



UGANDA CHRISTIAN
UNIVERSITY

A Centre of Excellence in the Heart of Africa

Department of Computing & Technology

ENHANCING FOOD SECURITY FORECASTING IN UGANDA:
A GEOSPATIAL ANALYTICS FRAMEWORK WITH CROP
YIELD INTEGRATION

By

Philip Arinaitwe

Reg No: M24M19/032, Student No: B29415

E-mail: philip.arinaitwe00@gmail.com, Phone No: +256784982250

A dissertation submitted to the Faculty of Engineering Design & Technology for the study leading to the Degree in partial fulfilment of the requirements for the award of **Master of Science in Data Science and Analytics** of Uganda Christian University

01st September 2026

Declaration

I, Philip Arinaitwe, declare to the best of my knowledge that the work in this dissertation is original and has not been submitted before to any university/institution. It is the result of my effort.



Signature: _____

09/09/2026

Date: _____

Ethical Generative AI Usage Declaration

Students may use generative artificial intelligence (AI) ethically as a “study buddy” or assistant rather than a replacement for critical thinking. This includes citing AI use and fact-checking all outputs to maintain academic integrity.

For purposes of this assignment, the following guidelines shall apply.

- i. Generative AI may be used for brainstorming, and generating ideas to improve your work. You may also use AI for literature search, and summarizing journal papers for quick analysis and assessing relevance of publications. However, once the relevant publications have been identified, a student is expected to fully review the selected publications, identifying problems addressed, methods used, gaps identified and recommendations made.
- ii. Where AI has been used as guided in (i) above, full disclosure of how it has been used should be made in the appendix A, including how AI was used, the AI tools (including mode versions), and a compilation on major AI prompts used.
- iii. NO GENERATIVE AI CONTENT IS ALLOWED IN THE FINAL SUBMISSION. The final submission SHOULD be your original work.

A. Declaring the non-use of AI for content generation

I, Philip Arinaitwe, declare that no content generated by AI technologies has been used in this dissertation.



Signature: _____ Date: 09/09/2026

B. Acknowledging the use of AI as a study buddy

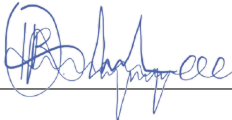
Please refer to Appendix A for the complete declaration of AI usage for generating ideas, brainstorming, literature search, and summarizing only to improve your work.

Approval

This dissertation has been submitted for examination with the approval of the following supervisor.

Ian Raymond Osolo

Department of Computing & Technology
Faculty of Engineering Design & Technology
Uganda Christian University

Signature:  _____

Date: 14 / 09 / 2026

Abstract

Food security is a crucial pillar for human wellbeing, societal stability and national development. Food security can be defined as a situation when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life [1]. In Uganda, nearly 20 percent of the population lives below the poverty line and over 16 million Ugandans face food insufficiency [2] due to various factors like unreliable climate, economic instability, population pressure, conflict and inadequate forecasting mechanisms [3]. Recent research studies have increasingly turned to predictive modelling using machine learning for food security forecasting, integration of climate data, socio-economic indicators and innovative geospatial analytics [4]. Whereas advances in machine learning and geospatial analytics have led to significant improvements in climatic modelling for food security, they often fall short by excluding actual agricultural outputs. Without integrating crop yield data, forecasts may misestimate food availability yet it can be a key driver of food crises. Particularly for Uganda, where over 70% of the population is employed by agriculture [5], the omission of crop yield data from forecasting models can result in inaccurate predictions/forecasts. To address this gap, this study developed and evaluated a geospatial machine learning framework for forecasting food security in Uganda by integrating crop yield data from FAOSTAT with multi-source environmental, market, conflict, demographic and historical Integrated Food Security Phase Classification (IPC) data from 2007 to 2020. A hybrid model combining a Long Short-Term Memory (LSTM) temporal encoder and an XGBoost classifier was trained and validated on 44,082 district-month observations. The hybrid LSTM–XGBoost model achieved a macro F1-score of 0.75 and an overall accuracy of 93.8% on the test set. Importantly, the model demonstrated exceptional performance in identifying severe food crises (IPC 3+), achieving a recall of 0.82, a ROC-AUC of 0.998, and a PR-AUC of 0.852. An ablation study confirmed that incorporating crop yield data significantly improved predictive power, increasing the IPC 3+ F1-score from 0.688 to 0.759. Furthermore, model comparisons demonstrated that a country-specific model trained purely on Ugandan observations significantly outperformed a consolidated regional model shown by a macro F1 score of 0.747 achieved by the country-specific model as compared to the consolidated regional model with a macro F1-score of 0.725. These results provide robust empirical evidence that integrating crop yield data into geospatial frameworks enhances food security early warning systems which enables timely, targeted interventions to mitigate severe hunger.

Dedication

This research is dedicated to my beloved parents and siblings, whose unwavering love, prayers and constant encouragement have been my greatest support.

List of Figures

2.1	2011 East African Drought by FEWS NET.	5
2.2	Indian Ocean Dipole Positive and Negative Phase [6].	6
3.1	Choropleth map showing the distribution of Mean annual rainfall in main drainage sub-basins in Uganda, 1940-2009.	12
3.2	Confusion matrix [7].	15
4.1	Confusion Matrix for the Model Performance.	20
4.2	ROC-AUC and PR-AUC for the different IPC classes.	22
4.3	Ablation Study Results.	24
4.4	SHAP Feature Importance by Category.	27
4.5	Model Comparison: Uganda-specific vs consolidated regional model.	28

List of Acronyms

ACLED	Armed Conflict Location and Event Data
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station
CNN	Convolutional Neural Network
COVID-19	Coronavirus Disease 2019
DL	Deep Learning
ENSO	El Niño Southern Oscillation
FAOSTAT	Food and Agriculture Organisation Statistics
FEWS NET	Famine Early Warning Systems Network
GIS	Geographic Information System
GYGA	Global Yield Gap Atlas
IOD	Indian Ocean Dipole
IPC	Integrated Food Security Phase Classification
IQR	Interquartile Range
KNN	K-Nearest Neighbour
LST	Land Surface Temperature
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MI	Mutual Information
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NDP IV	National Development Plan IV
NDVI	Normalized Difference Vegetation Index
PFC	Predicting Food Crises
PR-AUC	Precision-Recall Area Under the Curve
RF	Random Forest
RMSE	Root Mean Squared Error

ROC-AUC Receiver Operating Characteristic - Area Under the Curve

SDG Sustainable Development Goal

SHAP SHapley Additive exPlanations

SMOTE Synthetic Minority Over-sampling Technique

UBOS Uganda Bureau of Statistics

Contents

Declaration	i
Declaration	ii
Approval	iii
Abstract	iv
Dedication	v
List of Figures	vi
List of Acronyms	vii
Acknowledgments	xi
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Research Objectives	2
1.3.1 Specific Objective 1	2
1.3.2 Specific Objective 2	2
1.3.3 Specific Objective 3	2
1.4 Justification	2
2 Literature Review	4
2.1 Impact of climatic variables on Food Security Forecasting	4
2.2 Geospatial Analytics in Predictive Modelling	6
2.3 Crop Yield Data Integration	6
2.4 Machine Learning and Deep Learning in Food Security Prediction	7
3 Methodology	9
3.1 Study Site	9
3.2 Study Design	9
3.3 Study Population	10
3.4 Sample Size Determination	10
3.5 Data Management	10
3.6 Data Analysis and Preparation	11
3.6.1 Feature Engineering	11
3.6.2 Missing Data Treatment	11
3.6.3 Feature Importance	11

3.6.4	Outlier Treatment	11
3.6.5	Statistical Correlation	11
3.6.6	Machine Learning Algorithms	12
3.6.6.1	LSTM Temporal encoder	13
3.6.7	Performance Measurement and Evaluation	14
3.7	Ethical Considerations	15
4	Results	16
4.1	Introduction	16
4.2	Relationship between Geospatial Variables, Crop Yield and Food Security	16
4.2.1	Bivariate Correlations with Future IPC	16
4.2.2	Relationship between Crop Yield and Food Security	18
4.2.3	Non-linear dependencies	18
4.3	Machine Learning Model Development and Training	19
4.3.1	Data pre-processing	19
4.3.2	Training of the LSTM Encoder	19
4.3.3	Classification performance of the Model	20
4.3.4	ROC-AUC and Precision-Recall AUC	21
4.4	Model Evaluation, Ablation and Baseline Comparison	24
4.4.1	Ablation Study on the Impact of Crop Yield data	24
4.4.2	Baseline Comparison	25
4.4.3	Temporal Cross-validation	26
4.4.4	Feature Importance by Group	27
4.4.5	Model Comparison: Uganda-specific vs consolidated regional model	28
5	Discussion	30
5.1	Relationship between Geospatial variables, Crop yield and Food Security	30
5.2	Model Performance	31
5.3	The Impact of Crop Yield data	32
5.4	Limitations	33
6	Conclusion and Recommendation	34
6.1	Recommendations	34
6.2	Conclusion	35
	REFERENCES	39
	APPENDICES	39
A	Declaration of AI use	40
B	List of 10 Major Journal Publications reviewed	41
C	Research alignment with specialization, SDG, NDPIV & Vision 2040	42
C.1	Alignment with programme specialization	42
C.2	Alignment with Sustainable Development Goals (SDG)	42
C.3	Alignment with the National Development Plan IV (NDPIV)	42
C.4	Alignment with the Uganda Vision 2040	43
D	Research poster layout	44

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Mr. Ian Raymond Osolo, for his invaluable guidance, patience and encouragement throughout this dissertation. I am incredibly fortunate to have benefited from his expertise.

I am also profoundly grateful to my parents and siblings. Thank you for your unconditional love, support and belief in my abilities. Your constant encouragement during the late nights and stressful deadlines provided the foundation upon which this work was built.

This achievement belongs to you as much as it does to me.

Chapter 1

Introduction

1.1 Background

Food security is a crucial pillar for human wellbeing, societal stability and national development.

Food security can be defined as a situation when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life [1]. Food security is related to nutrition, and conversely food insecurity is related to malnutrition. Not all malnourishment arises from food insecurity, as households may have access to healthy diets but choose to eat unhealthily or it may arise from illness. However, in many parts of the world, poverty is linked to poor diets [8].

In Uganda, nearly 20 percent of the population lives below the poverty line and over 16 million Ugandans face food insufficiency [2] due to various factors like unreliable climate, economic instability, population pressure, conflict and inadequate forecasting mechanisms [3].

Recent research studies have increasingly turned to predictive modelling using machine learning models for food security forecasting, integration of climate data, socio-economic indicators, and innovative geospatial analytics [4]. Machine Learning employs statistical and computational techniques to enable algorithms to learn patterns from data and make predictions or decisions without being explicitly programmed for each task. ML systems optimize model parameters by minimizing a loss function, often through gradient-based methods or ensemble approaches, thereby generalizing from observed examples to unseen instances [9].

However, there is still a critical gap in food security prediction frameworks for Uganda, which is, the integration of agricultural production data, in particular, crop yield data. This research aims to develop a more comprehensive and accurate forecasting framework that can contribute to more effective food security forecasting in Uganda through integrating crop yield data to provide better insight into food availability.

1.2 Problem Statement

Current geospatial food security prediction models including regional multi-model frameworks such as that developed by [4] place substantial emphasis on climatic and socio-economic indicators while crop yield which is a direct measure of agricultural output has not been systematically incorporated into these forecasting frameworks. Given that over 70% of Uganda's population is employed in agriculture [5] and that the effect of climatic variability on crop yield differs

across regions and crop types [10], the extent to which the omission of crop yield data affects the accuracy of food security forecasts for Uganda remains insufficiently understood.

There is a need to systematically integrate crop yield data into existing food security prediction models for Uganda, particularly for staple crops such as cassava, bananas and maize and to empirically evaluate whether such integration measurably improves forecasting performance. This study addresses that need by developing a hybrid geospatial machine learning framework that incorporates crop yield data alongside environmental, market, conflict and historical food security indicators and by testing the specific contribution of yield data to forecast accuracy through a controlled ablation study.

1.3 Research Objectives

To optimize food security forecasting in Uganda through the development and implementation of a geospatial analytics framework integrated with crop yield data.

1.3.1 Specific Objective 1

To determine the relationship between key geospatial variables, crop yield and food security indicators.

1.3.2 Specific Objective 2

To develop and train a machine learning model that can forecast food security basing on crop yield data.

1.3.3 Specific Objective 3

To evaluate the performance of the machine learning model and test whether including yield data improves forecasts.

1.4 Justification

Integrating crop yield data into food security forecasts has the potential to enhance early warning, policy response and planning. A forecasting framework that predicts food crises in Uganda more accurately will give humanitarian agencies, policymakers, and farmers a basis to effectively allocate resources and prepare interventions to mitigate food insecurity and reduce the possibility of hunger which aligns with Sustainable Development Goal 2 which is Zero Hunger [11]. It also fosters value addition in agriculture, technology and innovation as well as effective response to climate change in order to reduce its adverse effect on agricultural production, which are some of the main goals of Uganda's National Development IV framework [12]. NDP IV identifies a set of food security commodities that include crops such as maize and bananas to be targeted through area-based planning across diverse agro-ecological zones. By using crop yield data to forecast potential shortages for these staple crops, the proposed framework will equip planners with data-driven insights to meet these targets [12].

Additionally, the project also provides for technological advancement by demonstrating how machine learning can be used to develop more advanced models that can give better and more accurate results compared to traditional forecasting models [4]. NDP IV emphasizes value chain development, mechanization and efficient use of technology in agriculture. The proposed

project embodies this modernization thrust by employing cutting-edge data analytics such as AI and GIS to enhance crop production forecasting.

Academically, this work seeks to extend the geospatial and multi-model approaches of [4] by adding crop yield data which will fill a literature gap on food security forecasting. The creation of a machine learning model further represents an advancement in food security forecasting.

Chapter 2

Literature Review

2.1 Impact of climatic variables on Food Security Forecasting

Variability in climate, especially due to global warming is a primary cause of food insecurity, as it significantly impacts agricultural productivity, market stability and resilience of households [13].

The frequency and occurrence of extreme events has also affected food security in the various regions. Droughts which are characterized by prolonged precipitation deficits and high evapotranspiration rates have been shown to reduce cereal yields by up to 40% in semi-arid regions like the Sahel [14]. Examples include the East African drought in 2011 that led to a 50% decline in maize production in Somalia which in turn triggered famine-level food insecurity [15] and crop failure rates exceeding 80% for maize and sorghum in the Teso and Karamoja subregions that led to acute food insecurity for over 1.6 million in 2016-17 [16].

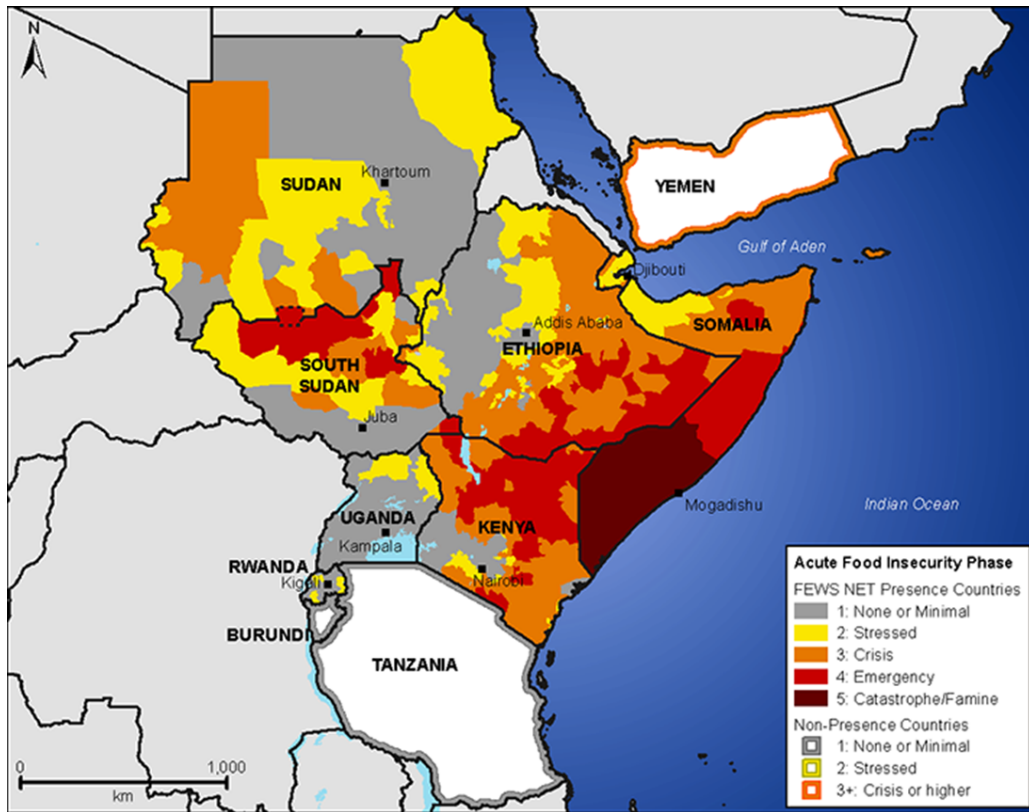


Figure 2.1: 2011 East African Drought by FEWS NET.

Satellite-derived indicators have become crucial for early warning systems. The Normalized Difference Vegetation Index (NDVI) which measures photosynthetic activity has been integrated with Climate Hazards Group Infrared Precipitation with Station (CHIRPS) data to detect vegetation stress up to 12 weeks before harvest failures in Ethiopia [17] [18]. Similarly, Land Surface Temperature (LST) anomalies from MODIS sensors have been used to identify heat stress in India's wheat belt, where a 1°C temperature rise correlates with a 5% yield loss [19] which shows how significant the effect of a slight change in temperature is on crop yield.

Upcoming frameworks also try to incorporate oceanic-atmospheric oscillations, such as the Indian Ocean Dipole (IOD) to predict multi-seasonal rainfall variability in East Africa. Regional rainfall predictions have been integrated and studied together with Indian Ocean Dipole (IOD) and El-Nino Southern Oscillation (ENSO) and the results showcase an improvement in seasonal outlooks for Africa which is an important step towards more accurate multi-seasonal forecasting [20].

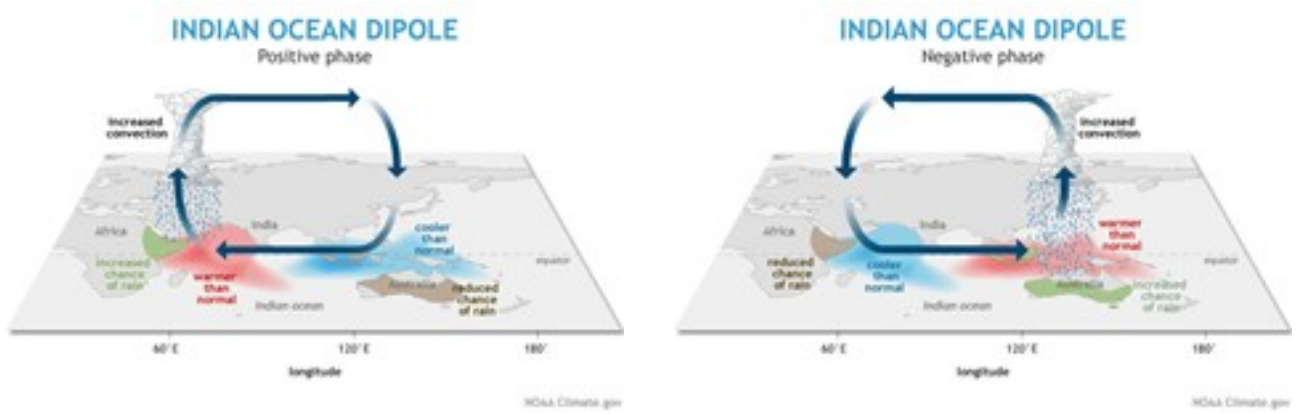


Figure 2.2: Indian Ocean Dipole Positive and Negative Phase [6].

2.2 Geospatial Analytics in Predictive Modelling

Geospatial technologies provide for spatially explicit risk assessments by combining and integrating data from multiple sources such as climatic indicators and socioeconomic factors. Remote sensing platforms like Sentinel-2 (10–20m resolution) provide high-frequency monitoring of the stages of crop growth. GIS-based spatial interpolation techniques such as kriging and spline interpolation address data gaps in regions with few meteorological stations.

In Uganda, studies have shown that non-climatic factors explained more variability in crop yield than rainfall and temperature alone, basing on regression models that have been developed [21]. This necessitates the need for geospatial frameworks that combine both climatic and socioeconomic data.

[4] has extended these approaches by developing a multi-model ensemble that combines climatic indicators with socioeconomic variables to create a more robust framework.

New innovations such as participatory GIS (PGIS) integrate indigenous knowledge into risk models. In Malawi, community mapping was integrated with satellite-derived flood information and hydrological modelling to highlight the value of hybrid top-down/bottom-up approaches [22]. Further, geotagged data on crop health and the spread of crop diseases has been used and tested with cassava in Uganda and this has improved the spatial accuracy of pest and disease models [23].

Hotspot mapping through spatial autocorrelation methods has advanced. This has been used to identify chronic food insecurity clusters such as in the Karamoja and Bukedi sub-regions [4]. However, these rely on static agricultural zones thus overlook the cultivation patterns that are ever changing due to climate adaptation which is addressed in dynamic land-use models such as CLUE-S [24].

2.3 Crop Yield Data Integration

Crop yield data bridges the gap between climatic forecasts and food availability. FAOSTAT provides standardized yield metrics for 245 countries but relies on national agricultural surveys which introduces reporting biases, for example cassava yields in Nigeria were underreported due to informal smallholder markets [25]. In Uganda, crop yield statistics are often aggregated, underreported or sampled and collected without consistency [21]. In order to mitigate this, researchers have fused agricultural census data with satellite proxies. Sentinel-1 SAR data achieved 89% accuracy in estimating rice yields in Vietnam by correlating backscatter signals

from radar remote sensing with panicle density [26] which explains the potential suitability for staple crops like maize and cassava in Uganda. [27] integrated Landsat-8 NDVI with household survey data which greatly improved the prediction of maize yields in the districts of Eastern Uganda.

The use of Machine learning frameworks provides for further enhancement of yield prediction. Convolutional Neural Networks (CNNs) trained on Landsat-8 and soil organic carbon maps predicted U.S. maize yields with a root mean square error (RMSE) of 0.8 t/ha [28].

Yield gap analysis has emerged as a critical tool for identifying underperforming regions. Such analysis requires high-resolution agroclimatic zoning which is shown by the Global Yield Gap Atlas (GYGA). A study carried on staple crops in the Sub-Saharan Africa region showcased significant yield gaps for Uganda’s staple crops, with maize achieving only 30% of its potential yield due to a multitude of factors [29].

2.4 Machine Learning and Deep Learning in Food Security Prediction

Machine learning (ML) and deep learning (DL) have emerged as transformative tools for food security prediction due to their capacity to model nonlinear relationships and high-dimensional datasets. Random Forest (RF) and XGBoost algorithms, for example, have achieved high accuracy in predicting crop failures by synthesizing climatic, topographic and phenological variables [28] [30].

Deep learning approaches, in particular, Long Short-Term Memory (LSTM) networks, excel in capturing temporal dependencies in climatic time series, such as lagged effects of rainfall on soil moisture [31]. Transformers which were initially developed for natural language processing are now being adapted for spatial-temporal forecasting where attention mechanisms are being leveraged to prioritize critical climatic variables [32]. However, ML/DL models face challenges in low-data environments which are usually associated with over-fitting and poor generalizability.

Transfer learning solves the generalization by pre-training models on global datasets before fine-tuning them with local data. [33] pre-trained a deep learning model on a large global dataset, after which it was fine-tuned using locally collected data to make it more accurate and robust. Ethical considerations such as algorithmic bias bring in the need for transparent model architectures and participatory evaluation [34].

Despite the growing adoption of transformer architectures for time series and sequence modelling tasks [33], studies suggest that their advantages are not universal, particularly in data-constrained settings. [35] evaluated several state-of-the-art transformer models for long-term time series forecasting and found that substantially simpler linear models frequently achieved superior performance across multiple benchmark datasets. They argued that, unlike recurrent models, transformers do not inherently preserve temporal order, relying instead on positional encodings to recover sequential information, which may limit their effectiveness for certain forecasting tasks. Similarly, [36] reported that transformer-based forecasting models lack the strong temporal inductive bias naturally provided by recurrent and convolutional architectures. As a result, transformers must learn temporal dependencies directly from the training data which makes them more susceptible to underperformance when training datasets are comparatively small.

These observations are consistent with broader findings that transformer architectures which

were initially developed for natural language processing applications trained on extremely large data, generally require substantially larger datasets than recurrent models to fully realise their representational advantages [33]. Consequently, their superior performance is not guaranteed in applied forecasting problems where training data is comparatively limited.

[37] in a comprehensive study covering 45 datasets demonstrated that gradient-boosted tree ensembles including XGBoost consistently outperformed deep learning models including attention-based tabular architectures on medium-sized datasets. This was attributed primarily to the stronger inductive bias of tree-based methods for structured tabular data rather than to inherent limitations of deep learning itself.

Machine learning excels in modelling non-linear relationships that are complex and exist in food systems. Random Forest algorithms use decision trees to handle high-dimensional data and have been widely adopted. XGBoost is a scalable tree-based gradient-boosting framework that has also been adopted heavily [38].

Deep learning models, particularly Long Short-Term Memory (LSTM) networks capture temporal dependencies in climatic time series.

However, ML/DL models face a number of ethical and operational challenges. First is algorithmic bias which occurs when the outputs of an algorithm benefit or disadvantage certain individuals or groups more than others without a justified reason for such unequal impacts [39], for example models trained on global datasets may underperform in certain regions.

Another is explainability which comes about due to a difference between what a model can explain and what a decision maker wants to know. These issues also arise for well-functioning models that fail on few data instances [30] [31].

Finally, there is scalability which refers to the ability of a system to accommodate an increasing number of elements or objects, process growing volumes of work gracefully, and/or be susceptible to enlargement [40].

Chapter 3

Methodology

3.1 Study Site

This study was carried out in Uganda, which is situated in East Africa and is characterized by a tropical climate with bimodal rainfall patterns, that run from March–May and September–November and are critical for its economy that’s dominated by agriculture. Agriculture employs over 70% of the country’s population [9]. The country’s agro-ecological zones experience significantly different climate conditions which in turn influence crop production and food security outcomes [41]. Focus districts were selected based on their prominence in staple crop production and vulnerability to climatic changes.

3.2 Study Design

This study adopted a quantitative, correlational and predictive design. It integrated multi-source geospatial and agricultural datasets. The framework was built on methodologies from Atuhaire et al. [4] who combined geo-intelligence and machine learning (ML) for food security forecasting and extends this through the integration of crop yield.

The study used secondary data. Crop yield data was obtained from the Food and Agriculture Organisation Statistics (FAOSTAT) Database [42]. For Food Security data, the Predicting Food Crises 2020 dataset was used [43].

A three-phase design was used, corresponding to the specific objectives of the study. Phase I examined the bivariate and multivariate relationships between geospatial variables, crop yield data and food security indicators.

Phase II involved the development and training of a machine-learning model that combined a Long Short-Term Memory (LSTM) encoder with an XGBoost classifier. The LSTM was tasked with learning the temporal patterns from the geospatial data with the final multi-class IPC phase classification performed by the XGBoost classifier [31] [44].

Phase III evaluated the performance of the model through temporal cross-validation, baseline comparisons and ablation studies in order to understand the specific contribution of crop yield data to the accuracy of food security forecasting as recommended by Roberts et al. [44].

3.3 Study Population

The study population comprised 56 districts across Uganda with validated agricultural, climatic and food security data from 2009–2020. District observations by month were used, each characterised by a combination of geospatial, demographic, conflict, market attributes and lag features, obtained from a combination of two datasets namely:

The Predicting Food Crises 2020 dataset is a global dataset from the World Bank that is made up of data from 1162 districts in 21 developing countries from the period 2009 to 2020 [45]. The dataset is made up of a target variable for predicting food security on an ordinal scale with five stages: (1) minimal, (2) stressed, (3) crisis, (4) emergency and (5) famine.

The second dataset is from the Food and Agriculture Organisation Statistics (FAOSTAT) Database showing food production and yield, collected annually in 245 countries from 2009 to 2020. Crop and livestock statistics for 278 products in different categories are maintained in this dataset [42]. The two datasets were merged using country-year as the primary key to create a unified data frame.

For the final analysis, district-month observations were used with the target variable defined as the IPC phase in the three months ahead ($t + 3$), maintained as 1–3 in order to focus on the range within which Uganda lies.

3.4 Sample Size Determination

A total of about 44082 records were derived which aligns with sample size recommendations for ML models in geospatial studies [46]. In order to handle the categorical imbalance for the target variable, the Synthetic Minority Over-sampling Technique (SMOTE) was used.

For the LSTM encoder, a minimum of 12-month sequence was required per district. For the XGBoost, a minimum of 10–20 predictor events were required as proposed by Riley et al. [47]. The sample comfortably satisfied this threshold with 139 features for use after feature engineering was done. Further, an 18-month window was ultimately adopted, exceeding this minimum, in order to capture both annual rainfall seasons within a single sequence

For the temporal cross-validation, the sample was partitioned into temporal folds with a minimum size of 5% of Uganda’s observations, in order to ensure the observations were sufficient for stable performance estimation [48].

3.5 Data Management

This project utilized secondary data provided on the World Bank Repository and Food and Agriculture Statistics (FAOSTAT) database. The datasets were stored in Comma-Separated Value format spreadsheets which are compatible with several data analytics tools.

The analytics were carried out in Python using a Google Colab notebook with the packages Pandas for data manipulation, NumPy for numerical operations, scikit-learn for baseline models and preprocessing, XGBoost for gradient boosting, PyTorch for the LSTM encoder, SHAP for model interpretability and GeoPandas for the geospatial analysis.

3.6 Data Analysis and Preparation

3.6.1 Feature Engineering

Feature engineering was used to preserve all 28 raw PFC attributes while creating derived features for use in predictive performance improvement. The engineering pipeline followed a tiered structure.

For each district, temporal features were created. Lagged values were generated for IPC phases, environmental variables such as NDVI, conflict indicators and staple food prices. Next, IPC-specific features were derived such as IPC trend, velocity, stress flags, phase change indicators and recovery signals.

Interaction features were created to capture compound risk scenarios for example ($\text{NDVI} \times \text{rainfall}$) to capture the vegetation-water stress.

From FAOSTAT, raw yield values for maize, beans and bananas were supplemented with yield shock indicators made up of z-scores relative to the expanding mean.

K-nearest-neighbour spatial lags were computed for key variables using district centroids in order to capture spatial autocorrelation in food security outcomes.

The final feature set comprised 139 engineered features, all traceable to the original 28 attributes from the datasets.

3.6.2 Missing Data Treatment

Missing data which is a common challenge in multi-source geospatial datasets, was addressed through a tiered approach. K-Nearest Neighbor (KNN) imputation was applied using scikit-learn's KNNImputer, which estimated missing values based on similarity metrics across neighbouring observations [49], together with Simple statistical techniques like mean, median and mode imputation.

3.6.3 Feature Importance

SHAP (SHapley Additive exPlanations) was used to determine feature importance as it attaches a value to each variable that was used in the model's development. This value is called the Shapley Value which is a fair profit allocation among many stakeholders basing on their contribution [50].

3.6.4 Outlier Treatment

Outliers which can distort model performance, were identified from the data using the Interquartile Range (IQR) method which flagged values that exhibited a normal distribution. Mean and standard deviation were also used to flag those values that exhibited a skewed distribution.

3.6.5 Statistical Correlation

Pearson and Spearman correlation analyses were carried out to quantify linear and non-linear relationships between variables. Visualizations such as heat maps were generated using Seaborn and matplotlib to provide a graphical representation of the interaction between variables. Choropleth maps were also developed using GeoPandas.

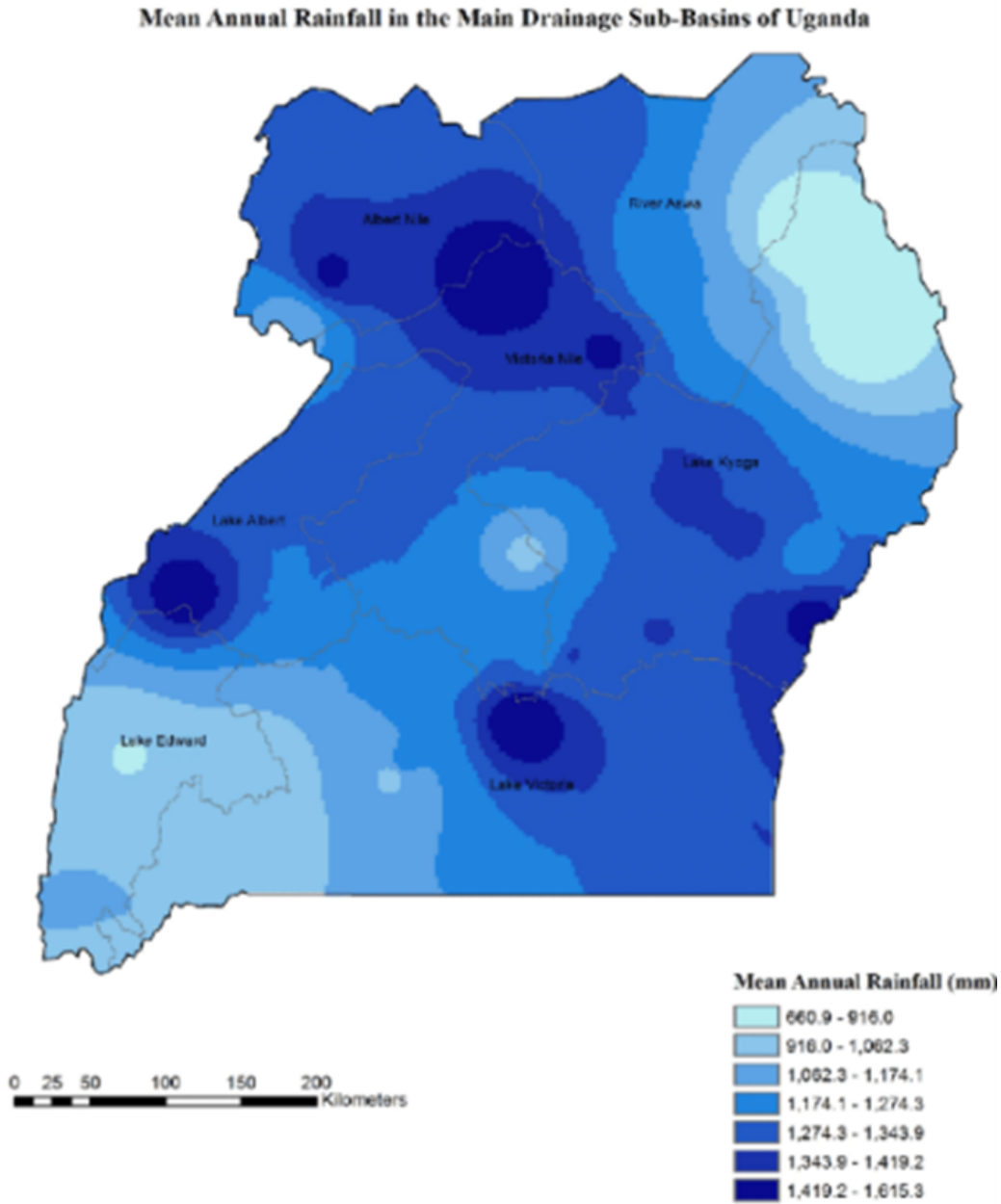


Figure 3.1: Choropleth map showing the distribution of Mean annual rainfall in main drainage sub-basins in Uganda, 1940-2009.

3.6.6 Machine Learning Algorithms

A number of machine learning (ML) and deep learning (DL) algorithms were selected to balance interpretability and predictive power. XGBoost, a gradient-boosting framework optimized for high-dimensional data was used alongside a Deep learning 2-layer Long Short-Term Memory (LSTM) network model which captured temporal dependencies in climate-yield sequences [32] from the 18-month sequence of engineered features on which it was trained. Class weights were used to address the imbalance in classes. The XGBoost classifier was tasked with predicting amongst three classes, IPC 1, 2 and 3.

The hybrid LSTM-XGBoost architecture was selected because it better matched the characteristics of the dataset than a transformer-based alternative. LSTM networks are well established for capturing non-linear temporal dependencies in climate and crop yield time series [32], while XGBoost consistently performs well on high-dimensional heterogeneous tabular data and has

been widely applied in crop yield and food security prediction [30][31][39]. Accordingly, the LSTM, augmented with an additive attention mechanism, was used to learn temporal characteristics from 18-month sequences, while XGBoost performed the final IPC classification using the learned embeddings together with the engineered static features.

A transformer-based encoder, using self-attention in place of recurrence, was considered as an alternative to the LSTM but was not adopted due to the nature of the dataset [32]. Although transformer architectures have demonstrated state-of-the-art performance across a wide range of sequential learning tasks through their self-attention mechanism, they generally require larger training datasets and greater computational resources than recurrent models to achieve stable generalisation, particularly in time-series applications, due to their larger parameter space and weaker sequential inductive bias [51][35]. The approximately 43,000 training sequences available in this study, together with an 18-month input window, were therefore considered more appropriate for a compact bidirectional LSTM, whose recurrent structure is well suited to modelling relatively short temporal sequences [32].

Moreover, the principal advantage of transformers lies in capturing long-range temporal dependencies through global self-attention [32][35], whereas the temporal dependencies in this study could be effectively represented using an LSTM augmented with additive attention. Restricting the deep learning component to feature extraction also allowed the final classification to remain within the XGBoost framework, preserving robust class-imbalance handling, ordinal-aware training and SHAP-based interpretability, while reducing the computational burden associated with training and tuning a full transformer model. This made the hybrid LSTM–XGBoost architecture a more practical, interpretable and computationally efficient choice for the study.

3.6.6.1 LSTM Temporal encoder

The temporal encoder was implemented as a two-layer, bidirectional Long Short-Term Memory (LSTM) network with 48 hidden units per direction per layer, followed by an additive (Bahdanau-style) attention mechanism applied over the sequence of hidden states and a projection head producing a fixed-length embedding per district-month observation. During encoder training, the embedding was passed through an auxiliary linear classification head predicting the three IPC classes directly, so that the encoder learned temporally and class-discriminative representations. Only the embedding itself was retained and passed forward to the XGBoost classifier once training was complete.

Each input sequence spanned 18 consecutive monthly observations per district, zero-padded at the start where fewer than 18 prior months were available. Each time step was described by 54 engineered temporal features, drawn from a candidate list of 55 covering IPC history and trend indicators, environmental variables, conflict indicators, interaction terms, crop yield and yield-shock variables, market indicators, spatial lag features and static geospatial and demographic attributes.

This produced 44,082 candidate sequences, of which 43,245 carried a valid training target. An 18-month window was chosen in preference to the 12-month minimum identified in the literature, in order to allow the encoder to observe both of Uganda’s bimodal rainfall seasons (March–May and September–November) at least once within a single input sequence, together with a preceding buffer period, rather than the single-season exposure a strict 12-month minimum would provide.

The bidirectional LSTM used 48 hidden units per direction across two stacked layers, with recurrent dropout of 0.2 applied between layers. The attention layer computed a scalar score for each time step via a linear projection of the concatenated (forward and backward) hidden

state, normalised across the sequence with a softmax function and used to compute a weighted sum of hidden states as the sequence’s context vector. This allowed the encoder to assign greater importance to the time steps most predictive of the eventual food security outcome, rather than treating all 18 months equally. The context vector was passed through a projection head consisting of layer normalisation, dropout of 0.2 to prevent over-fitting, a linear layer and a Rectified Linear Unit (ReLU) activation function to produce the final 24-dimensional embedding.

A 24-dimensional output was chosen to provide a sufficiently expressive summary of temporal dynamics without allowing the temporal embedding to dominate the substantially larger set of remaining static and engineered tabular features once passed to the XGBoost classifier.

The encoder was trained for 30 epochs using the AdamW optimiser with a learning rate of 1×10^{-3} , weight decay of 1×10^{-4} batch size of 256 and a cosine-annealed learning rate schedule decaying from 1×10^{-3} to approximately zero over the course of training. Gradients were clipped to a maximum norm of 1.0 to stabilise training. The loss function was class-weighted cross-entropy, with weights computed as the inverse class frequency further boosted by manually specified factors of 1.0, 1.8 and 3.0 for IPC 1, 2 and 3+ respectively, consistent with the class-weighting scheme used in the XGBoost classifier. Training loss declined from 1.003 at epoch 1 to 0.920 at epoch 30.

SMOTETomek oversampling was applied to the training set, adjusted to the minority class size, with severe misclassification punished.

3.6.7 Performance Measurement and Evaluation

Four models were compared as part of the final evaluation: a naive baseline predicting no change, Logistic Regression with lag features, class-weighted Random Forest made up of 500 trees and the full hybrid LSTM-XGBoost model.

Model performance was evaluated using regression and classification metrics. For crop yield and IPC score prediction (regression tasks), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the coefficient of determination (R^2) were calculated, with lower RMSE/MAE and higher R^2 indicating superior performance [52]. For IPC crisis classification, precision, recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used to account for class imbalance. Bootstrap 95% Confidence Intervals were generated for all metrics in order to quantify uncertainty [53].

Stratified 3-fold cross-validation which ensures robustness and fine-tuned hyperparameters was used to balance model complexity and generalization. The confusion matrix measured the performance of the model based on a set of test data. It displays several accurate and inaccurate (True Positives, False Positives, True Negatives and False Negatives) using the model’s prediction [54].

The ablation study was conducted to quantify the marginal contribution of crop yield information to food security forecasting performance independently of all other engineered feature groups. Yield-related variables were identified programmatically by matching feature names against predefined naming patterns associated with crop yields and yield anomalies ensuring that all yield-derived variables including raw crop yields, yield-shock indicators and crop-specific interaction terms were consistently identified. This process classified 14 of the 139 engineered features as yield-related, with the remaining 125 features forming the reduced feature set used for comparison.

To minimise dependence on a single train–test split, the evaluation was repeated across three

independent temporal splits. The original training cutoff date (7 April 2017) was randomly perturbed by an offset between -30 and $+30$ days for each run which generated three closely related but distinct temporal partitions. For each split, observations prior to the perturbed cutoff were used for training while observations from the cutoff onward formed the test set.

For each temporal split, two XGBoost classifiers were trained independently. One was trained using the complete 139-feature dataset and the other using the reduced 125-feature dataset with all yield-related variables removed. Both models employed identical hyperparameters, class-weighting strategies and decision thresholds in order to ensure that crop yield information was the only experimental variable. Performance was evaluated on the common held-out Uganda test set using macro F1-score, IPC 2 F1-score, IPC 3+ F1-score and overall accuracy.

The statistical significance of the observed performance differences was assessed using paired t -tests, comparing the *withyield* and *withoutyield* configurations across the three matched temporal splits for each evaluation metric. Statistical significance was assessed at the conventional $\alpha = 0.05$ level.

After training, the model was evaluated using temporal cross-validation which provides for multiple (t) time validation of the performance of the model. First, the dataset was split into subsets with past observations used to train the model and then future used to evaluate its performance [55].

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Figure 3.2: Confusion matrix [7].

3.7 Ethical Considerations

The study used secondary data accessed from sources that are available to the public, with the sources properly cited.

Chapter 4

Results

4.1 Introduction

The findings of this study were obtained from data pre-processing, exploratory data analysis, model development, evaluation and ablation study. Further, these findings aim to meet the objectives of the research described above.

4.2 Relationship between Geospatial Variables, Crop Yield and Food Security

4.2.1 Bivariate Correlations with Future IPC

The study used Spearman Rank (ρ) and Pearson (r) coefficient correlation to determine the relationship between each predictor feature and the Future Integrated Phase Classification (IPC) outcome in three months ($t + 3$). The variables were classified into five distinct groups namely Geospatial, Conflict, Market, Crop Yield and IPC History.

Table 4.1: Spearman rank and Pearson correlation between selected predictor variables and the Future three-month IPC phase.

Category	Variable	Spearman (ρ)	Pearson (r)	p -value	Significance
Geospatial	spatial_fews_ipc	0.769	0.774	0.0000	Significant
Geospatial	pasture_pct	0.511	0.496	0.0000	Significant
Geospatial	ndvi_mean	-0.432	-0.420	0.0000	Significant
Conflict	acled_fatalities_max6	0.228	0.126	0.0000	Significant
Conflict	acled_fatalities_sma3	0.181	0.090	0.0000	Significant
Conflict	acled_count_max6	0.135	0.119	0.0000	Significant
Crop Yield	beans_shock	-0.219	-0.147	0.0000	Significant
Crop Yield	maize_yield	-0.200	-0.207	0.0000	Significant
Crop Yield	bananas_yield	-0.147	-0.155	0.0000	Significant
Market	p_staple_food	0.183	0.394	0.0000	Significant
Market	price_diff_6	0.252	0.373	0.0000	Significant
Market	price_diff_3	0.233	0.339	0.0000	Significant
IPC History	fews_ipc_adjusted	0.888	0.877	0.0000	Significant
IPC History	ipc_lag_1	0.867	0.860	0.0000	Significant
IPC History	ipc_roll3_mean	0.866	0.857	0.0000	Significant

From Table 4.1, **fews_ipc_adjusted** had a strong positive relationship with future IPC of 0.888 which suggests that regions surrounding areas experiencing high food insecurity are more likely to experience food insecurity as well in the subsequent period. **ipc_lag_1** is the IPC level from the previous period and its strong positive correlation with future IPC of 0.867 represents that current temporal events are highly predictive of near-future conditions.

Amongst the geospatial variables, the positive correlation of 0.511 between **pasture_pct** and future IPC indicates that areas dominated by pasture lands are more likely to experience higher IPC levels in the future. **ndvi_mean** is derived from the Normalised Difference Vegetation Index and is an indicator of the vegetation health. Its negative correlation of -0.432 with Future IPC indicates that healthier vegetation is associated with lower future IPC levels.

For the conflict variables, **acled_fatalities_max6** measures the maximum number of fatalities recovered over a 6-month period as a result of conflict. Its positive correlation of 0.228 with future IPC signifies an increase in fatalities results into higher future IPC, therefore signifying food insecurity. **acled_fatalities_sma3** represents the three-month rolling average of fatalities due to conflict. Its positive relationship of 0.181 with future IPC signifies that sustained conflict incidents result in higher future IPC thereby resulting in food insecurity.

For the market variables, **p_staple_food** represents the price of the staple food. The positive correlation of 0.183 with future IPC signifies that increasing food prices are associated with higher future IPC phases, signifying food insecurity as purchasing power reduces especially amongst low-income households. **price_diff_6** represents the six-month price difference. Its positive correlation of 0.252 with future IPC shows that sustained increases in staple food prices are associated with higher future IPC leading to food insecurity.

Amongst the crop yield variables, **maize_yield** and **bananas_yield** have negative correlations of -0.200 and -0.147 respectively which signifies that an increase in yield for either crop is associated with lower future IPC phases which signifies food security.

4.2.2 Relationship between Crop Yield and Food Security

Partial correlations were computed to identify the direct relationship between crop yield and future IPC, excluding the influence of variables like NDVI, rainfall, conflict count and seasonality. Regression analysis was conducted for both the crop yield and future IPC to exclude the influence of the other variables and identify the unique correlation between crop yield and future IPC.

Table 4.2: Partial correlations between crop yield variables and IPC 3-months in the future.

Yield Variable	Partial r	p -value
maize_yield	-0.269	0.0000
beans_yield	0.105	0.0000
bananas_yield	-0.198	0.0000
maize_shock	0.009	0.0648
beans_shock	-0.167	0.0000
bananas_shock	-0.096	0.0000

Maize yield has a negative correlation of -0.269 with future IPC which signifies that an increase in maize yield is associated with a lower future IPC. The p -value of 0.0000 also indicates that maize yield is a significant predictor of future IPC. Banana yield also follows suit with a negative correlation of -0.198 with future IPC which shows that an increase in banana yield is associated with a lower future IPC. The p -value of 0.0000 also indicates that banana yield is a significant predictor of future IPC. Beans yield exhibits a positive correlation of 0.105 with future IPC after regression which points towards other underlying factors that may influence future IPC, regardless of the yield.

Bean shock and Bananas shock also have statistically significant impacts on the prediction of future IPC as shown by the p -value of 0.0000 and the negative correlations of -0.167 and -0.096 with future IPC respectively. On the other hand, maize shock exhibits a non-significant relationship with future IPC with a p -value of 0.0648 which points towards the possibility that the variability in maize production is influenced by some of the variables that were isolated during the regression analysis.

4.2.3 Non-linear dependencies

Mutual Information (MI) scores were computed to capture non-linear dependencies between the variables and future IPC.

Table 4.3: Non-linear dependencies between future IPC and select variables using Mutual Information (MI) score.

Variable	Category	Mutual Information (MI)
cropland_pct	Geospatial	0.4426
cropland_x_rugged	Geospatial	0.4411
pasture_pct	Geospatial	0.4402
acled_fatalities_max6	Conflict	0.0619
acled_fatalities_sma3	Conflict	0.0317
beans_shock	Crop Yield	0.3335
maize_shock	Crop Yield	0.3165
bananas_shock	Crop Yield	0.2825
p_staple_food	Market	0.1333
price_diff_6	Market	0.1092
price_diff_3	Market	0.1000
fews_ipc_adjusted	IPC History	0.6010
ipc_lag_1	IPC History	0.5511

IPC History variables like `fews_ipc_adjusted` have the highest MI scores which signifies that IPC History is a substantial non-linear predictor of future IPC. Geospatial variables also exhibited high MI scores which pointed towards the strong influence of geographical regions towards food security predictions.

Amongst the crop yield variables, the shock variables have higher MI scores than the actual yield variables which suggested that yield deviations from the historical values appear to capture non-linear food security dynamics more effectively than the raw yield values.

4.3 Machine Learning Model Development and Training

4.3.1 Data pre-processing

The final dataset comprised 44,082 district-month observations with 147 columns. Amongst these, 84 columns required imputation during different stages of the analysis.

The temporal split resulted in 25,480 rows for training and 1,736 rows for testing. SMOTE-Tomek was applied to the training dataset to deal with the class imbalance which resulted in the training set being expanded to 44,776 rows.

Due to the complex nature of food security trajectories, a Long Short-Term Memory (LSTM) encoder was used.

4.3.2 Training of the LSTM Encoder

An 18-step Long Short-Term Memory (LSTM) encoder with 54 input features was pre-trained to extract 24 district-level embeddings that were subsequently used as part of the features for the XGBoost model. The network was trained for 30 epochs with the training loss declining

from 1.0026 at epoch 1 to 0.9203 at epoch 30 with the cosine annealing learning rate reducing from 0.001 to 0.000.

4.3.3 Classification performance of the Model

The hybrid LSTM-XGBoost model was evaluated on the Ugandan test dataset split made up of 1,736 district-level observations from April 2017 onwards.

Table 4.4: Evaluation results of the hybrid LSTM-XGBoost model.

Class	Precision	Recall	F1-Score	Support
IPC 1 (Minimal)	0.99	0.93	0.96	1610
IPC 2 (Stressed)	0.41	0.76	0.54	104
IPC 3+ (Crisis)	0.69	0.82	0.75	22
Accuracy			0.938	1736
Macro Average	0.70	0.84	0.75	1736

The model achieves a high performance on the Majority IPC 1 Class with an F1-score of 0.96. It also performs very well on the IPC 3+ Crisis class with an F1-score of 0.75 and a recall of 0.82 which signifies that the model correctly identifies 82% of the crisis food security classifications. For IPC 2, the model has an F1-score of 0.54 which reflects the challenge of distinguishing marginal food stress from minimal food insecurity with precision 0.41. The recall of 0.76, however, shows that the model errs towards detecting stress rather than missing it completely.

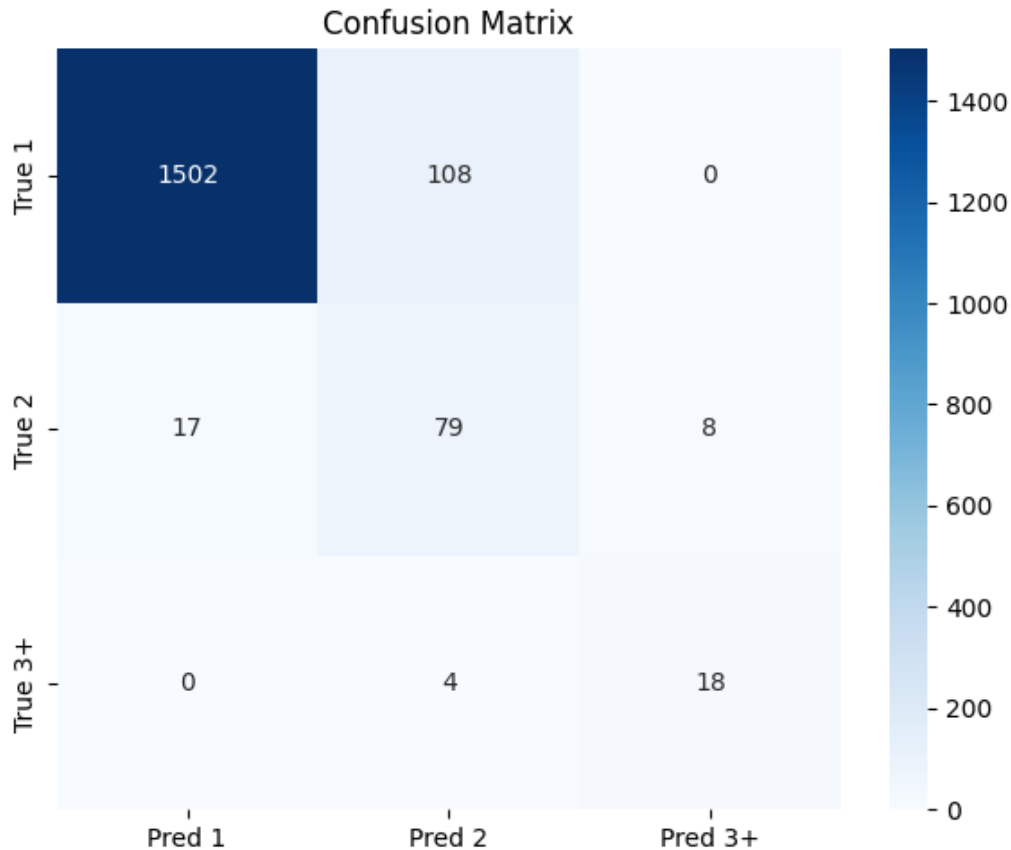


Figure 4.1: Confusion Matrix for the Model Performance.

To assess the stability and reliability of the classification metrics, bootstrap resampling with 500 iterations was performed to generate 95% confidence intervals for precision, recall, and F1-scores. For IPC 2 (Stressed), the F1-score ranged from 0.463 to 0.594, with precision between 0.338 and 0.476 and recall between 0.681 and 0.836. These intervals suggest that while the model consistently identifies a large proportion of stressed cases (high recall), its precision is more variable, indicating some uncertainty in distinguishing stressed households from other IPC categories.

For IPC 3+, the F1-score confidence interval ranged from 0.590 to 0.866, with precision between 0.507 and 0.882 and recall between 0.625 and 0.955. The wider intervals observed for this class are expected given the relatively small number of crisis observations in the test set at 22, giving way to greater variability.

The lower bound of the IPC 3+ F1-score confidence interval, 0.590 is substantially higher than what would be expected from random or chance-level classification, indicating that the model retains meaningful predictive capability for crisis-level food insecurity events. Furthermore, the consistently high recall estimates suggest that the model is able to identify the majority of crisis cases which is an important characteristic in food security early warning systems where failing to detect severe events may have significant humanitarian consequences.

Overall, the bootstrap analysis demonstrates that the model’s performance is robust across repeated resamples of the data. Although predictions for the rare IPC 3+ class exhibit greater uncertainty than those for the other classes, the confidence intervals provide evidence that the model captures genuine predictive signals rather than attributes of a particular train-test split which increases confidence in the model’s ability to support early identification of deteriorating food security conditions in Uganda.

4.3.4 ROC-AUC and Precision-Recall AUC

Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves were generated using a one-vs-rest strategy for each IPC class based on the isotonic-calibrated class probabilities produced by the final XGBoost classifier on the held-out Uganda test set made up of 1,736 district-month observations. Prior to evaluation, probability calibration was performed using isotonic regression through `CalibratedClassifierCV`. The trained XGBoost model was first fitted on 80% of the resampled training data, with the remaining 20% reserved as an independent calibration set. All reported class probabilities, ROC curves, PR curves and the reliability diagram were generated from the calibrated model rather than the raw XGBoost predictions.

For each IPC class, the true labels were binarised using a one-vs-rest approach and ROC and PR curves were computed by varying the classification threshold across the full range of calibrated predicted probabilities using the `roc_curve` and `precision_recall_curve` functions implemented in scikit-learn. Model calibration was further assessed using a reliability diagram which was constructed by partitioning predicted probabilities into ten equal-width bins and comparing the mean predicted probability within each bin to the observed proportion of positive outcomes against the ideal calibration line.

Predictive discrimination was quantified using the Area Under the Receiver Operating Characteristic Curve (ROC–AUC) and the Area Under the Precision–Recall Curve (PR–AUC). ROC–AUC measures the model’s ability to distinguish between classes across all classification thresholds while PR–AUC provides a more informative assessment of performance for imbalanced datasets by emphasising precision and recall for the positive class.

Table 4.5: ROC-AUC and PR-AUC for each IPC Class.

Class	ROC-AUC	PR-AUC	Support
IPC 1 (Minimal)	0.972	0.997	1610
IPC 2 (Stressed)	0.956	0.611	104
IPC 3+ (Crisis)	0.998	0.852	22
Macro-average	0.975	0.820	1736

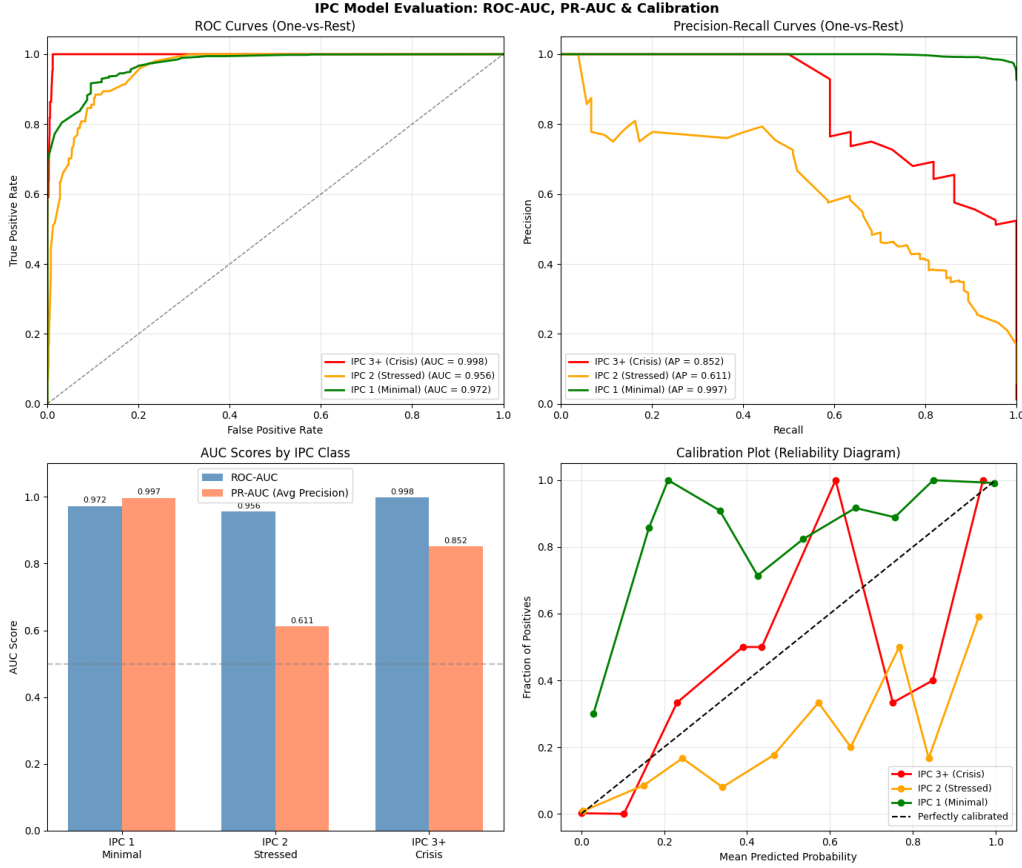


Figure 4.2: ROC-AUC and PR-AUC for the different IPC classes.

The class-specific ROC-AUC and PR-AUC results demonstrate that the proposed model effectively discriminates between different levels of food insecurity. For the crisis class IPC 3+, the ROC-AUC of 0.998 indicates near-perfect discrimination, which suggests that the model almost always assigns higher probability scores to true crisis observations than to non-crisis observations. This reflects a strong ability to rank districts according to their likelihood of experiencing severe food insecurity.

The corresponding PR-AUC of 0.852 further demonstrates strong predictive performance for IPC 3+, indicating that the model maintains a favourable balance between identifying true crisis events and limiting false alarms. This is particularly important in food security early warning systems where missed crisis events may delay humanitarian intervention while excessive false alarms can result in the inefficient allocation of limited resources.

For IPC 2, the ROC-AUC remained high at 0.956 which indicates good discriminatory ability. However, the corresponding PR-AUC of 0.611 was considerably lower than that of the other classes which suggests that although the model generally distinguishes stressed observations

from other IPC categories, it experiences greater difficulty in maintaining both high precision and high recall. This is consistent with IPC 2 representing a transitional state between minimal and crisis food insecurity where substantial overlap exists in both predictor patterns and predicted probabilities, making this class the most susceptible to misclassification.

The IPC 3+ ROC-AUC of 0.998 and PR-AUC of 0.852 should nevertheless be interpreted with caution because these metrics are based on only 22 positive crisis-class observations in the test set. Performance estimates derived from such a limited number of positive cases are inherently subject to high variance as a small number of additional correct or incorrect rankings could substantially alter the reported values. The bootstrap confidence interval for the IPC 3+ with F1-score range of 0.590–0.866 indicates that the model retains meaningful predictive capability well above chance, although the point estimates should not be interpreted as precise measures of population-level performance.

A distinction should also be made between the model’s discrimination ability assessed using ROC-AUC and PR-AUC and its probability calibration evaluated using the reliability diagram. The probabilities used for both evaluations were obtained after isotonic recalibration rather than directly from the raw XGBoost outputs. Despite this recalibration, the reliability diagram exhibits noticeable deviations from the ideal calibration line for IPC 2 and IPC 3+ especially in probability bins containing relatively few observations. This behaviour is most likely as a result of the limited number of samples of the minority-class available for fitting the isotonic calibration model rather than a limitation of the calibration method itself. Consequently, while the model demonstrates excellent capability in ranking districts according to relative food security risk, the predicted probabilities for the minority classes should be interpreted cautiously. For early warning systems, the model exhibits high risk prioritisation as compared to precise probabilistic forecasts.

It should also be noted that the classification results reported in Table 4.5 and Figure 4.2 were not obtained by assigning each observation to the class with the highest predicted probability. Instead, class-specific decision thresholds were optimised for the minority classes using the F-beta score ($\beta = 1.5$) which deliberately places greater emphasis on recall than precision. This threshold optimisation reflects the operational objective of minimising missed crisis detections within an early warning framework and therefore contributes to the strong recall observed for IPC 2 and IPC 3+ alongside the predictive capability of the underlying classifier.

Across all classes, the macro-average ROC-AUC of 0.975 and macro-average PR-AUC of 0.820 indicate consistently strong predictive performance when each IPC category is weighted equally. Importantly, the high macro-average PR-AUC demonstrates that the model performs well not only for the dominant IPC 1 class but also for the less frequent stressed and crisis categories. This is particularly significant because the practical value of a food security forecasting system depends not only on correctly identifying food-secure populations but also on reliably detecting districts at risk of deteriorating food security conditions.

Overall, these findings demonstrate that the proposed hybrid LSTM–XGBoost framework captures meaningful temporal and spatial patterns associated with future food insecurity and provides strong discriminatory performance across all IPC classes. Although the calibration of minority-class probabilities remains constrained by the limited number of crisis observations, the model exhibits excellent ranking capability and high sensitivity for severe food insecurity events which supports its suitability as a decision-support tool for food security early warning.

4.4 Model Evaluation, Ablation and Baseline Comparison

4.4.1 Ablation Study on the Impact of Crop Yield data

An ablation study was carried out across three random train-test splits to test whether including crop yield data improves food security forecasts.

Table 4.6: Ablation study results comparing the performance of the model with and without crop yield included.

Metric	With Yield	Without Yield	Δ Mean	p -value	Significant
macro_f1	0.763	0.729	0.034	0.0073	Yes
ipc2_f1	0.560	0.530	0.030	0.0635	No
ipc3_f1	0.759	0.688	0.071	0.0000	Yes
accuracy	0.937	0.934	0.003	0.2020	No

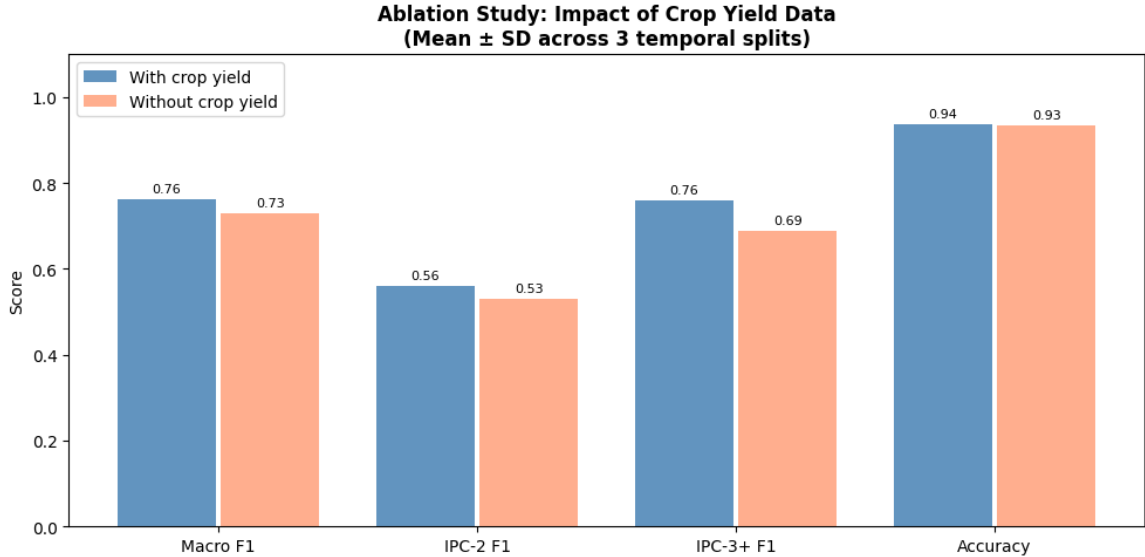


Figure 4.3: Ablation Study Results.

The results indicate that incorporating crop yield improved several key performance metrics. The macro F1-score increased from 0.729 to 0.763, with the improvement being statistically significant. This suggests that crop yield variables contribute meaningful predictive information across all IPC classes.

The largest improvement was observed for IPC 3+ where the F1-score increased from 0.688 to 0.759. This statistically significant gain indicates that crop yield data substantially enhances the model’s ability to identify severe food insecurity events.

For IPC 2, the F1-score increased from 0.530 to 0.560 with a p -value of 0.0635, which was not statistically significant at the 5% level. While the result suggests a positive contribution of crop yield variables, the evidence is insufficient to conclude that the improvement is unlikely to have occurred by chance.

Overall accuracy increased only marginally from 0.934 to 0.937 and this change was not statistically significant with a p -value of 0.2020, which suggests that accuracy alone may underestimate the practical value of crop yield data.

When combined, these findings indicate that crop yield variables provide significant added predictive value, particularly for identifying severe food insecurity conditions. While their impact on overall accuracy is limited, they contribute meaningfully to improvements in class-balanced performance metrics and crisis detection, on which basis they are important to be included in the final forecasting model.

4.4.2 Baseline Comparison

Four models were compared on the same Uganda test set: Baseline naïve model (Persistence), Logistic Regression on lag features, Random Forest, and the Full Hybrid Model.

Table 4.7: Performance comparison amongst the four models.

Model	Macro F1	IPC2 F1	IPC3+ F1	Accuracy
Persistence	0.768	0.618	0.711	0.946
Logistic Regression	0.698	0.508	0.622	0.925
Random Forest	0.735	0.542	0.700	0.927
Full Model	0.765	0.577	0.750	0.938

The Persistence model, which assumes that future IPC conditions remain unchanged from the most recent observation, achieved the highest overall accuracy of 0.946 and the highest IPC 2 F1-score of 0.618.

The Logistic Regression model produced the weakest performance across all metrics, achieving a Macro F1-score of 0.698, an IPC 2 F1-score of 0.508 and an IPC 3+ F1-score of 0.622.

The Random Forest model performed slightly better than the Logistic Regression, achieving a Macro F1-score of 0.735 and an IPC 3+ F1-score of 0.700.

The Full Hybrid Model achieved the strongest performance for crisis detection, recording the highest IPC 3+ F1-score of 0.750 while maintaining a high Macro F1-score of 0.765 and overall accuracy of 0.938.

The Persistence model, which predicts the district’s current IPC phase `fews_ipc` as the forecast three months ahead, achieved the highest overall accuracy of 0.946 and macro F1-score of 0.768 among the four models evaluated, marginally outperforming the proposed hybrid model on these aggregate metrics.

However, these results should not be taken as evidence that persistence is the superior forecasting approach. IPC classifications exhibit strong temporal persistence at the district level with food security conditions often remaining unchanged over a three-month period. Consequently, a persistence model correctly predicts all instances in which no transition occurs without modelling the underlying drivers of food insecurity. Its strong aggregate performance including an IPC 3+ recall of 0.73, therefore reflects the temporal stability of the target variable rather than genuine predictive capability.

The principal limitation of the persistence model and the primary justification for the operational value of the proposed hybrid framework is its inability to anticipate changes in food security status. By definition, persistence assumes that future conditions will be identical to

those currently observed and therefore cannot predict transitions from lower IPC phases into crisis. In contrast, the objective of an early warning system is not merely to reproduce existing conditions but to identify emerging deterioration before it occurs and enable timely intervention. This distinction is reflected in the class-specific performance metrics.

Although the persistence model achieved strong overall accuracy, the proposed hybrid model attained a higher IPC 3+ F1-score of 0.750 as well as higher precision for the crisis class which indicates that a greater proportion of predicted crisis events represented genuine future deteriorations rather than the continuation of existing crises. These improvements suggest that the hybrid model captures predictive signals associated with impending food insecurity that are unavailable to a persistence-based approach. Accordingly, aggregate performance measures such as overall accuracy and macro F1-score should be interpreted alongside the forecasting objective.

For food security early warning, the ability to forecast transitions into deteriorating conditions is considerably more valuable than accurately predicting districts whose food security status remains unchanged. From this perspective, the hybrid LSTM–XGBoost model provides greater practical utility despite the persistence baseline achieving marginally higher aggregate scores.

4.4.3 Temporal Cross-validation

Three-fold temporal cross-validation was conducted on the data at district-month level.

Table 4.8: Three-fold temporal cross-validation results on Uganda data.

Fold	Training Months	Test Months	Training End	Macro F1	IPC2 F1	IPC3+ F1	Accuracy
1	38	38	01/04/2013	0.622	0.511	0.375	0.956
2	76	38	01/06/2016	0.662	0.878	0.111	0.987
3	114	38	01/08/2019	0.649	0.512	0.473	0.912
Mean				0.644	0.633	0.320	0.952

Across the three folds, the model achieved a mean Macro F1-score of 0.644 indicating relatively stable performance across time. Overall accuracy remained high, averaging 0.952 suggesting that the model consistently classified the dominant IPC categories correctly.

Performance for IPC 2 varied substantially across folds, with F1-scores ranging from 0.511 to 0.878. This variability may reflect differences in the prevalence and characteristics of stressed conditions across time periods. Similarly, IPC 3+ exhibited the greatest instability, with F1-scores ranging from 0.111 to 0.473. The lower and more variable performance for crisis events is expected given their rarity and the evolving nature of severe food insecurity drivers over time.

Fold 2 achieved the highest overall accuracy of 0.987 and IPC 2 F1-score of 0.878 but recorded the lowest IPC 3+ F1-score of 0.111 indicating that excellent overall classification performance does not necessarily translate into effective crisis detection. In contrast, Fold 3 achieved the strongest IPC 3+ performance at 0.473 despite having the lowest overall accuracy of 0.912 illustrating the trade-off between optimizing for majority-class accuracy and detecting rare but operationally important crisis events.

Overall, the temporal cross-validation results demonstrate that the model retains predictive capability when applied to future time periods, although performance for rare crisis events remains more variable than for common IPC categories.

4.4.4 Feature Importance by Group

SHapley Additive exPlanations (SHAP) was used to quantify the contribution of feature groups to model importance.

Table 4.9: SHAP Analysis results for Feature Importance by Group.

Feature Group	Sum of SHAP	Percentage of Total Importance
IPC History	1.7049	36.00%
Environmental	0.8926	18.80%
Temporal	0.7002	14.80%
Market/Price	0.5804	12.20%
Crop Yield	0.3786	8.00%
Geospatial	0.2181	4.60%
Demographic	0.1685	3.60%
Conflict	0.0969	2.00%

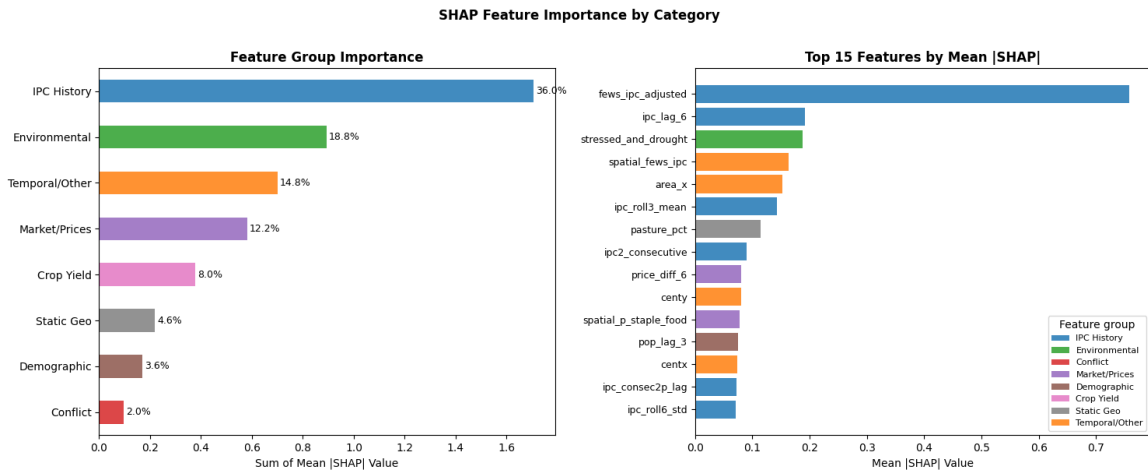


Figure 4.4: SHAP Feature Importance by Category.

IPC History was the most influential feature group, accounting for 36.0% of total model importance which is consistent with earlier correlation and mutual information analyses which demonstrated that past food security conditions are the strongest predictors of future IPC outcomes.

Environmental variables constituted the second most important group, contributing 18.8% of total importance, which highlights the critical role of vegetation, rainfall patterns and other environmental factors in shaping food security outcomes.

Temporal indicators contributed 14.8%, while market and price variables accounted for 12.2% of total importance. The importance of price-related indicators suggests that food affordability and market dynamics are significant drivers of food insecurity.

Notably, Crop Yield variables contributed 8.0% of total model importance, exceeding the contributions of Geospatial at 4.6%, Demographic at 3.6% and Conflict-related variables at 2.0%. This further provides evidence that crop yield data adds meaningful predictive value to the forecasting framework.

4.4.5 Model Comparison: Uganda-specific vs consolidated regional model

A comparison was carried out between two models evaluated on the same test data split for Uganda. Model A was trained purely on data for Uganda while Model B was trained on data for Uganda, Kenya, South Sudan and Ethiopia.

Table 4.10: Model Comparison results. Δ Mean = Model B – Model A. Significant = statistically significant at $\alpha = 0.05$.

Metric	Model A (Uganda)	Model B (Regional)	Δ Mean	p -value	Significant
macro_f1	0.747	0.725	-0.022	0.0361	Yes
ipc2_f1	0.603	0.523	-0.080	0.0014	Yes
ipc3_f1	0.662	0.680	0.018	0.3037	No
accuracy	0.950	0.938	-0.012	0.0023	Yes

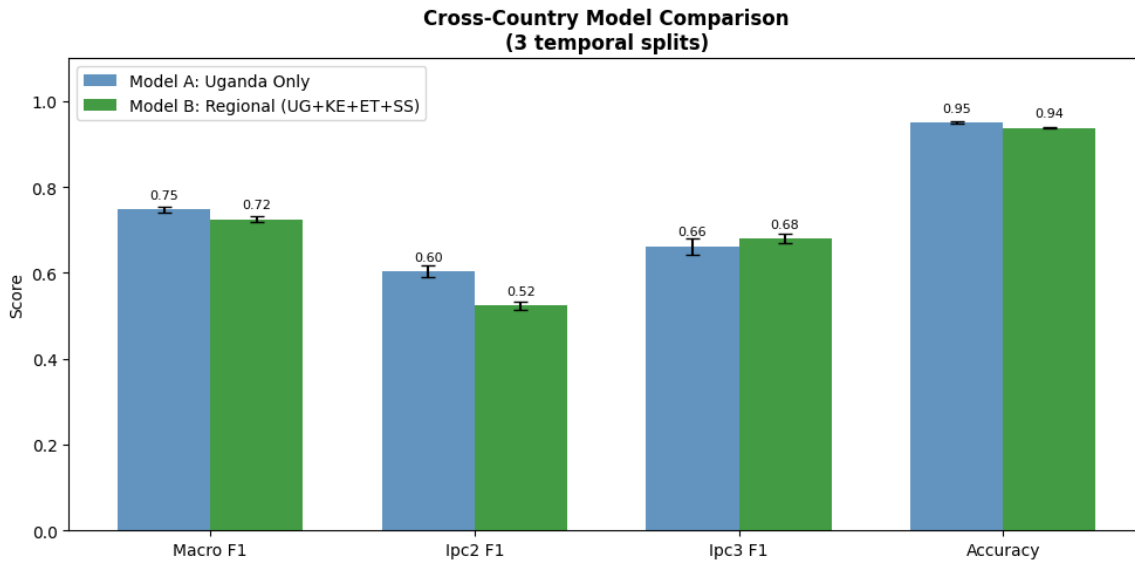


Figure 4.5: Model Comparison: Uganda-specific vs consolidated regional model.

The Macro F1-score decreased from 0.747 in Model A to 0.725 in Model B with a p -value of 0.0361 which indicates a statistically significant reduction in overall balanced classification performance when regional data were incorporated. Similarly, the IPC 2 F1-score declined substantially from 0.603 to 0.523 with a p -value of 0.0014 which suggests that the inclusion of neighbouring countries introduced additional heterogeneity that reduced the model's ability to identify stressed food security conditions in Uganda.

Overall accuracy also decreased significantly from 0.950 to 0.938 with a p -value of 0.0023. Although the magnitude of the reduction was relatively small, the statistical significance indicates that the Uganda-only model consistently achieved better classification performance across repeated evaluations.

In contrast, the IPC 3+ F1-score increased slightly from 0.662 to 0.680. However, this improvement was not statistically significant with a p -value of 0.3037 which suggested that the observed difference could have arisen from sampling variability.

Overall, the findings suggest that while regional training data may offer some benefits for crisis detection, the Uganda-only model provides superior performance for general food insecurity classification. Further, this implies that country-specific food security dynamics remain sufficiently distinct that locally trained models may be more effective than broader regional models for food security forecasting within Uganda.

Chapter 5

Discussion

5.1 Relationship between Geospatial variables, Crop yield and Food Security

The relationship analyses demonstrate that geospatial variables, crop yield and food security indicators are significantly interconnected in Uganda, consistent with established food systems research [56] [57]. However, the relative importance of these predictors differs from much of the remote-sensing literature, which often treats vegetation and rainfall indicators as the primary drivers of food security outcomes.

Among all variables examined, historical IPC indicators exhibited the strongest associations with future IPC. The adjusted current IPC phase with a correlation ρ of 0.888 and the three-month rolling IPC mean with a correlation ρ of 0.866 substantially outperformed environmental and agricultural variables. While the strong predictive power of IPC history reflects the temporal persistence of food insecurity, it also highlights an important methodological consideration that models that rely solely on environmental indicators while excluding historical IPC information are likely to omit a major predictive variable.

Spatial variables were also highly influential. The spatial lag of IPC with $\rho = 0.769$ emerged as the strongest geospatial predictor, indicating pronounced geographic clustering of food insecurity. Districts surrounded by food-insecure neighbours are themselves more likely to experience deteriorating food security conditions, reflecting shared exposure to climatic shocks, market disruptions, population movements and conflict spillovers. This justifies the inclusion of spatial predictors and emphasizes that food insecurity should be analysed as a spatially connected process rather than as independent observations across districts.

Environmental indicators also demonstrated meaningful predictive value. Vegetation health, represented by NDVI was negatively associated with future IPC, indicating that improved vegetation conditions are linked to reduced food insecurity risk.

Crop yield indicators showed significant associations with future IPC even after controlling for rainfall, NDVI, conflict intensity and seasonality. In particular, maize yield with partial $r = -0.269$ and banana yield with partial $r = -0.198$ remained significant predictors, which shows that crop production contributes information beyond that captured by environmental monitoring systems. This finding provides empirical support for integrating agricultural production statistics into food security forecasting frameworks. Furthermore, mutual information analysis revealed that yield-shock variables, particularly `beans_shock` with $MI = 0.334$ and `maize_shock` with $MI = 0.317$, contain substantial non-linear information about future food

insecurity, indicating that deviations from expected production levels may be more informative than absolute yield values.

5.2 Model Performance

The proposed hybrid LSTM–XGBoost framework demonstrated strong predictive performance, particularly for crisis-level food insecurity. On the held-out Uganda test set, the model achieved an IPC 3+ recall of 0.82, an IPC 3+ F1-score of 0.75, a ROC-AUC of 0.998 and a PR-AUC of 0.852. These results exceeded those of both the Logistic Regression baseline (IPC 3+ F1-score = 0.622) and the Random Forest baseline (IPC 3+ F1-score = 0.700), demonstrating the benefit of combining deep temporal representation learning with gradient-boosted decision trees for forecasting severe food insecurity.

Comparison with the persistence baseline requires careful interpretation. Although persistence achieved marginally higher aggregate metrics, including a macro F1-score of 0.768 compared with 0.765 for the proposed model and a higher IPC 2 F1-score (0.618 versus 0.577), these results should not be interpreted as evidence of superior forecasting capability. Persistence simply assumes that current IPC conditions remain unchanged and therefore performs well when food security conditions are temporally stable. It cannot, however, anticipate transitions into deteriorating conditions. In contrast, the proposed framework generates genuine three-month-ahead forecasts and successfully identified 82% of future crisis events, providing the predictive capability required for operational early warning.

While the hybrid framework consistently outperformed the simpler machine learning baselines, the magnitude of these improvements should be considered alongside the additional modelling complexity. Relative to the Random Forest baseline, the hybrid model improved the macro F1-score from 0.735 to 0.765 and the IPC 3+ F1-score from 0.700 to 0.750. Although these gains are operationally meaningful, particularly for crisis detection, they are achieved through a substantially more complex pipeline involving LSTM pretraining, embedding extraction, ordinal-aware XGBoost optimisation and isotonic probability calibration. By comparison, the Random Forest baseline comprises a single, comparatively straightforward model that is easier to train, deploy and maintain. Consequently, the hybrid architecture is most strongly justified in applications where improvements in minority-class detection outweigh the additional computational and engineering costs, whereas the Random Forest may remain an attractive alternative for resource-constrained operational environments.

The SHAP analysis provides further insight into the contribution of the temporal encoder. Temporal and sequence-based representations accounted for 14.8% of the overall model importance, compared with 36.0% attributed to IPC history features. This indicates that, although the LSTM-derived embedding contributes additional predictive information beyond the manually engineered lag variables, much of the temporal signal is still captured by explicit historical IPC features.

The hybrid model nevertheless consistently outperformed the Random Forest baseline, which was trained using the same engineered predictors but without the learned embedding, demonstrating that the temporal encoder contributes complementary predictive information. However, the relatively modest improvement suggests that the incremental value of the learned temporal representation may be smaller than the complexity of the encoder architecture alone might imply. Future work should therefore quantify this contribution more directly through an ablation study comparing XGBoost models trained with and without the 24 `lstm_feat_*` features while keeping all remaining predictors unchanged.

The comparatively weak and unstable performance observed for IPC 2 (Stressed) across the classification report, temporal cross-validation, and calibration analysis also warrants further consideration. As the intermediate category between minimal and crisis food insecurity, IPC 2 inherently shares characteristics with both neighbouring classes, making it the most ambiguous classification target. This ambiguity is reflected in the model’s precision of 0.41 and recall of 0.76 despite the application of additional class weighting during training. These findings suggest that further improvements may not be achieved through feature engineering alone. Instead, future research should investigate modelling approaches that better exploit the ordered nature of the IPC scale, such as fully ordinal classification methods or models designed to learn transition boundaries between adjacent IPC phases.

Probability calibration further demonstrated that the model’s predicted probabilities generally corresponded well with observed event frequencies, particularly for IPC 1. Although calibration was less stable for IPC 2 and IPC 3+, this behaviour is likely attributable to the limited number of minority-class observations available for fitting the isotonic calibration model rather than deficiencies in the calibration procedure itself. Consequently, the model is best viewed as a tool for ranking and prioritising districts according to relative food security risk rather than producing highly precise probability estimates for minority classes. The use of class-specific decision thresholds further enhanced the framework’s operational relevance by improving crisis detection while limiting unnecessary false alarms, thereby reflecting the practical requirements of humanitarian early warning systems.

Overall, the findings demonstrate that the proposed hybrid LSTM–XGBoost framework successfully integrates temporal sequence learning with powerful tabular classification to produce accurate and operationally relevant food security forecasts. Although the performance gains over simpler machine learning models are incremental rather than transformative, they are concentrated in the detection of minority crisis events, where forecasting accuracy is of greatest practical importance. These results therefore support the use of hybrid temporal–tabular modelling for food security early warning while also highlighting opportunities to simplify the architecture, strengthen ordinal modelling, and further evaluate the unique contribution of learned temporal representations in future research.

5.3 The Impact of Crop Yield data

The ablation study provides strong empirical evidence of the value of crop yield information for food security forecasting. Models that integrated crop yield variables achieved significantly better predictive performance than models that excluded them. Specifically, the macro F1-score increased from 0.729 to 0.763, which showed an overall improvement in classification performance across all IPC categories. More importantly, the model’s ability to identify crisis-level food insecurity IPC-3+ improved substantially with the IPC-3+ F1-score rising from 0.688 to 0.759. This represents a 10.3% improvement in crisis detection performance which highlights the practical importance of crop yield information for identifying populations at greatest risk of acute food insecurity.

Enhanced IPC-3+ detection capability can facilitate earlier identification of emerging food insecurity hotspots which enables timely preparation of humanitarian assistance and resource allocation. Furthermore, the observed improvements were consistent across multiple temporal train–test splits which suggests that the contribution of crop yield information is robust.

While crop yield information significantly improved overall model performance and crisis detection, its effect on the classification of stressed food security conditions IPC-2 was less pronounced. Although the IPC-2 F1-score increased, the improvement did not reach statistical sig-

nificance with a p -value of 0.064. This finding suggests that crop yield information contributes more strongly to the prediction of severe food insecurity outcomes than to the distinction between minimal and stressed conditions.

Similarly, the absence of a statistically significant improvement in overall accuracy was anticipated given the substantial class imbalance within the dataset where IPC-1 observations made up the majority of cases.

Conclusively, by demonstrating statistically significant improvements in both macro F1-score from 0.729 to 0.763 and crisis-level detection performance from 0.688 to 0.759, this study provides robust evidence that crop yield information enhances forecasting accuracy and should be integrated into operational food security early warning systems.

5.4 Limitations

Several limitations should be considered when interpreting the findings of this study.

Firstly, the number of IPC-3+ observations in the test set was relatively small with 22 district-month observations, resulting in wide bootstrap confidence intervals for crisis-class performance metrics [0.590, 0.866]. Although the lower bound remains substantially above chance level, estimates of crisis detection performance should be interpreted with caution and validated using larger hold-out samples or prospective operational data.

Secondly, the crop yield data obtained from FAOSTAT are annual national-level aggregates rather than district-level or seasonal observations. This can mean that important sub-national variations in agricultural production can be miscalculated such as poor harvests in one region may be offset by stronger performance elsewhere, reducing the sensitivity of yield variables to localised food security shocks. Access to district-level yield estimates would likely enhance predictive performance.

While the LSTM encoder demonstrated stable learning behaviour, the final training loss of 0.9606 suggests that temporal representations may not have been fully optimised within the 30-epoch training budget. Additional experimentation with longer training schedules, larger hidden dimensions or deeper architectures such as transformers may yield further performance improvements, although such modifications would increase the risk of over-fitting.

Finally, the study period ends in December 2019 and therefore does not capture the disruptions associated with the COVID-19 pandemic and subsequent economic shocks.

Chapter 6

Conclusion and Recommendation

6.1 Recommendations

Based on the findings of this study, several recommendations are proposed for researchers, practitioners, and policymakers involved in food security forecasting in Uganda.

To improve data quality and forecasting accuracy, investment in district-level crop yield monitoring is strongly recommended. The ablation analysis demonstrated that crop yield information significantly improves crisis detection performance, indicating that more granular agricultural production data could generate substantial forecasting benefits. Integration of seasonal climate forecasts from sources such as ICPAC would further enhance the model’s ability to anticipate emerging food security risks.

Future research should focus on several methodological improvements. These include investigating regional pre-training followed by Uganda-specific fine-tuning, extending LSTM training with validation-based early stopping, incorporating district-level agricultural statistics from UBOS and resolving the spatial weight matrix issue to enable formal spatial autocorrelation testing. Such enhancements could further strengthen predictive performance while improving model robustness.

From a policy perspective, model outputs should complement rather than replace existing expert-led assessment systems. Integration within the FEWS NET early warning framework would allow forecasts to support analyst decision-making while benefiting from contextual validation. Given the model’s strong crisis recall at 82% and crisis PR-AUC at 0.852, it is particularly well suited as a screening and prioritisation tool for identifying districts that warrant closer monitoring and early intervention.

With respect to model architecture, this study deliberately adopted a compact LSTM–XGBoost hybrid rather than a transformer-based encoder because of the available dataset size and the relatively short 18-month input window. While the LSTM provided an effective and computationally efficient means of capturing temporal dependencies under these conditions, transformer-based architectures are generally expected to realise their full potential when trained on substantially larger datasets with longer input sequences and higher temporal resolution.

As more extensive district-level historical records become available, particularly higher-frequency satellite-derived climatic, market and socio-economic data spanning longer time horizons and a broader range of countries, the suitability of transformer-based or hybrid attention–LSTM–transformer architectures should be re-evaluated. Future research should benchmark these architectures against the LSTM–XGBoost framework developed in this study using the same evaluation pro-

toocol in order to determine whether the additional modelling complexity yields statistically and operationally significant improvements in food security forecasting, particularly for the detection of minority crisis classes.

6.2 Conclusion

This study developed and evaluated a geospatial machine learning framework for forecasting food insecurity in Uganda by integrating environmental remote sensing indicators, crop yield information, conflict data, market prices, demographic characteristics and historical IPC observations within a hybrid LSTM–XGBoost architecture.

In addressing the first research objective, the study demonstrated that future food insecurity is strongly associated with historical IPC conditions, environmental factors, agricultural production and market dynamics. Historical IPC indicators emerged as the most influential predictors, reflecting the persistence of food insecurity over time. Nevertheless, crop yields retained significant predictive value even after controlling for environmental and conflict-related factors, which confirms that agricultural production contributes information beyond that captured by satellite-derived indicators alone.

For the second research objective, the proposed hybrid model achieved strong three-month-ahead forecasting performance, including an IPC-3+ recall of 0.82, a crisis-class ROC-AUC of 0.998, a crisis-class PR-AUC of 0.852 and a macro F1-score of 0.75. The integration of an attention-based bidirectional LSTM encoder enabled the model to capture complex temporal patterns that would not be fully represented by conventional lag features, which contributed meaningfully to overall predictive performance.

For the third research objective, the ablation study provided statistically robust evidence that crop yield information enhances food security forecasting. The inclusion of yield variables significantly improved both macro F1-score and crisis detection performance. These findings demonstrate the practical value of incorporating agricultural production information into operational food security monitoring systems.

An additional contribution of the study was the comparison between Uganda-only and regional training strategies where results showed that models trained exclusively on Ugandan observations significantly outperformed models trained on pooled regional datasets. This suggests that country-specific food security dynamics remain sufficiently distinct to outweigh the potential benefits of additional training data from neighbouring countries.

Overall, the study demonstrates that integrating crop yield information, environmental monitoring, conflict indicators, market conditions and historical IPC data within a hybrid deep learning framework can produce accurate and operationally relevant food security forecasts for Uganda. The framework provides interpretable predictions, supports district-level early warning and has been validated through temporal cross-validation, baseline comparison, feature attribution analysis and ablation testing. Future work should prioritise the incorporation of district-level agricultural data, seasonal climate forecasts and enhanced spatial modelling techniques to further improve forecasting performance and operational applicability.

Bibliography

- [1] F. R. Declaration, “World food summit plan of action.” FAO Rome Italy, 1996.
- [2] L. Scott, V. Diwakar, and M. Okech, “Ensuring escapes from poverty are sustained in uganda,” *CPAN ODI Prod. Leo Rep.*, vol. 25, 2016.
- [3] UNICEF, “The state of food security and nutrition in the world 2022,” 2022.
- [4] R. Atuhaire, G. Ssemaganda, E. Kaboggoza, G. Marvin, and R. Nakibuule, “Transparent multi-modelling for climatic geo-intelligence in predicting food security,” in *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, pp. 271–281, 2024.
- [5] UBOS, “2022 statistical abstract,” 2022.
- [6] N. Oceanic and A. Administration, “Meet enso’s neighbor, the indian ocean dipole — noaa climate.gov.” Accessed: Dec. 19, 2025.
- [7] Encord, “What is a confusion matrix? — machine learning glossary — encord — encord.” Accessed: May 01, 2025.
- [8] Fao, “Food and agriculture organization of the united nations,” *Rome URL Httpfaostat Fao Org*, vol. 403, 2018.
- [9] T. M. Mitchell and T. M. Mitchell, *Machine learning*, vol. 1. McGraw-hill New York, 1997.
- [10] D. P. Dee, “Bias and data assimilation,” *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 613, pp. 3323–3343, 2005.
- [11] U. Nations, “Goal 2 — department of economic and social affairs.” Accessed: Apr. 29, 2025.
- [12] Private Sector Development Unit, “Fourth national development plan (ndp iv),” 2025. Accessed: 2026-06-08.
- [13] B. Egamberdiev, I. Bobojonov, L. Kuhn, and T. Glauben, “Household resilience capacity and food security: evidence from kyrgyzstan,” *Food Secur.*, vol. 15, pp. 967–988, Aug. 2023.
- [14] B. Sultan, D. Defrance, and T. Iizumi, “Evidence of crop production losses in west africa due to historical global warming in two crop models,” *Sci. Rep.*, vol. 9, no. 1, p. 12834, 2019.
- [15] D. Maxwell and M. Fitzpatrick, “The 2011 somalia famine: Context, causes, and complications,” *Glob. Food Secur.*, vol. 1, no. 1, pp. 5–12, 2012.

- [16] T. I. F. S. P. C. (IPC), “Uganda: Acute food insecurity situation for november 2017 - february 2018 — ipc - integrated food security phase classification.” Accessed: Sep. 30, 2025.
- [17] M. E. Brown and C. C. Funk, “Food security under climate change,” *Science*, vol. 319, no. 5863, pp. 580–581, 2008.
- [18] C. Funk *et al.*, “Recognizing the famine early warning systems network: over 30 years of drought early warning science advances and partnerships promoting global food security,” *Bull. Am. Meteorol. Soc.*, vol. 100, no. 6, pp. 1011–1027, 2019.
- [19] D. B. Lobell and S. M. Gourdji, “The influence of climate change on global crop productivity,” *Plant Physiol.*, vol. 160, no. 4, pp. 1686–1697, 2012.
- [20] S. E. Nicholson, “Climate and climatic variability of rainfall over eastern africa,” *Rev. Geophys.*, vol. 55, no. 3, pp. 590–635, 2017.
- [21] T. E. Epule, J. D. Ford, S. Lwasa, B. Nabaasa, and A. Buyinza, “The determinants of crop yields in uganda: what is the role of climatic and non-climatic factors?,” *Agric. Food Secur.*, vol. 7, no. 1, p. 10, 2018.
- [22] W. Gumindoga, C. Liwonde, D. T. Rwasoka, P. Kowe, A. Maviza, J. Magidi, L. Chikwiramakomo, M. d. J. P. Mavaringana, and E. Tshitende, “Urban flash floods modeling in mzuzu city, malawi based on sentinel and modis data,” *Frontiers in Climate*, vol. 6, p. 1284437, 2024.
- [23] D. Mutembesa, E. Mwebaze, S. Nsumba, C. Omongo, and H. Mutaasa, “Mobile community sensing with smallholder farmers in a developing nation; a scaled pilot for crop health monitoring,” *ArXiv Prepr. ArXiv190807047*, 2019.
- [24] P. H. Verburg, W. Soepboer, A. Veldkamp, R. Limpiada, V. Espaldon, and S. S. Mastura, “Modeling the spatial dynamics of regional land use: the clue-s model,” *Environ. Manage.*, vol. 30, pp. 391–405, 2002.
- [25] M. Jerven, *Poor numbers: how we are misled by African development statistics and what to do about it*. Cornell University Press, 2013.
- [26] N. Torbick, W. Salas, D. Chowdhury, P. Ingraham, and M. Trinh, “Mapping rice greenhouse gas emissions in the red river delta, vietnam,” *Carbon Manag.*, vol. 8, no. 1, pp. 99–108, 2017.
- [27] A. McNally *et al.*, “Hydrologic and agricultural earth observations and modeling for the water-food nexus,” *Front. Environ. Sci.*, vol. 7, p. 23, 2019.
- [28] S. Khaki, L. Wang, and S. V. Archontoulis, “A cnn-rnn framework for crop yield prediction,” *Front. Plant Sci.*, vol. 10, p. 1750, 2020.
- [29] M. K. V. Ittersum *et al.*, “Can sub-saharan africa feed itself?,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 52, pp. 14964–14969, 2016.
- [30] J. H. Jeong *et al.*, “Random forests for global and regional crop yield predictions,” *PloS One*, vol. 11, no. 6, p. e0156571, 2016.
- [31] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

- [33] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018.
- [34] J. B. Cousins and L. M. Earl, “The case for participatory evaluation,” *Educ. Eval. Policy Anal.*, vol. 14, no. 4, pp. 397–418, 1992.
- [35] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 11121–11128, 2023.
- [36] Y. Zhu, S. Luo, D. Huang, W. Zheng, F. Su, and B. Hou, “Drcnn: Decomposing residual convolutional neural networks for time series forecasting,” *Scientific Reports*, vol. 13, p. 15901, 2023.
- [37] L. Grinsztajn, E. Oyallon, and G. Varoquaux, “Why do tree-based models still outperform deep learning on typical tabular data?,” in *Advances in Neural Information Processing Systems*, vol. 35, pp. 507–520, 2022.
- [38] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
- [39] N. Kordzadeh and M. Ghasemaghaei, “Algorithmic bias: review, synthesis, and future research directions,” *Eur. J. Inf. Syst.*, vol. 31, no. 3, pp. 388–409, 2022.
- [40] A. B. Bondi, “Characteristics of scalability and their impact on performance,” in *Proceedings of the 2nd international workshop on Software and performance*, pp. 195–203, 2000.
- [41] D. N. Mubiru *et al.*, “Climate trends, risks and coping strategies in smallholder farming systems in uganda,” *Clim. Risk Manag.*, vol. 22, pp. 4–21, 2018.
- [42] FAO, “Faostat.” Accessed: Apr. 30, 2025.
- [43] W. Bank, “Open knowledge repository.” Accessed: Apr. 30, 2025.
- [44] D. R. Roberts *et al.*, “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure,” *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.
- [45] B. P. J. Andree, A. Chamorro, A. Kraay, P. Spencer, and D. Wang, “Predicting food crises 2020: Dataset for reproducing working paper results,” 2020.
- [46] T. Hengl, M. Nussbaum, M. N. Wright, G. B. Heuvelink, and B. Gräler, “Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables,” *PeerJ*, vol. 6, p. e5518, 2018.
- [47] R. D. Riley *et al.*, “Minimum sample size for developing a multivariable prediction model: Part ii-binary and time-to-event outcomes,” *Stat. Med.*, vol. 38, no. 7, pp. 1276–1296, 2019.
- [48] T. Hastie, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
- [49] O. Troyanskaya *et al.*, “Missing value estimation methods for dna microarrays,” *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
- [50] A. E. Roth, *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.

- [51] B. Lim and S. Zohren, “Time-series forecasting with deep learning: A survey,” *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200209, 2021.
- [52] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance,” *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.
- [53] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. Chapman and Hall/CRC, 1994.
- [54] B. J. Erickson and F. Kitamura, “Magician’s corner: 9. performance metrics for machine learning models,” *Radiol. Artif. Intell.*, vol. 3, no. 3, p. e200126, 2021.
- [55] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, “The ‘k’ in k-fold cross validation,” in *Proceedings of the ESANN*, pp. 441–446, 2012.
- [56] P. K. Krishnamurthy, K. Lewis, and R. Choularton, “A methodological framework for rapidly assessing the impacts of climate risk on national-level food security through a vulnerability index,” *Glob. Environ. Change*, vol. 25, pp. 121–132, 2014.
- [57] S. Shukla *et al.*, “Improving early warning of drought-driven food insecurity in southern africa using operational hydrological monitoring and forecasting products,” *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 4, pp. 1187–1201, 2020.
- [58] N. Burkart and M. F. Huber, “A survey on the explainability of supervised machine learning,” *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, 2021.

Appendix A

Declaration of AI use

Declaration Statement:

I, Philip Arinaitwe, declare that I used

- i. *Claude AI* to *integrate geospatial and choropleth map visualizations into the code.*
- ii. *Claude AI* to *build the LSTM encoder.*

Signature: _____  Date: 09/09/2026 _____

Appendix B

List of 10 Major Journal Publications reviewed

1. Atuhaire et al. (2024). *Transparent multi-modelling for climatic geo-intelligence in predicting food security*. IC3 (International Conference on Contemporary Computing). [4]
2. Chen and Guestrin (2016). *XGBoost: A Scalable Tree Boosting System*. ACM SIGKDD Conference on Knowledge Discovery and Data Mining. [38]
3. Hochreiter and Schmidhuber (1997). *Long Short-Term Memory*. *Neural Computation*. [31]
4. Khaki et al. (2020). *A CNN-RNN Framework for Crop Yield Prediction*. *Frontiers in Plant Science*. [28]
5. McNally et al. (2019). *Hydrologic and Agricultural Earth Observations for the Water-Food Nexus*. *Frontiers in Environmental Science*. [27]
6. Shukla et al. (2020). *Improving Early Warning of Drought-Driven Food Insecurity in Southern Africa*. *Natural Hazards and Earth System Sciences*. [57]
7. Epule et al. (2018). *Determinants of Crop Yields in Uganda: Climatic and Non-Climatic Factors*. *Agriculture & Food Security*. [21]
8. Jeong et al. (2016). *Random Forests for Global and Regional Crop Yield Predictions*. *PLOS ONE*. [30]
9. Kamilaris and Prenafeta-Boldú (2018). *Deep Learning in Agriculture: A Survey*. *Computers & Electronics in Agriculture*. [33]
10. Burkart and Huber (2021). *A Survey on the Explainability of Supervised Machine Learning*. *Journal of Artificial Intelligence Research*. [58]

Appendix C

Research alignment with specialization, SDG, NDPIV & Vision 2040

C.1 Alignment with programme specialization

This study directly aligns with the Artificial Intelligence and Machine Learning specialisation. It develops a hybrid deep learning and gradient-boosting architecture that combines an LSTM encoder with an XGBoost classifier. The pipeline encompasses the full data science workflow starting with multi-source data integration, feature engineering across 139 variables, class imbalance handling via SMOTETomek, temporal cross-validation, SHAP-based model interpretability and bootstrap uncertainty quantification.

By applying these methods to food security forecasting, the project demonstrates how AI and ML can generate actionable and explainable predictions from complex geospatial and agricultural data.

C.2 Alignment with Sustainable Development Goals (SDG)

The study directly supports SDG 2 (Zero Hunger) by building a forecasting framework that enables earlier and more targeted responses to food insecurity. By predicting IPC food crisis phases three months in advance with a crisis recall of 82%, the model equips humanitarian agencies, policymakers and farmers with data-driven lead time to allocate resources and prepare interventions before conditions deteriorate. This directly advances SDG 2's targets on ending hunger, improving food access for vulnerable populations and building resilient agricultural systems in the face of climate variability.

C.3 Alignment with the National Development Plan IV (NDPIV)

NDP IV identifies a set of priority food security commodities — including maize and bananas, to be developed through area-based planning across Uganda's diverse agro-ecological zones. This study directly supports that agenda by using crop yield data for these staple crops to forecast potential shortfalls at the district level, providing planners with the data-driven insights needed to meet production and food security targets. The hybrid LSTM-XGBoost framework also aligns with NDP IV's emphasis on technology, innovation and value addition in agriculture,

demonstrating how cutting-edge AI and GIS analytics can modernise food security monitoring and strengthen Uganda’s agricultural planning infrastructure.

C.4 Alignment with the Uganda Vision 2040

Uganda Vision 2040 describes a transformation from a peasant to a modern, prosperous economy with agriculture serving as a key driver of growth through commercialisation, mechanisation and technology adoption. This study contributes to that vision by demonstrating how advanced data analytics can be embedded into agricultural and food security systems to drive evidence-based decision-making. The integration of machine learning with geospatial and crop yield data exemplifies the knowledge-driven, technology-enabled development that Vision 2040 envisions.

Appendix D

Research poster layout

