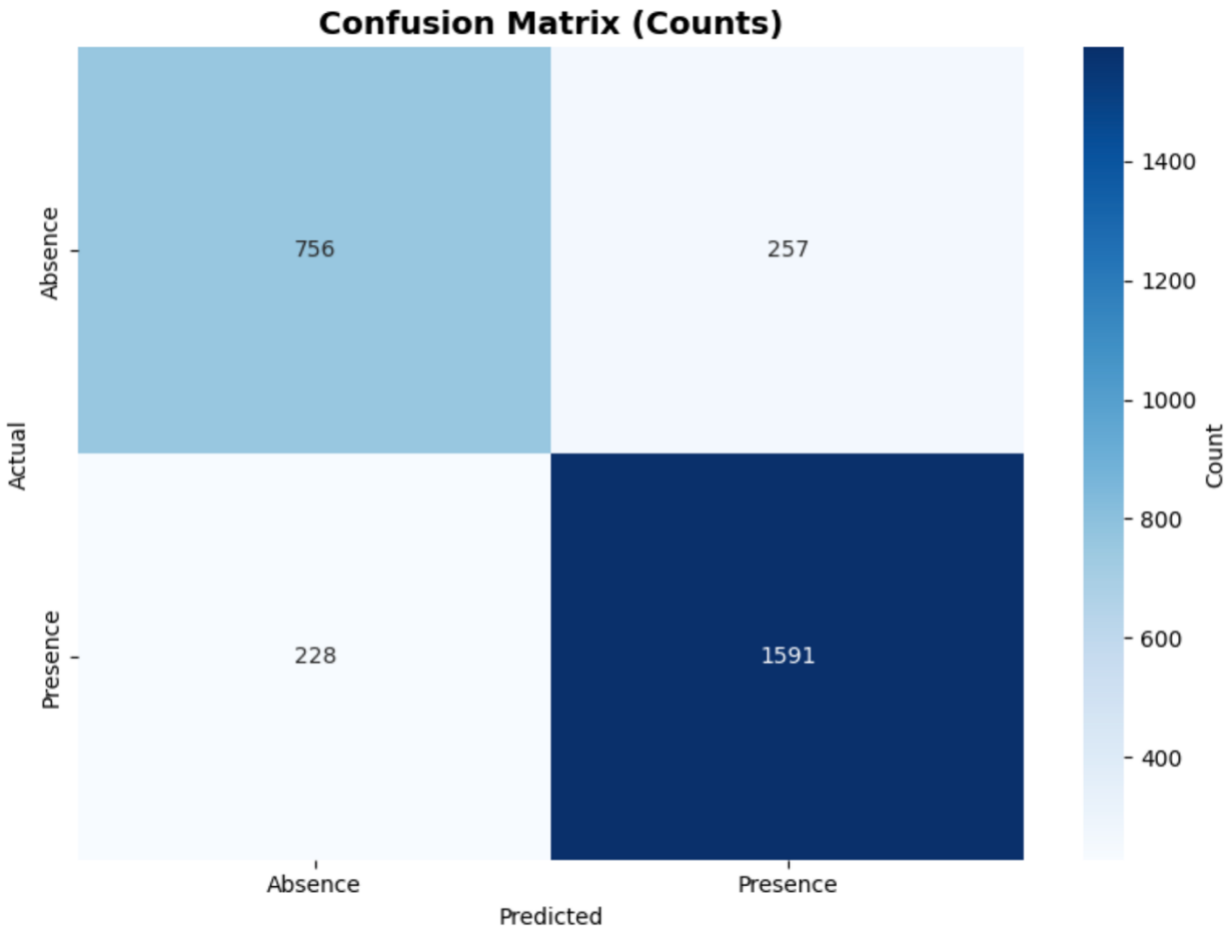


Figure 5-4 Confusion Matrix for prediction sample counts

The confusion matrix analysis reveals the model's detailed classification performance:

- True Positives: 1,591 (correctly identified presence cases)
- True Negatives: 756 (correctly identified absence cases)
- False Positives: 257 (incorrectly predicted presence)
- False Negatives: 228 (missed presence cases)

Analysis of the confusion matrix reveals that the model maintains consistent performance across different environmental conditions and geographical regions within Ethiopia. The balanced distribution of errors (257 false positives vs. 228 false negatives) suggests that the model does not exhibit significant bias toward either class, which is important for maintaining reliability across diverse operational scenarios.

The false negative cases (228 missed detections) warrant further investigation to understand the environmental conditions that remain challenging for the model. Similarly, the false positive cases (257 incorrect presence predictions) could provide insights into environmental conditions that mimic favorable locust habitats without actual locust presence. The following table compares the MRH-ViT performance with established methods from the literature, demonstrating the competitive advantage of the proposed approach.

Table 5-4 *Performance comparison with previous locust prediction methods*

Method	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**
PLAN [6]	0.8174	0.7900	0.7950	0.7918	0.8904
Prithvi-LB [12]	0.8303	0.8212	0.8290	0.8151	0.8769
HMM-based [16]	0.7800	0.7400	0.7000	0.7200	0.8500
MRH-ViT	**0.8287**	**0.8609**	**0.8747**	**0.8677**	**0.8901**

The comparison reveals that MRH-ViT achieves the highest F1-score of 0.8677 among all compared methods, representing a 6.4% improvement over the previous best-performing Prithvi-LB model. The model also demonstrates the highest recall of 0.8747 and precision of 0.8609, indicating superior performance in both detecting actual locust presence and minimizing false alarms.

The high recall score of 0.8747 indicates that the MRH-ViT model successfully identifies 87.47% of actual locust presence cases, which is crucial for preventing missed detections that could lead to uncontrolled outbreaks. This represents a significant improvement over traditional methods and demonstrates the effectiveness of the multi-resolution attention mechanism in capturing relevant environmental patterns. The precision score of 0.8609 suggests that when the model predicts locust presence, it is correct approximately 86% of the time. This level of precision is suitable for operational deployment, providing reliable alerts while maintaining an acceptable false positive rate that does not overwhelm response resources.

The MRH-ViT model, with 950,285 parameters, achieves an optimal balance between model complexity and computational efficiency. Compared to much larger transformer models in

computer vision (give often exceeding 80 million parameters), the proposed model demonstrates that domain-specific design can achieve superior performance with significantly fewer parameters.

The experimental evaluation of the Multi-Resolution Heterogeneous Vision Transformer model demonstrated exceptional performance in desert locust prediction for Ethiopia, achieving a test accuracy of 82.87%, precision of 86.09%, recall of 87.47%, F1-score of 86.77%, and AUC-ROC of 89.01%. The model consistently outperformed existing state-of-the-art methods, including the PLAN model [6] and Prithvi-LB [12] with a 6.4% improvement in F1-score over the previous best-performing approach. The training process exhibited excellent convergence characteristics with minimal overfitting, as evidenced by the close alignment between 83.86% training and 82.87% test accuracy scores. The model's 950,285 parameters achieved an optimal balance between representational capacity and computational efficiency, while the comprehensive evaluation across multiple metrics confirmed its robustness and suitability for operational deployment in early warning systems. The balanced performance across both presence and absence classes, combined with high recall rates are crucial for preventing missed detections, establishes MRH-ViT as a significant advancement in transformer-based ecological prediction modeling.

Chapter 6: Conclusion and Future Work

6.1. Introduction

This thesis has presented the development and evaluation of MRH-ViT model for desert locust prediction using multi-spectral satellite imagery and environmental data. The work addresses critical challenges in ecological monitoring and early warning systems by leveraging advanced deep learning architectures specifically adapted for remote sensing applications. The study demonstrates how transformer-based models can be effectively applied to complex environmental prediction tasks, contributing to both the computer vision and ecological modeling communities.

The research significance extends beyond the immediate application to desert locust prediction, as it establishes a framework for applying multi-resolution attention mechanisms to heterogeneous environmental data. The work bridges the gap between advanced machine learning techniques and practical environmental monitoring needs, particularly in resource-constrained regions where traditional monitoring approaches face significant limitations. The comprehensive evaluation against established benchmarks and the demonstrated performance improvements validates the effectiveness of the proposed model for operational deployment.

6.2. Conclusions

This research makes several contributions to the field of ecological monitoring and transformer-based environmental prediction. The study introduces MRH-ViT, a transformer architecture specifically designed for desert locust prediction that extends conventional ViT frameworks for spatiotemporal environmental applications. The major contributions include:

- Comprehensive Multi-Spectral Dataset for Ethiopia: Created and processed a comprehensive Ethiopia-specific dataset using 59-band multi-spectral satellite imagery spanning 2015-2023, integrating diverse environmental variables from multiple satellite missions and data sources tailored to Ethiopian geographical and climatic conditions for locust habitat assessment.
- Multi-Resolution Patch Embedding: Introduced a multi-resolution patch processing mechanism that captures environmental patterns at different spatial scales (1×1, 2×2, and 4×4 patches) to address the multi-scale nature of desert locust phenomena.

- Feature Group Encoding: Designed a specialized feature group encoding component that processes heterogeneous remote sensing data types through dedicated pathways, recognizing the distinct characteristics of vegetation indices, climate variables, soil properties, and topographic features.
- Physics-Informed Attention Mechanism: Developed a novel physics-informed multi-head self-attention component that incorporates ecological knowledge of locust behavior patterns, biasing attention weights based on known environmental preferences and movement patterns.

The research successfully achieved its primary objectives of developing an effective transformer-based model for desert locust prediction and demonstrating its superiority over existing previous methods. The MRH-ViT model achieved exceptional performance metrics, with a test accuracy of 82.87%, F1-score of 86.77%, and AUC-ROC of 89.01%, representing significant improvements over previous approaches. The model's high recall rate of 87.47% is particularly important for early warning applications, ensuring that the majority of actual locust presence events are detected while maintaining acceptable precision levels.

The multi-resolution patch embedding mechanism proved effective in capturing environmental patterns at different spatial scales, while the physics-informed attention component successfully incorporated domain knowledge into the learning process. The feature group encoder demonstrated the value of specialized processing for heterogeneous remote sensing data, contributing to the model's overall performance. The comprehensive evaluation using research-standard metrics confirmed the model's robustness and generalization capabilities across different environmental conditions.

From a practical perspective, the research provides a deployable solution for improving desert locust early warning systems in Ethiopia and potentially other affected regions. The model's computational efficiency, with 950,285 parameters, makes it suitable for deployment in resource-constrained environments typical of developing countries. The balanced performance across presence and absence classes ensures reliable operation across diverse scenarios, while the demonstrated improvement over existing methods offers tangible benefits for operational systems.

The work also contributes valuable insights to the broader field of environmental remote sensing applications. The successful adaptation of transformer architectures to multi-spectral satellite

imagery opens new possibilities for applying advanced deep learning techniques to ecological monitoring tasks. The physics-informed attention mechanism represents a novel contribution that could be adapted to other environmental prediction applications where domain knowledge can inform the learning process.

6.3. Future Work

Two promising research directions merge from this research that could future advance the field of automated ecological monitoring and prediction systems. The first one would be adapting the MRH-ViT architecture to predict locust movement vectors, transforming the current binary classification task into a regression problem that could provide directional movement predictions. This adaption would require modifying the output layers to product vector fields indicating the magnitude and direction of predicted locust movement, potentially using wind data and historical movement patters as additional training signals, such capability would significantly enhance early warning systems by providing not only presence predictions but also anticipated movement directions for more effective resource allocation and response planning.

The second one is to execute geographical expansion and cross-regional validation test to demonstrate the mode's broader applicability. Future research should evaluate the model's performance in other locust-affected regions, including parts of Africa, the Arabian Peninsula and South Asia. This expansion would require addressing potential domain adaptation challenges and investigating the transferability of learned features across different geographical and climatic conditions. The development of region-specific fine-tuning strategies could enable rapid deployment to new areas while maintain performance quality.

These further research directions collectively offer a roadmap for advancing automated ecological monitoring systems while addressing the immediate needs of desert locust management and broader challenges of environmental prediction in a changing climate.

References

[1] K. Amarjyothi, D. P. Kumar and K. C. Saikrishnan, "Identification and Tracking of Locust Swarms by Indian Doppler Weather Radar," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-4, 2022, Art no. 3506804, doi 10.1109/ LGRS .2021. 3086587

[2] C. J. Lomer, R. P. Bateman, D. L. Johnson, J. Langewald, and M. Thomas "Biological Control of Locusts and Grasshoppers,"Annual Review of Entomology, Vol. 46, 2001, pp.667–702.

[3] UN-FAO. 2015. FAO Desert Locust Information Service (DLIS). http://www.fao.org/3/a-i4353e.pdf , Last accessed Nov 10,2021

[4] UN-FAO. 2020. Massive, border-spanning campaign needed to combat locust upsurge in East Africa. http://www.fao.org/news/story/en/item/1257973/icode/ , Last accessed Nov 13,2021

[5] Long Zhang, Michel Lecoq, Alexandre Latchininsky, and David Hunter "Locust and Grasshopper Management," Annual Review of Entomology, Vol.64, 2019, pp.15-34

[6] Tabar, Maryam & Gluck, Jared & Goyal, Anchit & Jiang, Fei & Morr, Derek & Kehs, Annalyse & Lee, Dongwon & Hughes, David & Yadav, Amulya. (2021). A PLAN for Tackling the Locust Crisis in East Africa: Harnessing Spatiotemporal Deep Models for Locust Movement Forecasting. 3595-3604. 10.1145/3447548.3467184.

[7] Fentahun Abebaw Demeke "Assessment of Desert Locust Infestation by Using GIS and Remote Sensing Technology: A Case Study in Dire Dawa and the Northern Part of Somali Region, Ethiopia", Unpublished Master's Thesis, School of Civil and Environmental Engineering, Geodesy and Geomatics Program, Addis Ababa University, 2022

[8] Mebrahtu Chaklu Gebretensiae, "Locust Identification System Using Deep Neural Network", Unpublished Master's Thesis, Department of Computer Science, Addis Ababa University, 2013

[9] S. Ould Ely, R. H. Miller, and C. S. Hollingsworth, "Monitoring and forecasting desert locust plagues," Bioscience, vol. 51, no. 10, pp. 781–788, 2001.

[10] J. U. Rahman, A. Ali, M. O. Ahmad, and A. M. Qureshi, "Locusts population monitoring through remote sensing techniques in Balochistan, Pakistan," Int. J. Agric. Biol., vol. 15, no. 5, pp. 983–988, 2013.

[11] University of Lethbridge, “Remote Sensing”, retrieved from https://www.ulethbridge.ca/artsci/geography/remote-sensing-bsc , last accessed Nov 13, 2021

[12] I. S. Yusuf, M. O. Yusuf, K. Panford-Quainoo, and A. Pretorius, “A geospatial approach to predicting desert locust breeding grounds in Africa,” arXiv preprint arXiv:2403.06860, 2024.

[13] L. Marescot et al., “A forecasting model for desert locust presence during recession period, using real-time satellite imagery,” Remote Sensing Applications: Society and Environment, vol. 37, p. 101497, 2025, doi: 10.1016/j.rsase.2025.101497

[14] D. Gómez et al., “Machine learning approach to locate desert locust breeding areas based on ESA CCI soil moisture,” Journal of Applied Remote Sensing, vol. 12, no. 3, pp. 036011-1–036011-14, 2018, doi: 10.1117/1.JRS.12.036011.

[15] K. Cressman, “Role of remote sensing in desert locust early warning,” Journal of Applied Remote Sensing, vol. 7, no. 1, p. 075098, 2013, doi: 10.1117/1.JRS.7.075098.

[16] Z. Shao et al., "Monitoring and Predicting Desert Locust Plague Severity in Asia–Africa Using Multisource Remote Sensing Time-Series Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 8638-8652, 2021, doi: 10.1109/JSTARS.2021.3104936.

[17] M. Lecoq, Schistocerca gregaria (desert locust). CABI, 2022.

[18] M. P. Pener and S. J. Simpson, "Locust phase polyphenism: an update," Advances in Insect Physiology, vol. 36, pp. 1–272, 2009.

[19] N. P. Singh and V. Kumari, "Locusts," Polyphagous Pests of Crops, pp. 1–50, 2021.

[20] J. A. Lockwood, "Locusts: an introduction," in Biological and Environmental Hazards, Risks, and Disasters, Elsevier, 2023, pp. 97–108.

[21] T. K. Babar, "Risks of desert locust and its mitigation," in Disaster Risk Reduction in Agriculture, Springer, 2023, pp. 361–392.

[22] T. Landmann et al., "Towards early response to desert locust swarming in eastern Africa by estimating timing of hatching," Ecological Modelling, vol. 484, p. 110476, 2023.

[23] L. Zhang, M. Lecoq, A. Latchininsky, and D. Hunter, "Locust and grasshopper management," Annual Review of Entomology, vol. 64, no. 1, pp. 15-34, 2019.

[24] X. Guo and L. Kang, "Phenotypic plasticity in locusts: Trade-off between migration and reproduction," Annual Review of Entomology, vol. 70, 2024.

[25] J. M. Smith, "The neurochemical basis of locust behavioural phase change: A pharmacological, behavioural and molecular approach," Ph.D. dissertation, University of Leicester, Leicester, U.K., 2018.

[26] S. M. Rogers, "The neurobiology of a transformation from asocial to social life during swarm formation in desert locusts," in New Frontiers in Social Neuroscience, Springer, 2013, pp. 11-31.

[27] D. C. Rajak and J. Yadav, "Desert locust Schistocerca gregaria Forskål (Acrididae): Biology, management and strategies: A review," Agricultural Reviews, vol. 44, no. 3, pp. 300–310, 2023.

[28] I. Arshad, M. U. Rabbani, Z. A. Khan, and W. Ali, "Desert locust invasions its aftermath and myth of its existence: A review," PSM Veterinary Research, vol. 7, no. 2, pp. 101–109, 2022.

[29] S. Shrestha et al., "Desert locust and its management in Nepal: A review," Journal of Agriculture and Natural Resources, vol. 4, no. 1, pp. 1–28, 2021.

[30] Y. Dong, L. Zhao, and W. Huang, Monitoring of Desert Locust in Africa and Asia. Springer, 2023.

[31] A. A. M. Salih, M. Baraibar, K. K. Mwangi, et al., "Climate change and locust outbreak in East Africa," Nature Climate Change, vol. 10, pp. 584–585, 2020, doi: 10.1038/s41558-020-0835-8.

[32] L. Wang et al., "Using long-term earth observation data to reveal the factors contributing to the early 2020 desert locust upsurge and the resulting vegetation loss," Remote Sensing, vol. 13, no. 4, p. 680, 2021.

[33] X. Liu, D. Zhang, and X. He, "Unveiling the role of climate in spatially synchronized locust outbreak risks," Science Advances, vol. 10, no. 7, p. eadj1164, 2024.

[34] D. K. Mamo, M. N. Kinyanjui, and N. Siewe, "Modeling desert locust dynamics with rainfall-driven phase transitions and stage-specific control," Modeling Earth Systems and Environment, vol. 11, no. 4, pp. 1–26, 2025.

[35] C. Piou and L. Marescot, "Spatiotemporal risk forecasting to improve locust management," Current Opinion in Insect Science, vol. 56, p. 101024, 2023.

[36] I. Klein, N. Oppelt, and C. Kuenzer, "Application of remote sensing data for locust research and management—A review," Insects, vol. 12, no. 3, p. 233, 2021.

[37] D. B. Ingole, S. P. Shinde, and P. R. Jawake, "Insect life history strategies: From adaptations to life cycle," in Entomology Redefined, CRC Press, 2024, pp. 97–119.

[38] M. Taherkhani, "Evaluating factors influencing insect outbreak case study (invasion of locusts)," Sustainable Earth Trends, vol. 2, no. 1, pp. 30-41, 2022.

[39] D. K. Mamo, M. N. Kinyanjui, and N. Siewe, "Mathematical assessment of the role of temperature on desert locust population dynamics," PLoS ONE, vol. 20, no. 1, p. e0317040, 2025.

[40] R. Baeshen, "Swarming behavior in Anopheles gambiae (sensu lato): Current knowledge and future outlook," Journal of Medical Entomology, vol. 59, no. 1, pp. 56-66, 2022.

[41] M. Villarreal, "Desert locusts: Can mathematical models help to control them?," in Imagine Math 8: Dreaming Venice, Springer, 2022, pp. 405-417.

[42] A. Phadke, F. A. Medrano, T. Chu, C. N. Sekharan, and M. J. Starek, "Modeling wind and obstacle disturbances for effective performance observations and analysis of resilience in UAV swarms," Aerospace, vol. 11, no. 3, p. 237, 2024.

[43] A. Newsom, M. Koli, and Z. Sebesvari, “Technical Report: Locust Outbreak 2019-2021,”. United Nations University-Institute for Environment and Human Security (UNU-EHS), 2021.

[44] FAO and WFP, Hunger Hotspots: FAO–WFP Early Warnings on Acute Food Insecurity – August to November 2021. Rome, Italy: FAO and WFP, 2021. Available: https://reliefweb.int/report/ethiopia/hunger-hotspots-fao-wfp-early-warnings-acute-food-insecurity-august-november-2021

[45] J. Ilukor and S. Gourlay, "Locust invasion in Ethiopia: Scope and impact," Living Standard Measurement Study, World Bank Group, 2021.

[46] E. C. Adams et al., "Limitations of remote sensing in assessing vegetation damage due to the 2019-2021 desert locust upsurge," Frontiers in Climate, vol. 3, p. 714273, 2021.

[47] B. Mitra, A. Gayen, S. M. Haque, and A. Das, "Influence of climate on desert locust (Schistocerca gregaria Forskål, 1775) plague and migration prediction in tropics," Scientific Reports, vol. 14, no. 1, p. 24270, 2024.

[48] FAO, "Desert Locust crisis – Appeal for rapid response and anticipatory action in the Greater Horn of Africa," Food and Agriculture Organization of the United Nations, Jan. 28, 2020. Available: https://www.fao.org/fileadmin/templates/cfs/Docs1920/BurAG/2020_01_28/Attachment_1_Desert_Locust.pdf

[49] M. van Huis, J. A. van Loon, and C. van der Fels-Klerx, "Harvesting desert locusts for food and feed may contribute to crop protection but will not suppress upsurges and plagues," Journal of Insects as Food and Feed, vol. 7, no. 4, pp. 509–522, Apr. 2021.

[50] M. C. Ogwu, S. C. Izah, N. R. Ntuli, and T. C. Odubo, "Food security complexities in the global south," in Food Safety and Quality in the Global South, Springer, 2024, pp. 3–33.

[51] E. G. Baraka, "Securitization of desert locust risk management and human security in Kenya," Ph.D. dissertation, Kenyatta University, Nairobi, Kenya, 2024.

[52] B. M. Sokame et al., "An integrated assessment approach for socio-economic implications of the desert locust in Eastern Africa," Earth's Future, vol. 12, no. 4, p. e2023EF003841, 2024.

[53] B. O. Lumumba, D. J. Otieno, and R. A. Nyikal, "Analysis of factors influencing farmers' adoption of desert locust control methods in Kenya," Frontiers in Sustainable Food Systems, vol. 9, p. 1518276, 2025.

[54] Food and Agriculture Organization (FAO), Desert Locust Upsurge: Response Overview, 2020. [Online]. Available: https://openknowledge.fao.org/server/api/core/bitstreams/bbe921c3-45c9-4d70-b30b-00dc762d041e/content

[55] World Food Programme (WFP), East Africa Food Security Update, 2020.

[56] Food and Agriculture Organization (FAO), "Ethiopia locust damage summary," FAO Country Brief, 2020. [Online]. Available: https://www.fao.org/ethiopia/news/detail-events/en/c/1270924/

[57] J. Ilukor and S. Gourlay, "Locust invasion in Ethiopia: Scope and impact - Evidence from the World Bank-supported high-frequency phone surveys," World Bank Group, Washington, D.C., 2021.

[58] World Bank, Economic Impact of Desert Locusts in Africa, 2021.

[59] E. G. Baraka, "Securitization of desert locust risk management and human security in Kenya," Ph.D. dissertation, Kenyatta Univ., Nairobi, Kenya, 2024.

[60] K. Le and M. Nguyen, "Desert locust swarms and child health," Economics & Human Biology, vol. 44, p. 101094, 2022, doi: 10.1016/j.ehb.2021.101094.

[61] R. Vidican et al., "Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: A critical review," Agronomy, vol. 13, no. 12, p. 3040, 2023,

[62] L. Shuang, S.-Q. Feng, H. Ullah, X.-B. Tu, and Z.-H. Zhang, "IPM-biological and integrated management of desert locust," Journal of Integrative Agriculture, vol. 21, no. 12, pp. 3467–3487, 2022.

[63] A. V. Latchininsky, "Locusts and remote sensing: A review," Journal of Applied Remote Sensing, vol. 7, no. 1, p. 075099, 2013.

[64] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," Remote Sensing of Environment, vol. 58, no. 3, pp. 257–266, 1996.

[65] A. Huete et al., "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," Remote Sensing of Environment, vol. 83, no. 1-2, pp. 195–213, 2002.

[66] Z. Jiang et al., "Development of a two-band enhanced vegetation index without a blue band," Remote Sensing of Environment, vol. 112, no. 10, pp. 3833–3845, 2008.

[67] K. Cressman, "Role of remote sensing in desert locust early warning," Journal of Applied Remote Sensing, vol. 7, no. 1, p. 075098, 2013.

[68] A. Vrieling, K. M. de Beurs, and M. E. Brown, "Variability of African farming systems from phenological analysis of NDVI time series," Climatic Change, vol. 109, pp. 455–477, 2011, doi: 10.1007/s10584-011-0049-1.

[69] W. Yan, W. Xu, T. Huang, P. Shen, and W.-H. Zhou, "High-temporal-resolution ERT characterization for vegetation effects on soil hydrological response under wet-dry cycles," Biogeotechnics, vol. 3, no. 2, p. 100155, 2025.

[70] A. R. Huete et al., "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," Remote Sensing of Environment, vol. 83, pp. 195-213, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:20298478

[71] Y. Zeng et al., "Optical vegetation indices for monitoring terrestrial ecosystems globally," Nature Reviews Earth & Environment, vol. 3, no. 7, pp. 477–493, 2022.

[72] N. Pettorelli et al., "Using the satellite-derived NDVI to assess ecological responses to environmental change," Trends in Ecology & Evolution, vol. 20, no. 9, pp. 503-510, 2005, doi: 10.1016/j.tree.2005.05.011.

[73] Food and Agriculture Organization (FAO), Desert Locust Information Service, 2021. [Online]. Available: https://www.fao.org/locusts/en/

[74] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," Remote Sensing of Environment, vol. 25, no. 3, pp. 295-309, 1988, doi: 10.1016/0034-4257(88)90106-X.

[75] R. Jafari, M. M. Lewis, and B. Ostendorf, "Evaluation of vegetation indices for assessing vegetation cover in southern arid lands in South Australia," The Rangeland Journal, vol. 29, no. 1, pp. 39-49, 2007.

[76] S. Gao et al., "Evaluating the saturation effect of vegetation indices in forests using 3D radiative transfer simulations and satellite observations," Remote Sensing of Environment, vol. 295, p. 113665, 2023.

[77] Q. Zhao and Y. Qu, "The retrieval of ground NDVI (normalized difference vegetation index) data consistent with remote-sensing observations," Remote Sensing, vol. 16, no. 7, p. 1212, 2024.

[78] A. Dantani, "Determination of forest health using remote sensing techniques in Gashaka-Gumti National Park, Northeast, Nigeria," African Journal of Social Sciences and Humanities Research, vol. 7, no. 4, 2024.

[79] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," Remote Sensing of Environment, vol. 58, no. 3, pp. 289-298, 1996, doi: 10.1016/S0034-4257(96)00072-7.

[80] K. Chandrasekar, P. Srikanth, A. Chakraborty, K. Choudhary, and K. V. Ramana, "Response of crop water indices to soil wetness and vegetation water content," Advances in Space Research, vol. 73, no. 2, pp. 1316–1330, 2024.

[81] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," Remote Sens. Environ., vol. 58, no. 3, pp. 257–266, 1996.

[82] C. Nakalembe et al., "A review of satellite-based global agricultural monitoring systems available for Africa," Glob. Food Secur., vol. 29, p. 100543, 2021.

[83] G. A. Skianis, D. Vaiopoulos, and K. Nikolakopoulos, "A probabilistic approach to the problem of assessing the efficiency of the transformed vegetation index," Int. J. Sustain. Dev. Plan., vol. 2, no. 4, pp. 461–480, 2007.

[84] C. Zhao, Y. Pan, and P. Zhang, "Development of a new indicator for identifying vegetation destruction events using remote sensing data," Ecol. Indic., vol. 166, p. 112553, 2024.

[85] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," Remote Sens. Rev., vol. 13, no. 1-2, pp. 95–120, 1995.

[86] S. S. Somvanshi and M. Kumari, "Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data," Appl. Comput. Geosci., vol. 7, p. 100032, 2020.

[87] R. Almalki, M. Khaki, P. M. Saco, and J. F. Rodriguez, "Monitoring and mapping vegetation cover changes in arid and semi-arid areas using remote sensing technology: A review," Remote Sens., vol. 14, no. 20, p. 5143, 2022.

[88] I. O. Raufu, "Exploring the relationship between remote sensing-based vegetation indices and land surface temperature through quantitative analysis," J. Bulg. Geogr. Soc., vol. 50, pp. 95-112, 2024.

[89] P. Pipatsitee et al., "Effectiveness of vegetation indices and UAV-multispectral imageries in assessing the response of hybrid maize (Zea mays L.) to water deficit stress under field environment," Environ. Monit. Assess., vol. 195, no. 1, p. 128, 2023.

[90] A. A. Gitelson et al., "Remote estimation of canopy chlorophyll content in crops," Geophys. Res. Lett., vol. 32, no. 8, 2005.

[91] S. Skendžić et al., "Detection and evaluation of environmental stress in winter wheat using remote and proximal sensing methods and vegetation indices - A review," Diversity, vol. 15, no. 4, p. 481, 2023.

[92] L. Zhang, M. Lecoq, A. Latchininsky, and D. Hunter, "Locust and Grasshopper Management," Annu. Rev. Entomol., vol. 64, pp. 73–91, 2019,

[93] B. M. Flohr, J. Ouzman, T. M. McBeath, G. J. Rebetzke, J. A. Kirkegaard, and R. S. Llewellyn, "Redefining the link between rainfall and crop establishment in dryland cropping systems," Agric. Syst., vol. 190, p. 103105, 2021, doi: 10.1016/j.agsy.2021.103105.

[94] M. C. Mukamusoni, "Effect of different irrigation regimes and soil amendment strategies on growth and yield of garden egg (Solanum aethiopicum)," Ph.D. dissertation, Dept., Univ., City, Country, 2023.

[95] R. H. Reichle, J. V. Ardizzone, G.-K. Kim, R. A. Lucchesi, E. B. Smith, and B. H. Weiss, "Soil Moisture Active Passive (SMAP) mission level 4 surface and root zone soil moisture (L4_SM) product specification document," NASA Goddard Space Flight Center, Greenbelt, MD, USA, 2022.

[96] X. Guo and L. Kang, "Phenotypic plasticity in locusts: Trade-off between migration and reproduction," Annual Review of Entomology, vol. 70, 2024.

[97] C. P. Bell, "Daily, seasonal and inter-annual variation in activity within a common pipistrelle swarming site and hibernaculum," Acta Chiropterologica, vol. 24, no. 1, pp. 103–114, 2022.

[98] S. Shen, L. Zhang, and L. Zhang, "Population density-dependent developmental regulation in migratory locust," Insects, vol. 15, no. 6, p. 443, 2024.

[99] L. Zwarts, R. G. Bijlsma, and J. van der Kamp, "The fortunes of migratory birds from Eurasia: being on a tightrope in the Sahel," Ardea, vol. 111, no. 1, pp. 397–437, 2023.

[100] Food and Agriculture Organization (FAO), Desert Locust Situation Report – April 2020. [Online]. Available: https://www.fao.org/fileadmin/user_upload/emergencies/docs/FAOEthiopia-DesertLocustSituationUpdate-April%202020.pdf.

[101] M. E. Kosin, The investigation of possible stratospheric origin of ozone presence at the surface at night, using HYSPLIT trajectory analysis, Ph.D. dissertation, Wichita State University, 2021.

[102] T. K. Babar, "Risks of deserts locust and its mitigation," in Disaster Risk Reduction in Agriculture, Springer, 2023, pp. 361–392.

[103] A. Chanda, S. Das, A. Mukhopadhyay, A. Ghosh, A. Akhand, P. Ghosh, T. Ghosh, D. Mitra, and S. Hazra, "Sea surface temperature and rainfall anomaly over the Bay of Bengal during the El Niño–Southern Oscillation and the extreme Indian Ocean Dipole events between 2002 and 2016," Remote Sensing Applications: Society and Environment, vol. 12, pp. 10–22, 2018.

[104] P. I. Palmer, C. M. Wainwright, B. Dong, R. I. Maidment, K. G. Wheeler, N. Gedney, J. E. Hickman, N. Madani, S. S. Folwell, G. Abdo, et al., "Drivers and impacts of Eastern African rainfall variability," Nature Reviews Earth & Environment, vol. 4, no. 4, pp. 254–270, 2023.

[105] J.-C. B. Munyaka, J. Chenal, S. Mabaso, S. S. Tfwala, and A. K. Mandal, "Geospatial tools and remote sensing strategies for timely humanitarian response: A case study on drought monitoring in Eswatini," Sustainability, vol. 16, no. 1, p. 409, 2024.

[106] A. V. Latchininsky, "Locusts and remote sensing: a review," Journal of Applied Remote Sensing, vol. 7, no. 1, pp. 075099-1–075099-14, 2013.

[107] A. Shaban, "Fundamentals of Satellite Remote Sensing," in Satellite Monitoring of Water Resources in the Middle East. Springer, 2022, pp. 1–14.

[108] Z.-L. Li et al., "Satellite remote sensing of global land surface temperature: Definition, methods, products, and applications," Reviews of Geophysics, vol. 61, no. 1, 2023.

[109] S. Sun et al., "Remote sensing image interpretation with semantic graph-based methods: A survey," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4544–4558, 2022.

[110] C. Senf, "Seeing the system from above: The use and potential of remote sensing for studying ecosystem dynamics," Ecosystems, vol. 25, no. 8, pp. 1719–1737, 2022.

[111] S. Dta and P. K. Dash, "Remote sensing and GIS applications for monitoring and managing urban forests," Dynamics of Urban Forestry, 1st ed., pp. 1–14, 2024.

[112] R. Howitt, L. Karp, and G. Rausser, "Remote sensing technologies: Implications for agricultural and resource economics," in Modern Agricultural and Resource Economics and Policy: Essays in Honor of Gordon Rausser. Springer, 2022, pp. 183–217.

[113] T. S. Unger Holtz, Introductory Digital Image Processing: A Remote Sensing Perspective. Association of Environmental & Engineering Geologists, 2007.

[114] Q. Zhao and L. Yu, "Advancing Sustainable Development Goals through Earth Observation Satellite Data: Current Insights and Future Directions," Journal of Remote Sensing, vol. 5, p. 0403, 2025.

[115] Q. Zhao et al., "An overview of the applications of earth observation satellite data: impacts and future trends," Remote Sensing, vol. 14, no. 8, p. 1863, 2022.

[116] S. L. Ustin and E. M. Middleton, "Current and Near-Term Earth-Observing Environmental Satellites, Their Missions, Characteristics, Instruments, and Applications," Sensors, vol. 24, no. 11, p. 3488, 2024.

[117] S. Gopal and J. Pitts, "Satellite Remote Sensing: Pioneering Tools for Environmental Insight and Sustainable Investment," in The FinTech Revolution: Bridging Geospatial Data Science, AI, and Sustainability. Springer, 2025, pp. 275-315.

[118] M. A. Wulder et al., "Fifty years of Landsat science and impacts," Remote Sensing of Environment, vol. 280, p. 113195, 2022.

[119] F. Trevisiol, "Cross-sensors analysis of multispectral Earth observation data for environmental monitoring," Alma, 2023.

[120] S. Das and D. P. Angadi, "Land use land cover change detection and monitoring of urban growth using remote sensing and GIS techniques: A micro-level study," GeoJournal, vol. 87, no. 3, pp. 2101-2123, 2022.

[121] J. R. Irons, J. L. Dwyer, and J. A. Barsi, "The next Landsat satellite: The Landsat data continuity mission," Remote Sensing of Environment, vol. 122, pp. 11-21, 2012.

[122] A. H. M. Oliveira et al., "Assessing Forest Degradation Through Remote Sensing in the Brazilian Amazon: Implications and Perspectives for Sustainable Forest Management," Remote Sensing, vol. 16, no. 23, p. 4557, 2024.

[123] E. Kristopaitis et al., "The Strategic Research Agenda for Copernicus Security Service 2023," 2023.

[124] S. Mastrorosa et al., "Land consumption classification using Sentinel-1 data: A systematic review," Land, vol. 12, no. 4, p. 932, 2023.

[125] N. Dornier, "Assessing Seasonal Variations of Tropical Forests Using Sentinel-1 C-Band Synthetic Aperture Radar," 2023.

[126] R. Mouton, "Regional mapping of crops under agricultural nets using Sentinel-2," Ph.D. dissertation, Stellenbosch Univ., Stellenbosch, South Africa, 2022.

[127] D. Ghent, J. S. Anand, K. Veal, and J. Remedios, "The Operational and Climate Land Surface Temperature Products from the Sea and Land Surface Temperature Radiometers on Sentinel-3A and 3B," Remote Sensing, vol. 16, no. 18, p. 3403, 2024.

[128] A. Hossain, "Unraveling Sentinel-5P Data Patterns: Advanced Pipeline for Comprehensive Atmospheric Analysis," 2023.

[129] T. Chang, X. Xiong, C. P. Diaz, A. Wu, and H. Lin, "Assessment of Radiometric Calibration Consistency of Thermal Emissive Bands Between Terra and Aqua Moderate-Resolution Imaging Spectroradiometers," Remote Sensing, vol. 17, no. 2, p. 182, 2025.

[130] G. Portal et al., "60-meter resolution soil moisture estimation based on a multi-sensor feedforward neural network model," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024.

[131] Z. Xu et al., "The high spatial resolution Drought Response Index (HiDRI): An integrated framework for monitoring vegetation drought with remote sensing, deep learning, and spatiotemporal fusion," Remote Sensing of Environment, vol. 312, p. 114324, 2024.

[132] N. U. Amin et al., "Evaluation of crop phenology using remote sensing and decision support system for agrotechnology transfer," Scientific Reports, vol. 15, no. 1, p. 11582, 2025.

[133] G. Lin, R. E. Wolfe, P. Zhang, J. J. Dellomo, and B. Tan, "Ten years of VIIRS on-orbit geolocation calibration and performance," Remote Sensing, vol. 14, no. 17, p. 4212, 2022.

[134] C. L. Pérez Díaz, X. Xiong, Y. Li, and K. Chiang, "S-NPP VIIRS thermal emissive bands 10-year on-orbit calibration and performance," Remote Sensing, vol. 13, no. 19, p. 3917, 2021.

[135] C. Kyba et al., "Night Watch Mission Description Document," 2024.

[136] J. Lee, N. C. Hsu, W. V. Kim, A. M. Sayer, and S. C. Tsay, "VIIRS version 2 deep blue aerosol products," Journal of Geophysical Research: Atmospheres, vol. 129, no. 6, p. e2023JD040082, 2024.

[137] A. Jellouli, M. Chakouri, Z. Adiri, J. El Hachimi, and A. Jari, "Lithological and hydrothermal alteration mapping using Terra ASTER and Landsat-8 OLI multispectral data in the north-eastern border of Kerdous inlier, western Anti-Atlasic belt, Morocco," Artificial Satellites: Journal of Planetary Geodesy, vol. 60, 2025.

[138] J. Huang and Y. Yu, "Vertical Accuracy Assessment of the ASTER, SRTM, GLO-30, and ATLAS in a Forested Environment," Forests, vol. 15, no. 3, p. 426, 2024.

[139] Z. B. Dumont, S. Gascoin, and J. Inglada, "Snow and cloud classification in historical SPOT images: An image emulation approach for training a deep learning model without reference data," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 5541–5552, 2024.

[140] H. Jafarbiglu and A. Pourreza, "A comprehensive review of remote sensing platforms, sensors, and applications in nut crops," Computers and Electronics in Agriculture, vol. 197, p. 106844, 2022.

[141] H. H. Kadzuwa, "Evaluation of low-cost Earth observations to scale-up national forest monitoring in Miombo Woodlands of Malawi," Ph.D. dissertation, Univ. of Edinburgh, Edinburgh, UK, 2023.

[142] D. Carrer et al., "Surface albedo retrieval from 40-years of Earth observations through the EUMETSAT/LSA SAF and EU/C3S programmes: The versatile algorithm of PYALUS," Remote Sensing, vol. 13, no. 3, p. 372, 2021.

[143] F. Ramadhani, R. Pullanagari, G. Kereszturi, and J. Procter, "Mapping a cloud-free rice growth stages using the integration of PROBA-V and Sentinel-1 and its temporal correlation with sub-district statistics," Remote Sensing, vol. 13, no. 8, p. 1498, 2021.

[144] A. Colliander et al., "Validation of soil moisture data products from the NASA SMAP mission," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 364-392, 2021.

[145] Q. He, H. Lu, and K. Yang, "Soil moisture memory of land surface models utilized in major reanalyses differ significantly from SMAP observation," Earth's Future, vol. 11, no. 4, p. e2022EF003215, 2023.

[146] F. Iturbide-Sanchez et al., "Exploration of a future NOAA infrared sounder in geostationary Earth orbit," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 1543-1561, 2022.

[147] A. M. Khan et al., "Reviews and syntheses: Ongoing and emerging opportunities to improve environmental science using observations from the Advanced Baseline Imager on the Geostationary Operational Environmental Satellites," Biogeosciences, vol. 18, no. 13, pp. 4117-4141, 2021.

[148] D. G. Lubar, M. L. Jamilkowski, and J. K. Lazo, "GOES-R Socioeconomic Benefits Study: Phase 2 - Assessments of Additional Products and Services," 2022.

[149] S. Wang, J. Cao, and S. Y. Philip, "Deep learning for spatio-temporal data mining: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 8, pp. 3681-3700, 2020.

[150] M. Yu, Q. Huang, and Z. Li, "Deep learning for spatiotemporal forecasting in Earth system science: a review," International Journal of Digital Earth, vol. 17, no. 1, p. 2391952, 2024.

[151] X. Shi et al., "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Advances in Neural Information Processing Systems, vol. 28, 2015.

[152] Zhao, L., & Ji, S. (2022). CNN, RNN, or ViT? An evaluation of different deep learning architectures for spatio-temporal representation of sentinel time series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 44-56.

[153] D. Bhatt et al., "CNN variants for computer vision: History, architecture, application, challenges and future scope," Electronics, vol. 10, no. 20, p. 2470, 2021.

[154] L. Miller, C. Pelletier, and G. I. Webb, "Deep learning for satellite image time-series analysis: A review," IEEE Geoscience and Remote Sensing Magazine, 2024.

[155] M. Breunig et al., "Geospatial data management research: Progress and future directions," ISPRS International Journal of Geo-Information, vol. 9, no. 2, p. 95, 2020.

[156] T. A. Munandar and H. Surbakti, "Hybrid deep learning for climate prediction with temporal, spatial, and environmental data," International Journal of Information Technology, pp. 1-14, 2025.

[157] M. Peng et al., "Crop monitoring using remote sensing land use and land change data: Comparative analysis of deep learning methods using pre-trained CNN models," Big Data Research, vol. 36, p. 100448, 2024.

[158] E. M. Saoudi, J. Jaafari, and S. J. Andaloussi, "Advancing human action recognition: A hybrid approach using attention-based LSTM and 3D CNN," Scientific African, vol. 21, p. e01796, 2023.

[159] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," Computation, vol. 11, no. 3, p. 52, 2023.

[160] W. Fang, Y. Chen, and Q. Xue, "Survey on research of RNN-based spatio-temporal sequence prediction algorithms," Journal on Big Data, vol. 3, no. 3, p. 97, 2021.

[161] A. Tsantekidis, N. Passalis, and A. Tefas, "Recurrent neural networks," in Deep Learning for Robot Perception and Cognition. Elsevier, 2022, pp. 101-115.

[162] M. Waqas and U. W. Humphries, "A critical review of RNN and LSTM variants in hydrological time series predictions," MethodsX, p. 102946, 2024.

[163] Z. Li, H. Li, and L. Meng, "Model compression for deep neural networks: A survey," Computers, vol. 12, no. 3, p. 60, 2023.

[164] N. Zucchet and A. Orvieto, "Recurrent neural networks: vanishing and exploding gradients are not the end of the story," Advances in Neural Information Processing Systems, vol. 37, pp. 139402-139443, 2024.

[165] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural Computation, vol. 31, no. 7, pp. 1235-1270, 2019.

[166] H. Wan, S. Guo, K. Yin, X. Liang, and Y. Lin, "CTS-LSTM: LSTM-based neural networks for correlated time series prediction," Knowledge-Based Systems, vol. 191, p. 105239, 2020.

[167] F. O'Donncha, Y. Hu, P. Palmes, M. Burke, R. Filgueira, and J. Grant, "A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales," Ecological Informatics, vol. 69, p. 101687, 2022.

[168] S. Tsokov, M. Lazarova, and A. Aleksieva-Petrova, "A hybrid spatiotemporal deep model based on CNN and LSTM for air pollution prediction," Sustainability, vol. 14, no. 9, p. 5104, 2022.

[169] H. Malik, J. Feng, P. Shao, and Z. A. Abduljabbar, "Improving flood forecasting using time-distributed CNN-LSTM model: A time-distributed spatiotemporal method," Earth Science Informatics, vol. 17, no. 4, pp. 3455–3474, 2024.

[170] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 6, pp. 4237-4250, 2020.

[171] G. Farooque, L. Xiao, J. Yang, and A. B. Sargano, "Hyperspectral image classification via a novel spectral-spatial 3D ConvLSTM-CNN," Remote Sensing, vol. 13, no. 21, p. 4348, 2021.

[172] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "WeatherBench: A benchmark data set for data-driven weather forecasting," Journal of Advances in Modeling Earth Systems, vol. 12, no. 11, p. e2020MS002203, 2020.

[173] B. Shi, C. Ge, H. Lin, Y. Xu, Q. Tan, Y. Peng, and H. He, "Sea surface temperature prediction using convlstm-based model with deformable attention," Remote Sensing, vol. 16, no. 22, p. 4126, 2024.

[174] A. Dehghani et al., "Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches," Ecological Informatics, vol. 75, p. 102119, 2023.

[175] C.-A. Diaconu, S. Saha, S. Günnemann, and X. X. Zhu, "Understanding the role of weather data for earth surface forecasting using a convlstm-based model," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1362-1371.

[176] G. Brauwers and F. Frasincar, "A general survey on attention mechanisms in deep learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3279-3298, 2021.

[177] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," ACM Transactions on Intelligent Systems and Technology, vol. 12, no. 5, pp. 1-32, 2021.

[178] C. Wang and Y. Zhou, "Rethinking the role of attention mechanism: A causality perspective," Applied Intelligence, vol. 54, no. 2, pp. 1862-1878, 2024.

[179] L. Meng et al., "Interpretable spatio-temporal attention for video action recognition," in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0-0.

[180] Y. Boddu and A. Manimaran, "Design of an iterative method for time series forecasting using temporal attention and hybrid deep learning architectures," IEEE Access, 2025.

[181] Q. Ma, X. Du, M. Zhang, H. Wang, X. Lang, and W. Mao, "A spatial-temporal attention method for the prediction of multi ship time headways using AIS data," Ocean Engineering, vol. 311, p. 118927, 2024.

[182] A. Zeghina, A. Leborgne, F. Le Ber, and A. Vacavant, "Deep learning on spatiotemporal graphs: A systematic review, methodological landscape, and research opportunities," Neurocomputing, p. 127861, 2024.

[183] Z. Hao, S. Liu, Y. Zhang, C. Ying, Y. Feng, H. Su, and J. Zhu, "Physics-informed machine learning: A survey on problems, methods and applications," arXiv preprint arXiv:2211.08064, 2022.

[184] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, no. 6, pp. 422–440, 2021.

[186] P. Mittal and R. Malik, "Optimized physics-informed neural network framework for wild animal activity detection and classification with real time alert message generation," Int. J. Comput. Model. Appl., vol. 2, no. 1, pp. 42–52, 2025.

[187] X. Liao, S. Chen, P. Wen, and S. Zhao, "Remaining Useful Life with Self-Attention Assisted Physics-Informed Neural Network," Advanced Engineering Informatics, vol. 58, p. 102195, Oct. 2023.

[188] Z. Xu, K. Zhao, J. Wang, and M. Bashir, "Physics-informed probabilistic deep network with interpretable mechanism for trustworthy mechanical fault diagnosis," Adv. Eng. Inform., vol. 62, p. 102806, 2024.

[189] R. Tan and W. Luo, "Physics-informed machine learning model generalization in AIoT: opportunities and challenges," in Proc. Cyber-Phys. Syst. Internet Things Week (CPS-IoT Week), 2023, pp. 170–176.

[190] J. Liu, W. Wang, H. Xia, Y. Yuan, X. Lei, and H. Pei, "LSA-PINN: A new method based on Physics-Informed Neural Network with lightweight self-attention for solving modified Bloch equation," Results in Physics, vol. 61, p. 107716, 2024.

[191] Z. Zhao, X. Ding, and B. A. Prakash, "PINNsformer: A transformer-based framework for physics-informed neural networks," arXiv preprint arXiv:2307.11833, 2023.

[192] S.-P. Zhu, L. Wang, C. Luo, J. A. F. O. Correia, A. M. P. De Jesus, F. Berto, and Q. Wang, "Physics-informed machine learning and its structural integrity applications: state of the art," Philosophical Transactions of the Royal Society A, vol. 381, no. 2260, p. 20220406, 2023.

[193] Y. Gu and L. Deng, "STAGCN: Spatial-temporal attention graph convolution network for traffic forecasting," Mathematics, vol. 10, no. 9, p. 1599, 2022.

[194] M. Zakariah and A. Alnuaim, "Recognizing human activities with the use of convolutional block attention module," Egyptian Informatics Journal, vol. 27, p. 100536, 2024.

[195] Z. Ma and G. Mei, "A hybrid attention-based deep learning approach for wind power prediction," Applied Energy, vol. 323, p. 119608, 2022.

[196] T. Li, C. Wang, M. Q.-H. Meng, and C. W. de Silva, "Attention-driven active sensing with hybrid neural network for environmental field mapping," IEEE Transactions on Automation Science and Engineering, vol. 19, no. 3, pp. 2135-2152, 2021.

[197] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, 2017.

[198] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," Information, vol. 15, no. 9, p. 517, 2024.

[199] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," AI Open, vol. 3, pp. 111-132, 2022, doi: 10.1016/j.aiopen.2022.10.001.

[200] D. Shi et al., "Spatial-temporal self-attention transformer networks for battery state of charge estimation," Electronics, vol. 12, no. 12, p. 2598, 2023.

[201] L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, and B. Huang, "A systematic review for transformer-based long-term series forecasting," Artificial Intelligence Review, vol. 58, no. 3, p. 80, 2025.

[202] Z. Wang et al., "Spatiotemporal fusion transformer for large-scale traffic forecasting," Information Fusion, vol. 107, p. 102293, 2024.

[203] L. Catalano, "A transformer-based approach to air quality prediction in Milan through satellite imagery combined with meteorological and morphological data," Ph.D. dissertation, Politecnico di Torino, 2024.

[204] M. Farina et al., "Sparsity in transformers: A systematic literature review," Neurocomputing, p. 127468, 2024.

[205] V. Hassija et al., "Transformers for vision: A survey on innovative methods for computer vision," IEEE Access, 2025.

[206] S. Khan et al., "Transformers in vision: A survey," ACM Computing Surveys, vol. 54, no. 10s, pp. 1-41, 2022.

[207] A. Hatamizadeh, H. Yin, G. Heinrich, J. Kautz, and P. Molchanov, "Global context vision transformers," in Proceedings of the International Conference on Machine Learning, Jul. 2023, pp.

[208] M. Tarasiou, E. Chavez, and S. Zafeiriou, "ViTs for SITS: Vision transformers for satellite image time series," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 10418-10428.

[209] G. Mai et al., "Towards the next generation of geospatial artificial intelligence," International Journal of Applied Earth Observation and Geoinformation, vol. 136, p. 104368, 2025.

[211] Y. Wang and H. A. Karimi, "Advanced deep learning models and algorithms for spatial-temporal data," in Big Data, CRC Press, 2024, pp. 227-253.

[212] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021.