Predicting Child Malnutrition Risk and Humanitarian Challenges Resulting from Climate-Induced Crop Yield Declines in Sub-Saharan Africa

Image credit: https://www.google.com/url?sa=i&url=https%3A%2F%2Fmaritimefairtrade.org%2Fglobal-climate-change-trigger-decline-crop-yield%2F&psig=AOvVaw0TGWV0d6FF4gJNzZ8jIKML&ust=1760711700003000&source=images&cd=vfe&opi=89978449&ved=0CBUQjRxqFwoTCODDk674qJADFQAAAAAdAAAAABAE

By: Chenwei Pan, Aiza Khwaja

Project Timeline

September 2025

- **September 25-26:** Team Formation and Initial Planning
  - Formed a team for our science fair project. We decided to focus on using machine learning and data science to address global food security and malnutrition.
  - Our initial motivation came from wanting to work on a project that could have real humanitarian impact. We were particularly interested in how technology could help solve pressing global challenges related to hunger and nutrition.
- **September 27-30:** Research Topic Brainstorming
  - Spent time discussing different approaches to food security. We considered several angles:
    - Agricultural productivity and crop yields
    - Climate impacts on food systems
    - Malnutrition prediction and early warning
    - Integration of health and agricultural data
  - We decided to combine health data with agricultural data to create a more comprehensive project that could address multiple aspects of the food security challenge.

October 2025

- **October 1-5:** Background Research Begins
  - Started comprehensive literature review on food security and malnutrition. Key findings:
    1. 733 million people worldwide affected by malnutrition

    ii. Child malnutrition is particularly critical due to long-term developmental impacts
    iii. Sub-Saharan Africa has especially high rates (23.2% of children under 5 are stunted)
- Learned about different types of malnutrition:
    i. Wasting: Acute malnutrition, low weight-for-height, indicates recent weight loss
    ii. Stunting: Chronic malnutrition, low height-for-age, indicates long-term undernourishment
    iii. Underweight: Low weight-for-age

- **October 6-10:** Literature Review on Current Solutions
    - Researched existing early warning systems for food insecurity:
        i. FEWS NET (Famine Early Warning Systems Network): Used by humanitarian organizations, focuses on current crises
        ii. IPC (Integrated Food Security Phase Classification): Provides food security analysis and early warning
    - Gap identified: Most current systems are reactive rather than preventive. They respond to crises after malnutrition becomes widespread, rather than predicting and preventing it beforehand.
    - Reviewed academic papers on machine learning for malnutrition:
        i. Papua New Guinea study: 72.8% accuracy using gradient boosting
        ii. Bangladesh study: 59% accuracy with sensitivity of 61%
        iii. Malawi diagnostic model: 64% AUROC
    - Opportunity: Machine learning could potentially improve prediction accuracy and enable earlier intervention.
- **October 11-15:** Project Focus Refinement
    - Decided to focus specifically on child stunting in Sub-Saharan Africa.
    - Reasons for choosing this focus:
        i. Children are most vulnerable to long-term impacts

        ii. Stunting is largely irreversible after age 2, making early prediction crucial
        iii. Sub-Saharan Africa has highest malnutrition rates and is particularly vulnerable to climate change
        iv. Sufficient data available for this region
        v. Clear humanitarian impact - aligns with UN Sustainable Development Goal 2 (Zero Hunger)
    - Research question taking shape: Can we use machine learning to predict child stunting by integrating climate, agricultural, and socioeconomic data?
- **October 16-20:** Dataset Identification
    - Began identifying potential datasets for our project:
    - Climate Data:
        i. CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data): Rainfall estimates
        ii. World Bank Climate Portal: Temperature and precipitation data
        iii. NASA datasets: Climate projections and impacts
    - Agricultural Data:
        i. FAOSTAT (Food and Agriculture Organization): Crop production, yields, and food security indicators
        ii. Focused on key staple crops for Sub-Saharan Africa: maize, cassava, sorghum, wheat, yams
    - Health/Nutrition Data:
        i. UNICEF Joint Malnutrition Estimates (JME): Stunting and wasting prevalence by country
        ii. DHS (Demographic and Health Surveys): Household-level nutrition data
    - Socioeconomic Data:
        i. World Bank: GDP per capita, urbanization rates
        ii. FAOSTAT: Food consumption patterns, economic indicators
- **October 21-25:** Data Collection and Annotated Bibliography
    - Downloaded initial datasets from FAOSTAT:

        i. Selected Sub-Saharan Africa regions: Central, Eastern, Western, Southern Africa
        ii. Crops: Maize, Cassava, Sorghum, Wheat, Yams
        iii. Metrics: Production quantity, yield, and area harvested
        iv. Years: All available data (approximately 2000-2023)
    - Downloaded Food Security Indicators from FAOSTAT:
        i. GDP per capita (PPP, constant 2021 international dollars)
        ii. Percentage of children under 5 who are stunted
        iii. Average dietary energy supply
        iv. Access to basic drinking water services
        v. Access to basic sanitation services
        vi. Political stability index
    - Started annotated bibliography (can be seen at the end of the logbook) with 12+ sources including:
        i. Dataset sources (FAOSTAT, UNICEF, World Bank)
        ii. Research papers on ML for malnutrition prediction
        iii. Papers on climate impacts on agriculture
        iv. Early warning system documentation
    - This was created as part of background research to find potential datasets and gain base understanding of topic.

- **October 26-31:** Understanding Machine Learning Methods
    - Researched machine learning algorithms suitable for our project:
        i. Random Forest: Random Forest is a machine learning algorithm that makes predictions by combining many decision trees. Each tree analyzes the data independently, and the final prediction is made by averaging all the trees' answers (for numbers) or taking a majority vote (for categories). This approach reduces errors and improves accuracy. Random Forest works well with large datasets and helps avoid overfitting, which is when a model becomes too specific to training data and performs poorly on new

data. It's particularly good at handling many features and identifying which ones are most important.

ii. K-Nearest Neighbors (KNN): KNN is a simpler algorithm that makes predictions based on similarity. When predicting for a new data point, KNN finds the K most similar data points in the training data and uses their values to make a prediction. For example, if predicting stunting rates for Kenya in 2023, it would look at the K countries most similar to Kenya (based on GDP, rainfall, crop yields, etc.) and average their stunting rates. KNN works well when similar conditions lead to similar outcomes.

November 2025

- **November 1-7:** Initial Data Processing
  - Started coding in Google Colab to process and clean datasets.
    1. We created a shared notebook for collaboration. This allowed us to work on the code simultaneously and see each other's changes in real-time.
  - Data cleaning steps:
    1. Loaded CSV and Excel files using pandas
    2. Filtered data for Sub-Saharan Africa countries only
    3. Handled missing values using median imputation
    4. Standardized country names across datasets (some datasets used different naming conventions)
    5. Merged datasets based on country and year
  - Challenge: Different datasets had different time periods and not all countries had data for all years. Solution: Created a master dataset with only countries and years that had complete data across all sources.
- **November 8-14:** Feature Engineering Strategy
  - Developed our approach to creating features (input variables) for the machine learning model.
  - We decided to create three types of features:

    i. Current features: Present-year values (GDP, rainfall, crop production for the current year)
    ii. Lagged features: Previous years' values (lag-1 = 1 year ago, lag-2 = 2 years ago)
    iii. Change features: Year-over-year changes (how much did GDP change from last year?)
  - Rationale for lagged features: Stunting develops over time and is largely irreversible after age 2. This means conditions from 1-2 years ago can have more impact than current conditions. For example:
    i. 2021: Family experiences food insecurity due to drought
    ii. 2022: Child becomes malnourished during critical growth period (ages 0-2)
    iii. 2023: Child is stunted, even if food security improved in 2023
    iv. The 2021 drought (lag-2) is more predictive of 2023 stunting than 2023 conditions
- **November 15-21:** Downloading Additional Datasets
  - Downloaded climate data from World Bank Climate Portal:
    i. Format: JSON (required conversion to CSV for easier processing)
    ii. Variables: Temperature (degrees Celsius), Precipitation (mm)
    iii. Coverage: Sub-Saharan Africa, annual averages
    iv. Time period: 2000-2023
  - Downloaded UNICEF malnutrition datasets:
    i. Joint Malnutrition Estimates (JME) for stunting
    ii. Included point estimates and confidence intervals
    iii. Country-level data with some subnational breakdowns
  - Downloaded Children's Climate and Environment Risk Index (CCRI):
    i. Risk score based on children's exposure and vulnerability to climate change
    ii. Used as a feature in our model
  - Created a shared Google Drive folder to organize all datasets:

i. Raw data folder: Original downloaded files
ii. Processed data folder: Cleaned and merged datasets
iii. Code folder: Colab notebooks
iv. Documentation folder: Research notes and bibliography

- Decided on a prototype for our model that would function in the real world, outlining the key inputs, outputs, and intended use case for humanitarian organizations.

- **November 22-30:** Initial Model Building
  - Began data processing in Google Colab using pandas library.
  - Initial dataset contained 972 observations across 41 Sub-Saharan African countries. We needed to clean this data carefully:
  - Data cleaning approach:
    i. Standardized all datasets to country-year format
    ii. Combined into single unified dataset
    iii. Excluded countries with >30% missing data (following Alam et al. methodology)
    iv. Result: 852 observations across 36 countries
    v. Applied median imputation for remaining missing values
    vi. Final clean dataset: 780 complete observations
  - Why median imputation? Median is more robust to outliers than mean. For example, if one country has extremely high GDP but most have moderate GDP, median better represents the typical value.
  - Created master feature list documenting all 91 variables we would use for machine learning.

December 2025

- **December 1-7:** Feature Engineering
  - Worked on coding and debugging. Had to edit datasets that were being used to ensure consistency across sources.

    - Completed comprehensive feature engineering to create all input variables for the machine learning models.
    - Feature categories created:
        - Socioeconomic indicators (12 features): GDP per capita, water access, sanitation access, political stability
        - Climate variables (6 features): Temperature, precipitation
        - Crop production metrics (72 features): For 6 staple crops (cassava, maize, rice, sorghum, wheat, yams)
            - For each crop: production volume, area harvested, yield, volatility
            - Volatility = 3-year rolling standard deviation (measures harvest instability)
        - Environmental vulnerability (1 feature): CCRI score
    - For each relevant feature, we created three versions:
        - Current year value
        - Lag-1 (value from 1 year ago)
        - Lag-2 (value from 2 years ago)
    - Total: 91 features created
    - Why crop volatility matters: Unstable harvests (high volatility) make it harder for families to plan and can signal climate or economic stress. We calculated this as the rolling 3-year standard deviation of production.
- **December 8-14:** Understanding Nowcasting vs Forecasting
    - Realized we needed to build TWO different types of models to serve different purposes:
    - Nowcasting Model:
        - Purpose: Estimate CURRENT stunting rates when survey data unavailable
        - Uses: Current year features + lagged features (lag-1, lag-2)
        - Why useful: DHS surveys only happen every 4-5 years; this fills gaps between surveys
        - Example: Predict 2023 stunting using 2023, 2022, and 2021 data

    - Forecasting Model:
        - Purpose: Predict FUTURE stunting rates for early warning
        - Uses: ONLY lagged features (lag-1, lag-2), NO current year data
        - Why useful: True predictive capability - helps humanitarian organizations prepare in advance
        - Example: Predict 2024 stunting using only 2023 and 2022 data
    - This distinction is crucial - nowcasting helps when recent data is missing, forecasting enables proactive intervention.
- **December 15-21:** Algorithm Comparison and Hyperparameter Selection
    - Compared Random Forest vs K-Nearest Neighbors (KNN) for both nowcasting and forecasting:
    - Hyperparameter selection through sensitivity analysis:
        - Random Forest: 100 decision trees, maximum depth of 15
        - KNN: k=5 neighbors, distance-weighted
    - Used Scikit-Learn library to implement all models. Because KNN is sensitive to feature scale, standardized all input variables to zero mean and unit variance before training.
    - Realized that the model was nowcasting rather than truly forecasting, since current-year features were being used. Determined that nowcasting is still valuable for supplementing infrequent DHS surveys.
    - Created four different models by combining algorithms with temporal approaches:
        - RF Nowcasting
        - RF Forecasting
        - KNN Nowcasting
        - KNN Forecasting
    - Nowcasting results (using both current and lagged features):
        - Random Forest: $R^2 = 0.957$, MAE = 1.48%
        - KNN: $R^2 = 0.957$, MAE = 1.45%
        - Very similar performance, KNN slightly better

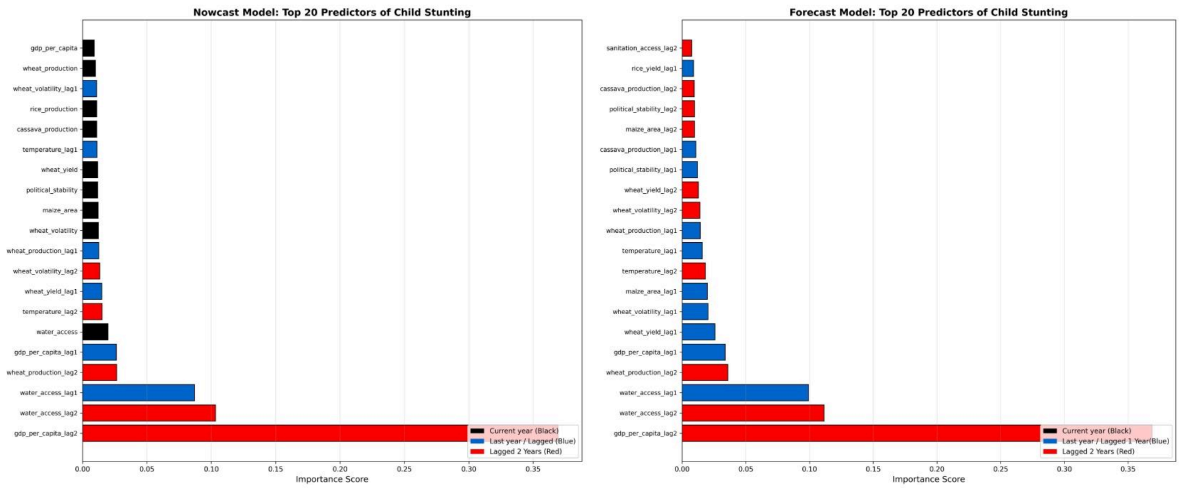
- Forecasting results (using only lagged features):
    - Random Forest: $R^2 = 0.957$, MAE = 1.52%
    - KNN: $R^2 = 0.950$, MAE = 1.58%
    - Random Forest performed slightly better
- Notably, KNN outperformed Random Forest for nowcasting, explained by KNN's distance-based calculations being more sensitive to correlated features, while Random Forest's approach naturally deprioritizes redundant features without hurting prediction accuracy.
- Key finding: Lag-1 features showed high importance in both models, consistent with the irreversible nature of stunting after age 2.

- **December 22-31:** Lag-2 Features
  - Tested whether using data from 2 years ago (lag-2) would improve predictions.
  - Hypothesis: Since stunting develops during critical window (ages 0-2), conditions from 2 years ago might be highly predictive.
  - Results: Lag-2 features showed HIGHER importance than lag-1 features!
  - Example: For predicting 2023 stunting rates:
    - 2021 GDP (lag-2) more important than 2022 GDP (lag-1)
    - 2021 water access (lag-2) more important than 2022 water access (lag-1)
  - Possible Explanation:
    - Critical development window: First 2 years of life most important for growth
    - Delayed effects: Malnutrition from 2021 manifests as stunting measured in 2023
    - Irreversibility: Stunting after age 2 cannot be reversed, so old conditions matter more
  - This discovery validated our approach and improved model performance!

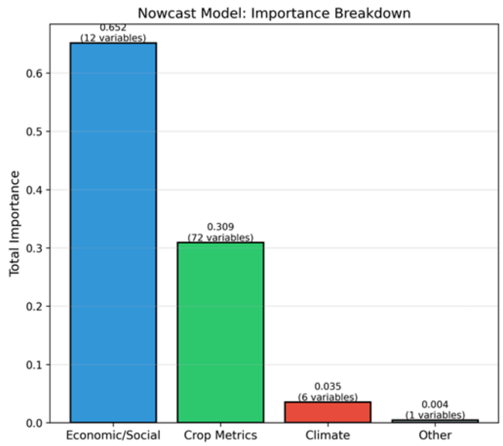
January 2026

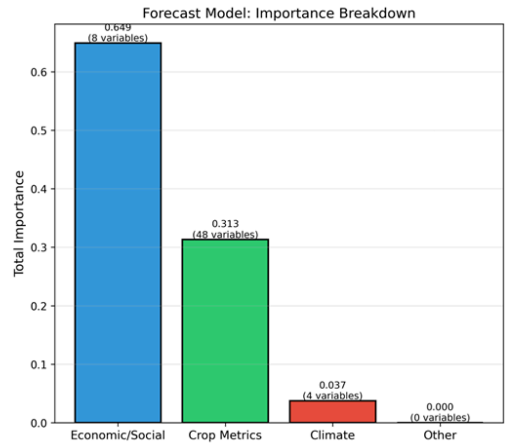
- **January 1-5:** Feature Importance Analysis

- Analyzed which features most strongly predict stunting rates.



-
- Key discovery: Socioeconomic factors (especially GDP and water access) are MORE important than climate or agriculture alone! This was surprising - we expected climate to dominate, but actually GDP and water infrastructure are the strongest predictors.
- Category-wise importance breakdown:

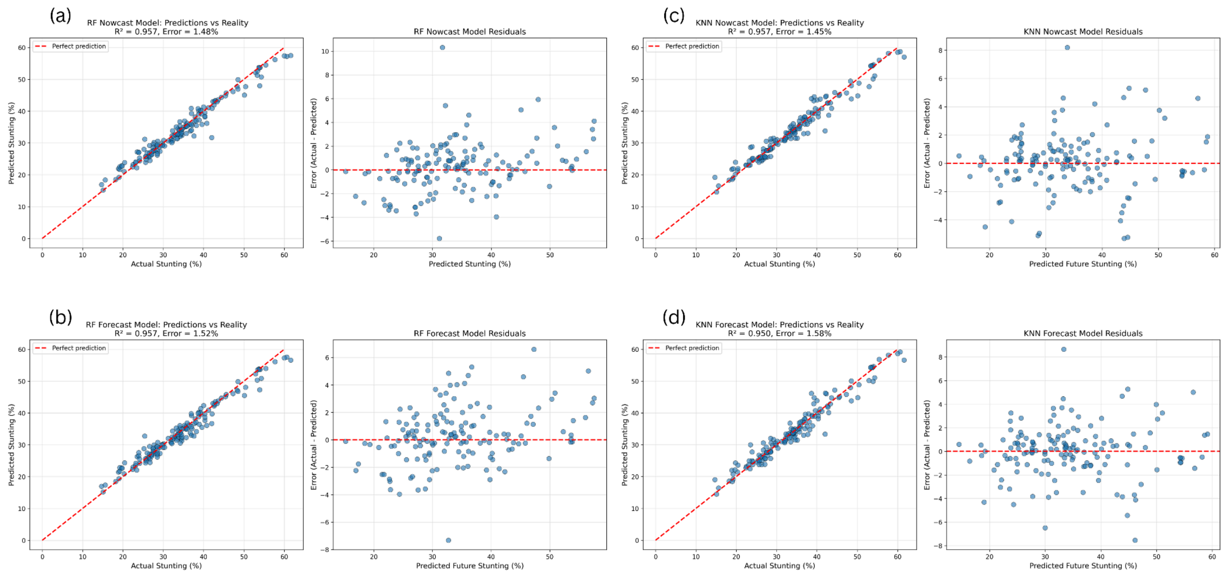




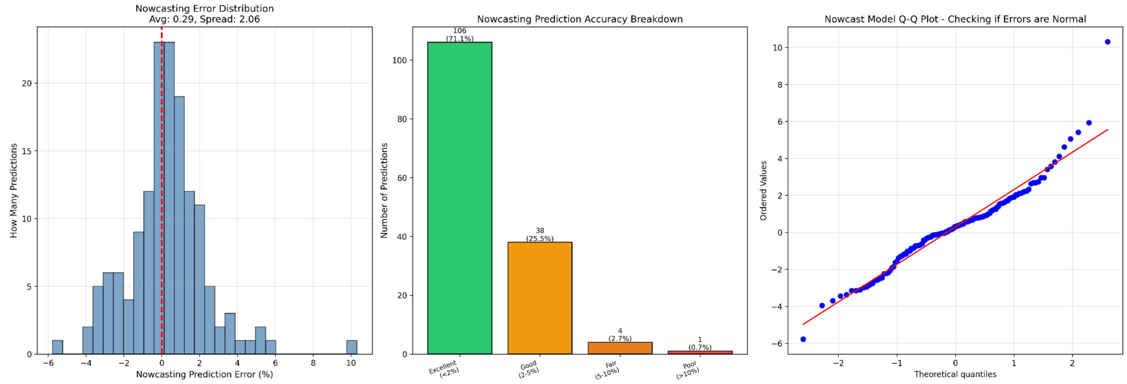
- This completely challenges our initial hypothesis! We thought climate and agriculture would dominate, but actually socioeconomic factors (especially GDP and water access) are FAR more important than climate alone.

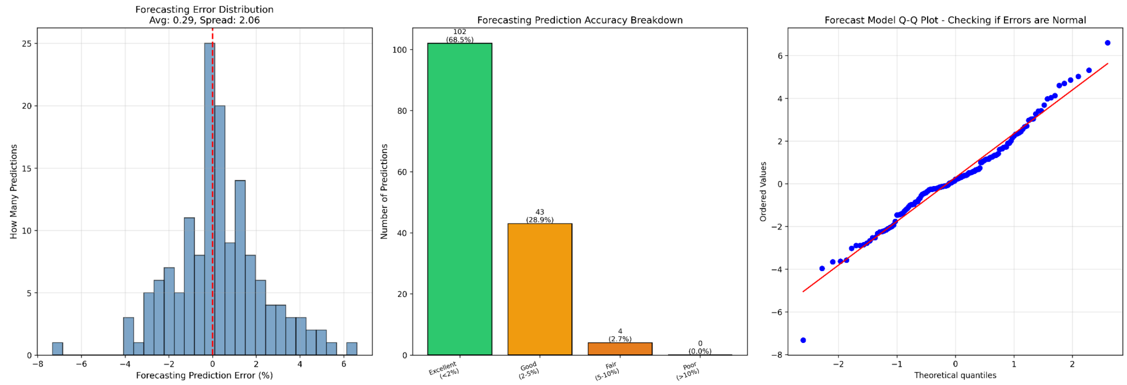
- **January 6-10:** Classification of Factor Types and Key Discoveries
    - Categorized our features into two groups based on their nature and timing:
    - Underlying Factors (structural, long-term causes):
        i. GDP per capita: Reflects overall economic development, takes years to change
        ii. Water access (lagged): Past infrastructure affects accumulated health over time
        iii. Wheat production and volatility: Long-term agricultural stability
        iv. Temperature (lagged): Long-term climate patterns These represent root causes of malnutrition that require sustained effort to address.
    - Immediate Factors (can impact within same year):
        i. Water access (current): New infrastructure can help immediately
        ii. Political stability: Current conflicts disrupt food systems now
        iii. Maize area: Short growing season, can buffer acute food shortages

        iv. Cassava production: Risk-buffer crop and famine reserve
        v. Rice production: Short growing season, immediate food availability
        vi. Current wheat variables: Annual harvest affects food supply now
    - SURPRISING WHEAT DISCOVERY: Wheat consistently outperformed other crops in importance! This was unexpected because:
        i. Maize is the primary staple crop in Sub-Saharan Africa
        ii. We expected maize to dominate agricultural features
        iii. But wheat showed higher importance across production, yield, and volatility
    - Why wheat matters so much:
        i. High nutritional value (more protein and nutrients than maize)
        ii. Significant in global trade (vulnerability to geopolitical events)
        iii. Climate-sensitive (indicator of broader agricultural stress)
        iv. Annual harvest makes it vulnerable to droughts and climate shocks
        v. Stability of wheat supply reflects overall food system resilience
    - This classification helps inform policy: Short-term interventions should target immediate factors (water access, political stability, cassava/maize/rice supply), while long-term development should address underlying factors (GDP, sustained water infrastructure, wheat resilience).
- **January 11-17:** Residual Analysis and Model Validation
    - Conducted residual analysis to verify our models were performing well and not overfitting.
    - What are residuals? Residual = Actual stunting rate - Predicted stunting rate
    - If a model is good, residuals should be:
        i. Randomly scattered around zero (no systematic patterns)
        ii. Normally distributed
        iii. Similar across all prediction ranges
    - Created residual plots for all four models:



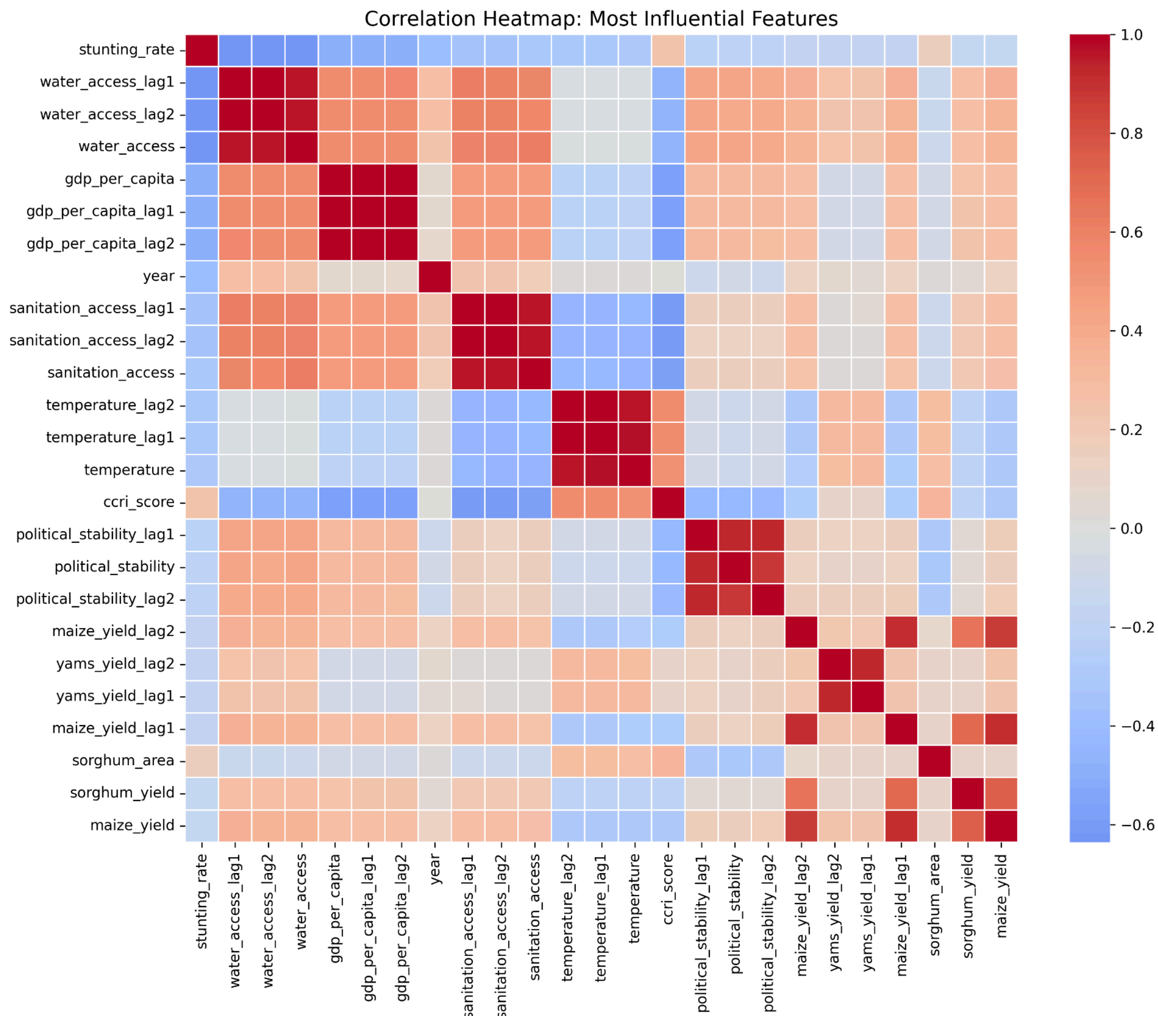
-
- Both RF and KNN showed similar residual distributions, indicating neither was overfitting (memorizing training data) or underfitting (missing important patterns).
- Also created Q-Q plots (Quantile-Quantile):

- 



  - These compare our residuals to theoretical normal distribution. Points falling along the diagonal line indicate normally distributed errors, which is ideal for statistical validity.
  - Result: More than 68% of predictions fell within ±2 percentage points of observed stunting rates. This confirms high reliability across both algorithms.

- **January 18-24:** Correlation Analysis and Poster Creation
  - Conducted correlation analysis: Created comprehensive correlation heatmap analyzing relationships between features and stunting rates.



-

- Strong negative correlations observed between stunting rate and:
    i. GDP per capita (current and lagged): Higher GDP causes lower stunting
    ii. Water access (current and lagged): Better water access causes lower stunting
    iii. Sanitation access (current and lagged): Better sanitation causes lower stunting
    iv. Political stability (current and lagged): More stability causes lower stunting
    v. Wheat production: Higher wheat production causes lower stunting
- This validates our feature importance findings.
- Started to create a scientific poster/trifold display for our project

- **January 25-31:** Manuscript Writing Begins
    - Started writing scientific manuscript to document our research findings.
    - Manuscript structure decided:
        i. Abstract (summary of entire project)
        ii. Introduction (background, problem statement, objectives)
        iii. Methods (data sources, processing, feature engineering, algorithms)
        iv. Results (model performance, feature importance, correlation analysis)
        v. Discussion (interpretation, policy implications, limitations)
        vi. Conclusion (summary of findings)
        vii. References (all cited sources)
        viii. Appendix (complete feature table)
    - Challenges encountered:
    - We tried to incorporate peer-reviewed sources only (not just websites or datasets)
    - Started with 22 peer-reviewed sources covering:
    - Child malnutrition statistics and impacts
    - Machine learning methods (Random Forest, KNN)
    - Data quality and survey limitations
    - Climate impacts on agriculture
    - Socioeconomic determinants of malnutrition
    - Specific crop biology (maize, wheat, cassava, rice growing seasons)

February 2026

- **February 1-7:** Starting to update the CYSF platform
    - Writing the different sections, adding most of what was in my manuscript to the platform.
    - Finishing up the poster design, buying materials for decorating the trifold
    - Finishing up the manuscript, using latex/overleaf to write it
    - Developed and deployed an interactive web application using Streamlit and Streamlit Community Cloud.

        i. The application allows users to select any Sub-Saharan African country and explore country-specific Random Forest feature importance, historical stunting rate trends, the top five most influential predictors with correlation direction, and a categorical breakdown of driver types. The app is publicly accessible at: https://stunting-prediction-app.streamlit.app/
        ii. Includes country-wise feature importance
        iii. Learned about and utilized recursive forecasting to tackle data unavailability issues.

- **February 8-15:** Polishing up everything
    - Practicing 3-5 minute “summaries” of our project
    - Practicing answering judge’s questions
    - Debugged several technical issues with the web application, including resolving an error caused by duplicate features in the column list.
- **February 16-23:**
    - Making sure we have all required elements
    - Continued refining the web application — fixed bar chart labels to clearly display lag timing (e.g., "Yams Production (lag 1)" vs "Yams Production") so duplicate-looking features are distinguishable. Updated the category breakdown pie chart to display only the top 3 categories for clarity.
    - Added an interactive map to click on countries
    - Added a predicted future stunting rate
- **February 23-28:**
    - Recording the 10-minute presentation
    - Emailing University of Calgary data science, health, nutrition, etc professors for feedback
        i. Essentially any relevant professors
    - We also finished updating the platform

i. There was an error on the platform where we couldn't select our project topics so we did have to email quite a few times about that

March 2026

- **March 1-4:**
  - We received a few responses back from the professors, and we will be updating our projects according to their feedback.
  - We will talk about this more at the actual CYSF - what we implemented, what feedback they gave us, etc.

Jonathan Hudson
to me, aizakhwaja2010@gmail.com

Hi,
This looks very impressive. What kind of feedback were you looking for Chenwei?
Thanks,
Jonathan Hudson

  -

C

Chel Hee Lee
to me

Mon, Mar 2, 9:53 AM (1 day ago)

Dear Chenwei,

I appreciate you reaching out to me, and I am very impressed by your work.

Can I ask how you found me?

Unfortunately, I am not the right person to review your work, as I do not have sufficient domain knowledge. My field is knowledge representation and engineering using biomedical data in an acute care environment. Which part do you need feedback on? Data engineering, predictive modeling, or economic utility of your work? My suggestion is to search for faculty mentors in the community health science or economics departments for health economics or equality.

Best,

Chel

Annotated Bibliography Of Potential Datasets and Sources: Created as part of Background Research

1.
Potential Impacts of Climate Change on World Food Supply: Datasets from a Major Crop Modeling Study | NASA Earthdata. NASA Earthdata [Internet]. 2025 Jun 18 [cited 2025 Oct 13]; Available from:
https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-cropclim-giss-db-1.00

Summary
This dataset provides projected changes in global and regional grain crop yields under multiple climate change scenarios from 1995 to 2110. Using outputs from General Circulation Models (GCMs) and varying levels of farmer adaptation, the data estimate yields for major crops, wheat, rice, coarse grains, and soybeans across 125 agricultural sites. Results were aggregated into global trade models to assess potential changes in food prices, shortages, and hunger risk. The data were developed by NASA's Goddard Institute for Space Studies (GISS) and distributed through Columbia University's CIESIN.

Relevance
This dataset will lay the foundation for our project for analyzing how climate change can alter agricultural productivity and food security. It supports sections of the report related to climate-driven changes in crop yields, regional food supply vulnerability, and economic impacts of climate stress on global trade. The dataset's large range of time periods allows for comparison between historical trends and future projections, making it valuable for time-series or predictive modelling. We can use this to benchmark our results against NASA's results, since part of our model is predicting the results of climate change on crop yield.

Limitations
The dataset is based on climate and crop models from the late 1990s, which may not reflect recent advances in climate projection or agricultural technology. It focuses primarily on four crop categories, potentially omitting other critical food sources. In addition, while it incorporates adaptation scenarios, these are simplified and may not capture the full range of farmer responses or socioeconomic dynamics. As a result, newer datasets may be required for cross-validation.

2.
FAOSTAT [Internet]. Available from: https://www.fao.org/faostat/en/#home

Summary
A comprehensive database providing statistics on various aspects of food and agriculture,

including crop yield data, malnutrition, and the impacts of climate change, based on year and country. The data is built from official national statistics using various published sources.

Relevance

FAOSTAT will be useful when connecting crop yield data with emission statistics. It provides datasets in relation to food security, production, emissions and other socioeconomic statistics, which are all key parts of our predictive model. The catalogue is based on a large period of time, allowing for comparisons between past and present trends. Overall, FAOSTAT will act as a crucial source during the training and testing process.

Limitations

Though presenting statistics for many different aspects of the issue, FAOSTAT provides data at a national level, potentially excluding regional and local trends within a country. Similarly, data is showcased on a year-to-year basis, meaning it might not be beneficial for short-term predictive models. This database relies heavily on officially reported data from different countries, meaning gaps could exist depending on the quality and accuracy of reports from different national statistics offices.

3.

Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data [Internet]. 2015 Dec 7 [cited 2025 Oct 13];2(1). Available from: https://www.nature.com/articles/sdata201566

Summary

This paper introduces CHIRPS, a global rainfall dataset that merges satellite imagery with ground station data to produce a long-term, high-resolution record of precipitation. It covers the period from 1981 to the present and is updated frequently, providing both near-real-time and finalized products. The dataset helps monitor droughts and rainfall variability, especially in regions that lack dense weather station coverage. The authors validate CHIRPS by comparing it with other precipitation datasets and show that it performs particularly well for tracking rainfall patterns that influence agriculture and food security.

Relevance

CHIRPS will be useful for analyzing how changing rainfall patterns might affect crop yields and food supply, especially in climate-vulnerable regions like Africa and South Asia. Because it offers detailed spatial and temporal data, it can help identify drought trends and extreme rainfall events over time. In our report, I can use CHIRPS as a key climate variable by using it to link precipitation changes to agricultural productivity or food insecurity risks in different countries.

Limitations
While CHIRPS is reliable and widely used, it only measures precipitation, so it doesn't capture other important climate variables like temperature or soil moisture. Its accuracy also depends on the availability of ground station data, meaning that in areas with limited observations, results may be less precise. Lastly, since the study dates back to 2015, newer methods may now provide slightly improved rainfall estimates, so cross-checking with updated datasets would be beneficial.

4.
The Climate Crisis is a Child Rights Crisis [Internet]. UNICEF. 2021. Available from: https://www.unicef.org/reports/climate-crisis-child-rights-crisis

Summary
The Children's Climate Risk Index (CCRI) is an index that uses various datasets to rank countries in terms of child exposure to climate risks. It is based on climate shocks and stresses as well as child susceptibility, combining the impacts of climate change with humanitarian challenges affecting children. The index is made up of two primary pillars, child vulnerability and climate risk, providing individual rankings for both of these as well as the combined CCRI score. The key target of this resource is to provide an estimate of the likelihood of climate change leading to a detrimental impact on the progress of children.

Relevance
This resource provides data on one of the key aspects of our predictive model, linking climate hazards, such as sudden floods and heatwaves, to child vulnerability. Each nation, including the entirety of the Sub-Saharan African region, is detailed, making it relevant to our cause. The CCRI rankings could potentially be used as features for the machine learning model.

Limitations
The CCRI only makes use of datasets until 2021, the year of its publication. This is a key limitation, as the climate crisis is one that is rapidly evolving. In addition, the data is aggregated at a national level, but regional and local trends must also be factored in as part of our predictive model. CCRI is an index, meaning it may need normalization in order to be compared to other variables, such as rainfall and crop yield.

5.
FAO Knowledge Repository [Internet]. Fao.org. 2025 [cited 2025 Oct 14]. Available from: https://openknowledge.fao.org/items/445c9d27-b396-4126-96c9-50b335364d01

Summary
The State of Food Security and Nutrition in the World 2023 is a comprehensive report by FAO, IFAD, UNICEF, WFP, and WHO that evaluates global progress toward ending hunger and all

forms of malnutrition, aligning with Sustainable Development Goal 2. It highlights that, despite some positive trends, the world is not on track to meet these targets by 2030.

Relevance

This source is essential for understanding the broader context of child malnutrition in Sub-Saharan Africa. It provides data that can justify the study's focus on climate-driven crop yield declines as a key factor influencing food insecurity and malnutrition. The regional statistics and trend analysis can help link projected crop yield impacts from climate models to observed nutritional outcomes.

Limitations

The data are aggregated at national or regional levels, which may obscure local variations in food insecurity or malnutrition. The report does not provide direct connections between crop yields and child malnutrition, so additional datasets (e.g., FAOSTAT, CHIRPS) are needed to model specific causal relationships. It also relies on available reporting, so areas with limited monitoring may be underrepresented.

6.

Constenla-Villoslada S, Liu Y, McBride L, Ouma C, Mutanda N, Barrett CB. High-frequency monitoring enables machine learning–based forecasting of acute child malnutrition for early warning. Proceedings of the National Academy of Sciences [Internet]. 2025 Jun 6;122(23). Available from: https://www.pnas.org/doi/10.1073/pnas.2416161122

Summary

This paper places emphasis on the impact of acute malnutrition worldwide. It explores the way high-frequency monitoring, as well as monthly time series child anthropometric data from sentinel site,s can be leveraged to develop a better predictive model for acute child malnutrition early warnings. The author details how advances in computational methods should allow for reliable early warning systems (EWS); however, current models have exhibited low skill in terms of predicting local acute malnutrition trends. The key goal of this study is to create a model that can accurately forecast malnutrition to enhance targeting of humanitarian response to impending food crises. Focusing specifically on the country of Kenya, this model can predict with respect to certain wards in the country.

Relevance

The study focuses specifically on early warning systems for malnutrition in order to aid humanitarian response, which is highly relevant to our project. This research provides a valuable perspective on the importance of local regions, specific communities, and vulnerable populations when creating an Early Warning System for malnutrition, something which could be applicable to our model. It provides datasets as well as information on machine learning methods used to create the machine learning tool, detailing key background information.

Limitations

The study relies heavily on data from sentinel sites, meaning model performance depends heavily on the quality and consistency of data collection. Though emphasizing the importance of vulnerable populations and local communities, some of the techniques leveraged might not work well when predicting for the entire region of Sub-Saharan Africa. This paper shows strong results for Kenya; however, it is uncertain whether the model will generalize well in other socioeconomic and geographic contexts where similar monitoring is not as widely available.

7.

Wheeler T, von Braun J. Climate Change Impacts on Global Food Security. Science [Internet]. 2013 Aug 2 [cited 2025 Oct 14];341(6145):508–13. Available from: https://www.science.org/doi/10.1126/science.1239402

Summary

This paper reviews the evidence linking climate change to global food security. The authors show that climate change is expected to disrupt crop productivity, particularly in tropical regions like Sub-Saharan Africa and South Asia, where hunger and undernutrition are already prevalent. Climate impacts are not limited to food availability; they also affect access, utilization, and stability through changes in incomes, food prices, health, and water availability. The paper emphasizes the need for climate-smart food systems and investments in adaptation to mitigate these effects.

Relevance

This source is directly relevant for understanding how climate-induced changes in crop yields can lead to child malnutrition in Sub-Saharan Africa. It provides a framework for linking biophysical crop impacts to socioeconomic consequences, including household food access and nutrition, which is essential for modelling malnutrition risk under climate scenarios. The discussion of regional vulnerabilities helps justify a focus on areas most at risk from climate-driven agricultural declines.

Limitations

The paper offers a broad, global overview rather than a fine-scale, country-specific analysis. Many conclusions are based on projections with inherent uncertainty, and impacts on specific crops, micro-level food access, and diet quality are less well quantified. It emphasizes theoretical linkages and general trends, so additional datasets (e.g., crop yields, child nutrition data) are needed to perform precise, quantitative modelling.

8.
Monitoring and Surveillance Nutrition and Food Safety (MNF). Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: key findings of the 2025 edition [Internet]. 2025. Available from: https://www.who.int/publications/i/item/9789240112308

Summary
This UNICEF/WHO/World Bank Group joint child malnutrition estimates edition presents worldwide data on child stunting, wasting, and overweight indicators. The report describes global, regional and national trends for the period between 2000-2024. It emphasizes the need for more extensive efforts to meet global stunting and wasting targets. Trends, percentages and prevalence according to region have all been provided.

Relevance
The report details statistics about malnutrition and is relevant for understanding key terms and trends with regard to child malnutrition in Sub-Saharan Africa. Undernourishment in children is one of the key dependent variables for our predictive model. Various socioeconomic factors have been highlighted, which could act as independent variables.

Limitations
The data for this edition is aggregated mostly at a regional scale, meaning local and country-specific trends are not emphasized. This is one of the central drawbacks of the report's relevance to this project. In addition, many charts, graphs, and figures are provided, but not much raw data is included.

9.
FAO. Climate change and food security: risks and responses [Internet]. 2015. Available from: https://openknowledge.fao.org/server/api/core/bitstreams/a4fd8ac5-4582-4a66-91b0-55abf642a400/content

Summary
This FAO report examines the multifaceted impacts of climate change on food security, emphasizing how disruptions in ecosystems, agriculture, and food systems can undermine nutrition and livelihoods. It highlights that climate-induced changes, such as altered precipitation patterns, increased temperatures, and extreme weather events, can lead to reduced crop yields, particularly in regions like Sub-Saharan Africa. The report underscores the importance of adaptive strategies, including climate-smart agriculture and resilient food systems, to mitigate these effects and ensure food security.

Relevance
This report talks about climate-induced crop yield declines and child malnutrition in Sub-Saharan Africa. It provides a detailed analysis of how climate change affects staple crops, such as maize and cassava, through altered rainfall patterns and increased temperatures. These crop-specific insights are crucial for modelling potential yield reductions and assessing the subsequent risks to food availability and child nutrition in the region.

Limitations
This FAO report examines how climate change impacts food security in Sub-Saharan Africa, focusing on staple crops like maize and cassava. It shows that shifts in rainfall, temperature, and extreme weather can reduce yields, threatening food availability and child nutrition. While the report is useful for linking climate impacts to malnutrition risk, it may lack detailed country-level data, may not reflect the very latest climate models, and some figures are aggregated, overlooking local variations.

10.
IPC - Integrated Food Security Phase Classification [Internet]. Available from: https://www.ipcinfo.org/

Summary
The Integrated Food Security Phase Classification (IPC) is a global initiative that provides clear analysis of food insecurity and malnutrition across different regions and countries. It is widely used by many governments, UN agencies and NGOs, acting as an early warning system for food insecurity. The IPC uses a standardized classification system, labelling each region and area on a scale of phases 1-5, ranging from minimal to famine. It provides two datasets on acute food insecurity and chronic food insecurity.

Relevance
The IPC acts as a central early warning system for many agencies, providing us with information of what is currently a norm in the industry. It serves as a foundational reference for food security and early warning. It acts as a benchmark for evaluating newer, machine learning based predictive models. The data provided by this source can also be utilized when developing our model.

Limitations
The IPC only publishes data from certain regions and areas, and may not be able to provide concrete data for the entire region of Sub-Saharan Africa at a given point of time. As illustrated in other papers and studies, this source only provides data on entire regions, but not specific communities and vulnerable populations. This can be a limitation in terms of tailoring

humanitarian aid to specific groups or identifying local hotspots. A well-developed predictive model could be very beneficial in this scenario, allowing for predictions months in advance.

11.
Hunger Hotspots FAO-WFP early warnings on acute food insecurity. 2024 [cited 2025 Oct 25]; Available from: https://docs.wfp.org/api/documents/WFP-0000162510/download/

Summary
This report provides an early-warning analysis of acute food insecurity across 22 countries, including multiple Sub-Saharan African regions. It identifies the main "hunger hotspots" where conflict, economic instability, and climate shocks, such as floods and droughts linked to La Niña, are projected to worsen food shortages between late 2024 and mid-2025. The report integrates rainfall forecasts, conflict data, and humanitarian access indicators to project where food insecurity is most likely to reach "Emergency" or "Catastrophe" levels on the IPC/CH scale. It highlights countries like Sudan, South Sudan, Mali, Chad, Niger, and Mozambique as facing the highest risk of famine or severe food insecurity due to the combined effects of conflict and climate variability.

Relevance
This source is directly relevant for validating the humanitarian implications of our predictive model. It provides recent, country-level projections of food insecurity in Sub-Saharan Africa and identifies the climatic and socioeconomic drivers behind them. The data can serve as a real-world benchmark to test the accuracy of our model's predictions of malnutrition or yield-related hunger risk. It also helps link modelled crop yield declines and rainfall variability to observed child nutrition outcomes, reinforcing the causal connection between climate stress and humanitarian crises.

Limitations
The report focuses on short-term (six-month) humanitarian forecasts rather than long-term climate or agricultural projections. Many figures are aggregated at the national or regional level and combine multiple drivers (conflict, displacement, economics), which can make it difficult to isolate the specific contribution of climate-driven crop failures. Data coverage may also vary by country depending on the availability of the analyses.

12.
Amegbor PM, Kaira Lapurga, Carr E, Guerrero M, Oyinkansola Babayode, Crisci SE, et al. Exploring the effect of early-life climate anomalies on child growth in the Sub-Saharan African context: Insights from the demographic & health survey. The Science of The Total Environment [Internet]. 2025 May 20 [cited 2025 Oct 25];983:179658–8. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0048969725012999

Summary

This study investigates how climate anomalies during prenatal and early childhood periods affect stunted growth among children under five across 30 Sub-Saharan African countries. Using data from the Demographic and Health Surveys (DHS) merged with temperature and precipitation anomaly data (2004-2021), the researchers applied Bayesian hierarchical models to assess both linear and nonlinear relationships. Results show that increases in maximum temperature anomalies significantly raise the likelihood of stunting, while children living in Tropical Monsoon zones had lower stunting rates. Approximately one-third (34.5%) of the sample children were stunted, highlighting the widespread prevalence of malnutrition linked to environmental variability.

Relevance

This paper provides strong empirical evidence connecting early-life climate shocks, particularly temperature anomalies, to child malnutrition outcomes in Sub-Saharan Africa. It demonstrates a measurable causal relationship between climate stressors and child growth. The use of DHS data aligns well with our goal of integrating health and climate variables for predictive modeling. We could use the study's statistical framework or variables (temperature deviations, ecological zone classification) to guide feature selection or validation for our model.

Limitations

While the study's large sample size and multi-country approach strengthen its findings, it still relies on historical data (2004-2021), which may not fully reflect recent climate extremes. The analysis focuses primarily on stunting and does not cover other forms of malnutrition like wasting or underweight. Additionally, the reliance on DHS and modeled climate data may introduce measurement inconsistencies across countries, and causal directionality, while suggested, cannot be fully confirmed.

13.
Gebeyehu L, Gemechu A, Bedemo A. Effects of climate change on food security in Africa: Meta-analysis. Journal of Agriculture and Food Research [Internet]. 2025 Oct [cited 2025 Oct 25];23:102214. Available from:
https://www.sciencedirect.com/science/article/pii/S266615432500585X

Summary

This paper performs a meta-analysis of existing studies on the impacts of climate change on food security across African countries. It synthesizes results from multiple studies to quantify how temperature increases, rainfall variability, and extreme weather events affect crop yields, livestock productivity, and overall food availability. The authors highlight regional differences, with Sub-Saharan Africa experiencing higher vulnerability due to reliance on rainfed agriculture. The study also discusses the lagged effects of climate shocks on food security and child nutrition outcomes.

Relevance

This source links climate change to food insecurity across Africa. It complements country-specific and local studies by offering synthesized effect sizes and trend estimates, which can inform feature selection for predictive modelling of malnutrition risk. The meta-analysis framework helps justify the importance of climate variables (temperature anomalies, rainfall deviations) in our model and provides context for regional disparities.

Limitations

Being a meta-analysis, the study relies on the quality and comparability of the original studies. Differences in study design, crop types, and timeframes may introduce heterogeneity. The results are aggregated at a regional level, so fine-scale, country-specific, or local effects on child malnutrition may be underrepresented. Moreover, the analysis focuses primarily on food security outcomes and may not provide direct, measurable indicators of child malnutrition.