AN ASSESSMENT OF THE ENVIRONMENTAL AND SOCIOECONOMIC FACTORS OF FOOD SECURITY OF CLIMATE CHANGE USING A LOGISTIC REGRESSION MODEL

Abubakar Usman[1*], Ibrahim Abubakar Sadiq[1], Aliyu Ismail Ishaq[1], Jamila Abdullahi[1], Buhari Ishaq[1,2], Reuben Oluwabukunmi David[1], Yakubu Aliyu[1] and Jamilu Yunusa Falgore[1]

[1]Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria
[2]Department of Military Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, Kaduna.

Corresponding Author: abubakarusman28@gmail.com

ABSTRACT

Understanding the environmental and socioeconomic drivers of food security is critical in the context of climate change. This study examines the impact of various environmental and socio-economic factors on food security and related health outcomes, utilizing a logistic regression model. A simulated dataset reflecting environmental and socioeconomic characteristics was analyzed using a multivariable logistic regression model. The Events per Variable (EPV) guideline was applied to ensure statistical adequacy, with 7 predictor variables requiring a minimum of 70 positive events. Classification performance was evaluated using accuracy, sensitivity, specificity, precision, F1 score, AUC, and the Hosmer–Lemeshow test. Variance Inflation Factor (VIF) was used to assess multicollinearity, and Least Absolute Shrinkage Selector (LASSO) regularization was implemented to improve model stability. The standard logistic regression model identified income level (OR=1.000, p=0.013), rainfall (OR = 1.514, p = 0.003), and temperature–rainfall interaction (OR = 0.982, p = 0.002) as significant predictor variables. The model achieved high accuracy (96.3%) and AUC (0.938) but exhibited low sensitivity (35.3%). Severe multicollinearity among environmental variables (VIF > 15) prompted the application of LASSO regression, which retained only temperature, its interaction with rainfall, and income level. The LASSO model improved interpretability and specificity (100%) but further reduced sensitivity (11.8%). Environmental variables—especially rainfall and its interaction with temperature play a significant role in food security outcomes. However, high multicollinearity undermines coefficient stability. Regularization via LASSO improved model robustness and parsimony but reduced the ability to detect food-insecure cases. These findings highlight the need for balanced modeling strategies that address both multicollinearity and class imbalance in food security research.

Keywords: *Climate change; Environmental variables; food security; logistic regression model*

1. INTRODUCTION

Climate change continues to reshape global food systems and health outcomes, especially among vulnerable populations. Rising temperatures, fluctuating rainfall patterns, and shifting environmental conditions increasingly threaten food production and public health (Myers et al., 2017). As agricultural yields decline in response to these stressors, food insecurity intensifies, often exacerbated by pre-existing socioeconomic inequalities. Understanding how environmental and socioeconomic variables interact is therefore crucial for informing policies and designing interventions aimed at improving food access and health equity (Schmidhuber & Tubiello, 2007).

Previous studies have demonstrated the direct and indirect effects of environmental variables, particularly temperature and rainfall, on agricultural productivity and food security (Lobell & Burke, 2010; Wheeler & von Braun, 2013). Higher temperatures can accelerate crop maturation while reducing yields, and erratic rainfall can impair soil moisture levels, both of which pose serious risks to food systems. In addition to environmental stressors, socioeconomic factors such as education, income, age, and gender influence how communities adapt to these risks and utilize available resources (Fanzo *et al*., 2020).

However, many of these studies focus either exclusively on environmental indicators or treat socioeconomic data as confounders, rather than exploring the full interaction between these two domains. Moreover, while descriptive analyses have been valuable, inferential modeling approaches, particularly logistic regression, offer a more nuanced understanding of how different factors predict food security outcomes under climatic stress (Beddington *et al*., 2012).

This study builds upon the existing body of research by assessing the independent and interactive effects of temperature, rainfall, and key socioeconomic indicators (education, income, age, and gender) on food security outcomes. Using logistic regression modeling on simulated yet representative data, we estimate odds ratios for each predictor to quantify its influence. By doing so, this study provides a statistically grounded approach to identifying the drivers of food security under changing environmental conditions, intending to guide evidence-based interventions.

2. LITERATURE REVIEW

Schmidhuber & Tubiello (2007) were among the earliest to highlight the potential consequences of climate change on global food security, noting that developing countries would bear the brunt of reduced agricultural productivity. They emphasized that both direct environmental changes such as altered rainfall patterns and rising temperatures, and indirect socioeconomic pressures could disrupt food access and nutritional health.

Building on this, Wheeler and von Braun (2013) examined how climatic shocks like droughts and floods significantly affect food availability and price volatility. Their findings underscored the importance of strengthening climate resilience, especially in regions with fragile food systems and poor adaptive capacity.

Lobell and Burke (2010) contributed empirical evidence showing that rising temperatures negatively impact crop yields, especially in tropical regions. Using regression-based crop models, they found that small increases in temperature during the growing season reduced yields of staples such as wheat and maize, which directly affects household food security.

Beddington *et al*., (2012) introduced a systems-based framework for assessing food and environmental security, advocating for integrated approaches that consider climate, population growth, and policy factors. They suggested that modeling tools, such as logistic regression, can help identify high-risk groups and inform adaptive strategies.

Fanzo *et al.,* (2020) stressed the multidimensional nature of food security, highlighting how education, income, gender, and geographic location influence people's ability to access nutritious food under changing environmental conditions. They recommended combining socioeconomic variables with climate data for better-targeted interventions.

Myers *et al.,* (2017) examined the health consequences of climate-induced changes in agriculture, particularly the reduction in nutritional quality of food crops due to elevated CO_2 levels. They proposed monitoring both environmental and health indicators as a unified framework.

In more recent studies, Béné *et al.,* (2021) applied logistic regression to assess how household characteristics and environmental stressors jointly influence food security across sub-Saharan Africa. The authors found strong associations between climatic variability and household food insecurity, with education and wealth serving as mitigating factors.

Lastly, Tigchelaar *et al.,* (2022) modeled future climate risks to staple crops using high-resolution climate projections. Their study suggested that unless adaptive strategies are adopted especially in low-income countries, climate extremes could double the risk of food insecurity by mid-century. Their statistical approach incorporated both environmental and socioeconomic factors, supporting the need for multidimensional analyses like logistic regression.

While these studies provide strong evidence of the relationship between climate and food security, there remains a gap in studies that integrate both environmental and socioeconomic factors using logistic regression on simulated yet representative datasets. This research therefore, contributes to the existing literature by quantifying the individual and interactive effects of temperature, rainfall, income, education, gender, and age on food security outcomes. By applying a logistic regression model in a controlled, data-driven setting. This study also, offers a nuanced practical framework for understanding how climate and socioeconomic dynamics jointly influence food security and health.

3. METHODOLOGY

This study employed a cross-sectional design using a simulated dataset comprising 200 observations to assess the effect of environmental and socioeconomic factors on food security outcomes. A logistic regression model was used to determine the likelihood of achieving adequate food security based on predictor variables. 5% level of significance was selected for the analysis.

3.1 Simulation Study

Simulated data were generated to reflect realistic distributions derived from previous empirical studies. The dependent variable is food security status (secure/insecure), and the independent variables included the temperature (°C)**,** rainfall (mm)**,** income (in monetary units)**,** Education (years of schooling)**,** Age (years)**,** Gender (0 = female, 1 = male), and temperature × rainfall interaction.

3.2 Model Specification

The logistic regression model is given in Equation (1) as:

$$\log\left(\frac{\mathrm{P}(\mathrm{Y}=1)}{1-\mathrm{P}(\mathrm{Y}=1)}\right)=\beta_0+\beta_1 x_1+\ldots+\beta_n x_n \qquad (1)$$

where:
Y=1 indicates probability of achieving food security,
β_0 = the intercept,
$\beta_1,\ldots,\beta_n$ are the unknown coefficients, and
$x_1,\ldots,x_n$ are the covariates.
This research utelised the following covariates - temperature, rainfall, income, education, age, gender and the interaction between temperature and rainfall. Thus:

$$\log\left(\frac{\mathrm{P}(\mathrm{Y}=1)}{1-\mathrm{P}(\mathrm{Y}=1)}\right)=\beta_0+\beta_1(temp)_1+\beta_2(rf)_2+\beta_3(income)_3+\beta_4(edu)_4+\beta_5(age)_5+\beta_6(gender)_6+\beta_7(temp\times rf) \qquad (2)$$

where: $\mathrm{P}(\mathrm{Y}=1)$ is the probability of achieving food security.

3.3 Sample Size Justification

Sample size selection was guided by the "events per variable" (EPV) rule, with at least 10 outcome events per predictor variable to ensure model stability.
Given the 7 predictor variables and a balanced binary outcome, 350 observations were deemed sufficient.

3.4 Model Evaluation

Model performance was assessed using the following metrics:
Accuracy**:** Proportion of correct classifications
Area Under the Curve (AUC)**:** Discriminatory power of the model
p-values and odds ratios**:** To assess predictor significance and effect size
All analyses were conducted using standard statistical software, and findings were reported with 95% confidence intervals.

4. RESULTS AND DISCUSSION

4.1 Standard Logistic Regression Results

This study fitted a multivariable logistic regression model to examine the influence of environmental and socioeconomic predictors on food security status. Table 1, summarises the coefficient estimates, odds ratios, confidence intervals, and p-values.

Coefficient Estimates and Statistical Significance

Table 1, Presents the Logistic Regression Results for Food Security Factors

Variable	Estimate (Log-Odds)	Odds Ratio	95% CI	P-Value
Intercept	-25.800	0.000	0.000–0.772	0.035
Temperature	0.848	2.335	0.865–5.581	0.069
Rainfall	0.415	1.514	1.148–2.005	0.003**
Temp × Rainfall	-0.018	0.982	0.971–0.994	0.002**
Education Level	0.211	1.234	0.683–2.254	0.484
Income Level	0.000	1.000	1.000–1.000	0.013*
Age	0.006	1.006	0.961–1.055	0.806
Gender	0.434	1.543	0.444–5.522	0.493

From Table 1, rainfall, income level, interaction between temperature and rainfall were statistically significant predictors of food security status. The negative sign on the interaction term suggests that higher rainfall mitigates the positive effect of temperature on the outcome.

4.2 Model Performance Metrics

The performance of the logistic regression model was evaluated using several classification metrics, shown in Table 2.

Table 2: Classification Performance (Standard Logistic Model)

Metric	Value
Accuracy	0.963
Sensitivity	0.353
Specificity	0.994
Precision	0.750
F1 Score	0.480
AUC	0.938
Hosmer–Lemeshow p-value	0.951

The model demonstrates high overall accuracy (0.963) and specificity (0.994), but low sensitivity (0.353). This indicates that it is better at identifying food-secure cases than food-insecure ones, likely due to class imbalance.

4.3 Multicollinearity Assessment

Table 3, presents the variance inflation factors (VIFs) calculated in order to assess multicollinearity among predictor variables.

Table 3: Variance Inflation Factors (VIF)

Variable	VIF
Temperature	15.18
Rainfall	94.41
Temp × Rainfall	70.55
Education Level	1.03
Income Level	1.03
Age	1.13
Gender	1.08

From Table 3, severe multicollinearity was observed among the environmental predictors, particularly rainfall, temperature, and their interaction. These high VIFs suggest unstable coefficient estimates and inflated standard errors, warranting the use of regularization techniques such as LASSO.

4.4 LASSO Logistic Regression

To mitigate multicollinearity and improve generalizability, the study applied the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression to select the important predictor variables. Table 4 reports the important retained predictor variables.

Table 4: LASSO Logistic Regression Coefficients

Variable	Estimate (Log-Odds)
Intercept	8.6901
Temperature	-0.4911
Temp × Rainfall	-0.0004
Income Level	0.0000

LASSO eliminated Education Level, Income Level, Age, and Gender by shrinking their coefficients to zero, retaining only a subset of predictors (Temperature, Temperature and rainfall interaction) that contributed most to model performance, thereby improving interpretability and reducing variance.

4.5 LASSO Model Performance

Table 5: Classification Metrics (LASSO Model)

Metric	Value
Accuracy	0.957
Sensitivity	0.118
Specificity	1.000
Precision	1.000
F1 Score	0.211
AUC	0.895

The LASSO model yielded slightly reduced AUC and sensitivity compared to the standard model, but perfect specificity and precision, suggesting it was highly conservative in predicting food insecurity.

5. CONCLUSION

This study investigated the effects of environmental and socioeconomic factors on food security status using both standard and regularized logistic regression models. The standard logistic regression model in Table 1, identified rainfall (p = 0.003), income level (p=0.013), and the interaction between temperature and rainfall (p = 0.002) as statistically significant predictors at 5% level of significance, suggesting that increased rainfall moderates the adverse effects of temperature on food insecurity. Despite the model's strong overall classification performance as shown in Table 2 (accuracy = 96.3%, AUC = 0.938) Table 2 also, exhibited low sensitivity (35.3%), indicating that a substantial number of food-insecure cases were not correctly classified. Moreover, multicollinearity diagnostics revealed extremely high VIF values for environmental predictors (VIF for Rainfall = 94.41), suggesting instability in the estimated coefficients and a risk of overfitting.

To address this, a LASSO logistic regression model was implemented, which selected only temperature, temperature–rainfall interaction, and income Level as important predictor variables, and shrunk the other variables to zero. While the LASSO model maintained a high specificity (100%) and precision (100%), its sensitivity dropped to 11.8%, and the F1 score decreased to 0.211, indicating a trade-off between model simplicity and the ability to detect positive cases.

REFERENCES

Beddington, J., Asaduzzaman, M., Clark, M., Fernández, A., Guillou, M., Jahn, M.. & Wakhungu, J. (2012). Achieving food security in the face of climate change: Final report from the Commission on Sustainable Agriculture and Climate Change. *CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).*

Béné, C., Fanzo, J., Prager, S. D., Acharya, T., & Mapes, B. R. (2021). Global drivers of food system (un)sustainability: A multi-country correlation analysis. *PLOS ONE*, *16*(4), e0249035.

Fanzo, J., Davis, C., McLaren, R., & Choufani, J. (2020). The effect of climate change across food systems: Implications for nutrition outcomes. *Global Food Security*, *26*, 100397.

Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, *150*(11), 1443–1452.

Myers, S. S., Smith, M. R., Guth, S., Golden, C. D., Vaitla, B., Mueller, N. D., ... & Huybers, P. (2017). Climate change and global food systems: Potential impacts on food security and undernutrition. *Annual Review of Public Health*, *38*, 259–277.

Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. *Proceedings of the National Academy of Sciences*, *104*(50), 19703–19708.

Tigchelaar, M., Battisti, D. S., Naylor, R. L., & Ray, D. K. (2022). Future warming increases probability of globally synchronized maize production shocks. *Nature Climate Change*, *12*(1), 36–43.

Wheeler, T., & von Braun, J. (2013). Climate change impacts on global food security. *Science*, *341*(6145), 508–513.