

# A machine learning-based exploration of resilience and food security

Alexis H. Villacis<sup>1</sup> | Syed Badruddoza<sup>2</sup> | Ashok K. Mishra<sup>3</sup>

<sup>1</sup>Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, Ohio, USA

<sup>2</sup>Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, Texas, USA

<sup>3</sup>Morrison School of Agribusiness, W.P. Carey School of Business, Arizona State University, Mesa, Arizona, USA

## Correspondence

Alexis H. Villacis, Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH 43210, USA.  
Email: [villacis.9@osu.edu](mailto:villacis.9@osu.edu)

**Editor in charge:** Gopinath Munisamy

[Correction added on 14 November 2024, after first online publication: The article classification has been updated in this version.]

## Abstract

Leveraging advancements in remote data collection and using the Food Insecurity Experience Scale (FIES) as a proxy measure of resilience, we show that machine learning models (such as Gradient Boosting Classifier, eXtreme Gradient Boosting, and Artificial Neural Networks), can predict resilience with relatively high accuracy (up to 81%). Key household-level predictors include access to financial institutions, asset ownership, the adoption of agricultural mechanization as evidenced by the use of tractors, the number of crops cultivated, and ownership of nonfarm enterprises. Our analysis offers insights to researchers and policymakers interested in the development of targeted interventions to bolster household resilience.

## KEYWORDS

Ethiopia, Food Insecurity Experience Scale, Malawi, Nigeria, predictive performance, Uganda

## JEL CLASSIFICATION

C52, C83, O12, Q18

*Life doesn't get easier or more forgiving; we get stronger and more resilient.*

Steve Maraboli (2009) *Life, the Truth, and Being Free.*

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Applied Economic Perspectives and Policy* published by Wiley Periodicals LLC on behalf of Agricultural & Applied Economics Association.

Resilience has emerged as a prominent policy priority for sustainability and development (Jones et al., 2021). Humanitarian and development agencies, along with researchers and practitioners, are increasingly emphasizing resilience to develop long-term strategies that tackle the effects of climate change, conflict, and epidemics (Knippenberg et al., 2019). Defined, from a normative standpoint, as “the ability to achieve and maintain an acceptable standard of well-being even in the face of shocks and stressors” (Barrett & Conostas, 2014), the concept of resilience is now prominently featured in large-scale sustainable development investments. These investments aim to support households and communities in coping with diverse shocks and stressors that undermine poverty reduction and food security efforts (Walsh-Dilley et al., 2016).

The increasing emphasis on sustainable development investments (Mullan et al., 2018) underscores the necessity for robust empirical evidence to elucidate the interplay between well-being and shocks. Advancements in remote data collection, earth observations, and big data analytics offer promising avenues for gaining new insights into the dynamics of resilience (Knippenberg et al., 2019). Prominently, the application of machine learning algorithms provides opportunities to identify more accurate predictors of resilience and vulnerable areas of concern (Jones et al., 2021; Lieslehto et al., 2022).

While machine learning algorithms have gained traction in predicting food insecurity (Balashankar et al., 2023; Foini et al., 2023; Hossain et al., 2019; Martini et al., 2022; Villacis et al., 2023; Yeh et al., 2020), their application in forecasting household resilience remains limited. However, leveraging machine learning models and big data holds the potential to improve the precision of resilience prediction, thereby offering valuable insights for decision-making processes aimed at supporting households in their recovery from adverse shocks.

The present study expands the currently limited knowledge of machine learning-based examinations of resilience (Garbero & Letta, 2022; Knippenberg et al., 2019) by presenting novel evidence from smallholder farmers from various African countries. By employing comprehensive data obtained from the Harmonized Phone Surveys conducted by the World Bank Living Standards Measurement Study (LSMS) in Ethiopia, Malawi, Nigeria, and Uganda in 2020, we leverage the exogenous shocks induced by the COVID-19 pandemic to explore the potential of machine learning in enhancing the understanding of farm-household resilience dynamics.

Building upon the definition from Barrett and Conostas et al. (2014), we proxy “an acceptable standard of well-being” with “an acceptable level of food security,” and subsequently utilize the Food Insecurity Experience Scale (FIES) for our purposes. Our efforts focus on employing various machine learning models to forecast resilience status and identify key predictors of resilience. Contrary to Knippenberg et al. (2019) but in accordance with Garbero and Letta (2022), we frame resilience prediction as a classification task rather than a regression problem. This choice is motivated by the primary objective of identifying households that demonstrate resilience in the face of exogenous shocks and determining the key features associated with their resilience. Results show that—within the context of our selected group of African nations—resilience can be predicted with relatively high accuracy (between 78% and 81%) using machine learning models. More importantly, we find that key household-level predictors of resilience include access to financial institutions, ownership of assets, risk and income diversification as evidenced by the number of crops cultivated and the ownership of nonfarm enterprises, and finally, the adoption of agricultural mechanization, as evidenced by the use of tractors for agricultural activities.

Our study makes two distinct contributions to the existing knowledge on resilience building strategies in the face of shocks. Firstly, from an academic standpoint, we expand upon existing

literature by employing novel proxy measures of resilience as well as novel machine learning models to forecast resilience. While previous studies utilized the Coping Strategy Index (CS) (Knippenberg et al., 2019) and the Ability to Recover from Shocks Index (ATR) (Garbero & Letta, 2022), we utilize the FIES as a proxy measure. Additionally, we incorporate robust machine learning models such as Gradient Boosting Classifier, eXtreme Gradient Boosting, and Artificial Neural Networks, which have not been extensively explored in previous resilience prediction research.<sup>1</sup> We train these models to use current household features to predict their resilience to shocks.

Secondly, the identification of significant predictors of resilience provides valuable insights for both researchers and policymakers. Our findings enhance our understanding of the predictors of resilience within our selected group of African countries, deepening knowledge regarding the factors associated with households' ability to withstand and recover from shocks. From a policy perspective, these results hold practical implications as policymakers can utilize the identified predictors to guide the development and implementation of research programs aimed at better understanding and enhancing household resilience.

The remainder of this paper is structured as follows. Section 1 outlines the context and essential definitions, describes the data utilized, and presents summary statistics. Section 2 examines the machine learning methods and approaches employed in this study. In Section 3, we present our research findings, emphasizing the use of machine learning methods for resilience prediction and identifying key predictors. Lastly, Section 4 concludes the paper and discusses the policy implications derived from our findings.

## CONTEXT, DEFINITIONS, AND DATA

To explore the potential of machine learning in advancing the analysis of household resilience, we leverage the shocks induced by the COVID-19 pandemic. In addition to the detrimental health outcomes such as morbidity and mortality, the pandemic led to travel restrictions, quarantine measures, business closures, and school suspensions in various regions (Hsiang et al., 2020). These adverse shocks had substantial economic implications, resulting in a global economic contraction (Blake & Wadhwa, 2020). Consequently, food security and access to essential medicines and staple foods were impacted in low-income countries (Josephson et al., 2021).

Of interest in this study is the food security status of households during exogenous shocks like the pandemic. Given that food security is one of the most widely recognized indicators of well-being (Pinstrup-Andersen, 2009), changes in food security status resulting from the adverse shocks of the pandemic provide an ideal context to investigate resilience.<sup>2</sup> To study changes in the food security status of households during the pandemic, we use data from high-frequency phone surveys conducted in Ethiopia, Malawi, Nigeria, and Uganda during 2020. The phone surveys were supported by the World Bank since the outset of the pandemic, motivated by the suspension of regular in-person data collection (Rudin-Rush et al., 2022).<sup>3</sup> Households interviewed by phone represent a subset of the complete study sample the World Bank team used to interview in person in each country as part of their Living Standards Measurement Study.

The choice to focus on Ethiopia, Malawi, Nigeria, and Uganda in this study was based on the next specific criteria: (i) the presence of an official first case of COVID-19 reported in the country during February, March, or April of 2020, (ii) the availability of publicly accessible survey data containing food security information from the early stages of the pandemic (May or

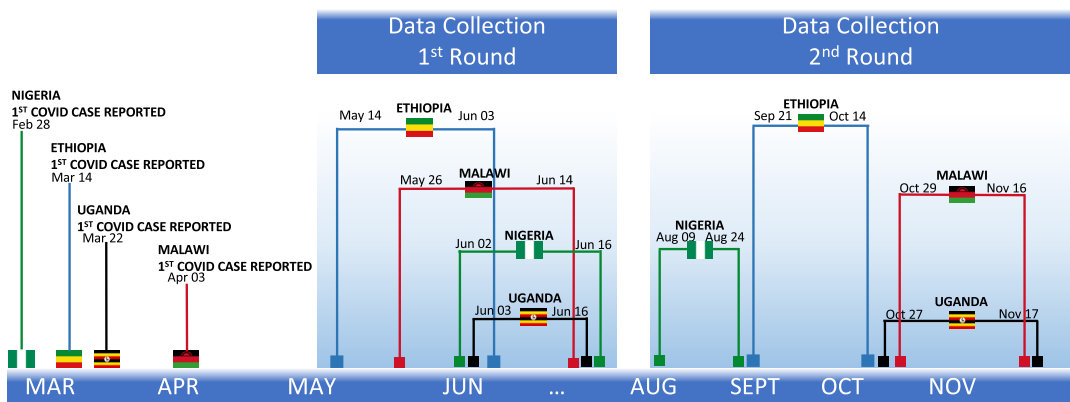
June 2020), and (iii) the existence of a survey follow-up that contained food security information and was close in timing with other countries surveyed—to maximize the sample size. Figure 1 illustrates how Ethiopia, Malawi, Nigeria, and Uganda fulfilled the criteria mentioned above, with a first round of data collection performed during May–June of 2020 and a follow-up round performed during August–November of 2020.<sup>4</sup>

To facilitate the utilization of data derived from the high-frequency phone surveys and enable cross-country comparisons, the World Bank LSMS team harmonized the variables obtained from these surveys. This harmonization process involved adhering to standardized definitions and ensuring consistent variable names. The variables encompassed various aspects such as demography, housing, household consumption expenditure, agriculture, and food security (World Bank, 2021).

To measure the food security status of households, the phone surveys used the FIES. The FIES is a metric used to assess the severity of food insecurity at the household or individual level. It relies on direct yes/no responses to eight questions regarding access to adequate food. The FIES serves as a scale encompassing a broad range of social, psychological, and health-related conditions, much like other established instruments used to measure unobservable traits. In Table 1, we provide the English version of the Food Insecurity Experience Scale Survey Module (FAO, 2017).<sup>5</sup>

The collective analysis of the eight questions of the FIES produces a quantitative tool to assess the prevalence of food insecurity. Specifically, the FIES methodology yields two indicators: (i) the prevalence of severe food insecurity and (ii) the prevalence of moderate or severe food insecurity (combining moderate and severe levels). Individuals experiencing moderate food insecurity often consume low-quality diets and may reduce the quantity of food they typically consume at certain times during the year. On the other hand, those experiencing severe food insecurity endure days without eating due to a lack of financial means or resources to acquire food (FAO, 2017).

In our analysis, households are deemed food insecure if their probability of being severely and/or moderately/severely food insecure is equal or greater than 50% (World Bank, 2021).<sup>6</sup> This criterion delineates four possible scenarios for households across two periods: (i) remaining food secure in both periods, (ii) transitioning from food security in the first period



**FIGURE 1** Timeline of events in Ethiopia, Malawi, Nigeria, and Uganda during 2020. The date of the official first case of COVID-19 reported in each country was sourced from Roberts et al. (2021). Data collection dates were obtained from descriptions provided by the Microdata Library of the World Bank.

**TABLE 1** Food insecurity experience scale survey module.

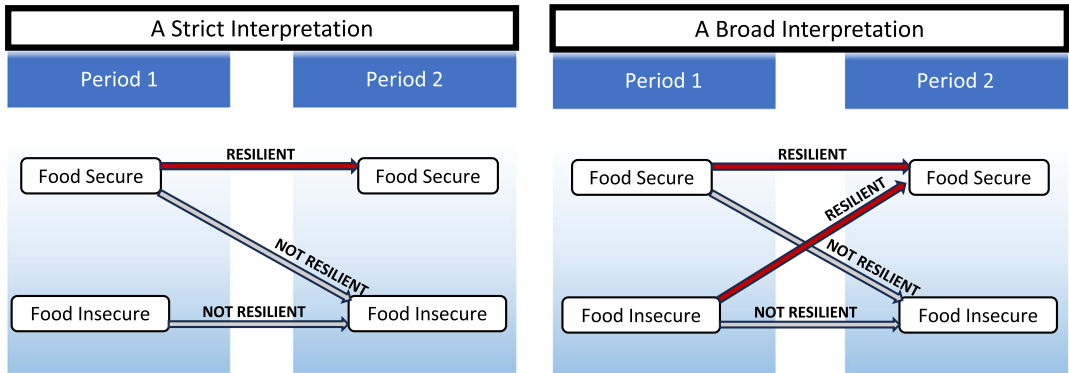
Question	Standard label	Question wording
1	WORRIED	During the last 30 DAYS, was there a time when You were worried you would not have enough food to eat because of a lack of money or other resources?
2	HEALTHY	Still thinking about the last 30 DAYS, was there a time when you were unable to eat healthy and nutritious food because of a lack of money or other resources?
3	FEWFOODS	Was there a time when you ate only a few kinds of foods because of a lack of money or other resources?
4	SKIPPED	Was there a time when you had to skip a meal because there was not enough money or other resources to get food?
5	ATELESS	Still thinking about the last 30 DAYS, was there a time when you ate less than you thought you should because of a lack of money or other resources?
6	RANOUT	Was there a time when your household ran out of food because of a lack of money or other resources?
7	HUNGRY	Was there a time when you were hungry but did not eat because there was not enough money or other resources for food?
8	WHOLEDAY	During the last 30 DAYS, was there a time when you went without eating for a whole day because of a lack of money or other resources?

**TABLE 2** Distribution of households according to their food security dynamics.

	(1) Continuous food security	(2) Recovery from food insecurity to food security	(3) Persistence in food insecurity	(4) Transition from food security to food insecurity	(5) Total
Ethiopia	1522	328	383	174	2407
Malawi	345	262	773	140	1520
Nigeria	304	188	1079	158	1729
Uganda	988	582	258	67	1895
Total	3159	1360	2493	539	7551

to food insecurity in the second period, (iii) remaining food insecure in both periods, and (iv) moving from food insecurity in the first period to food security in the second period. In Table 2, we show the distribution of households in our sample according to the four possible scenarios of food security dynamics described above.

We construct a binary indicator to represent resilience, as informed by the FIES indicators and the normative framework on resilience previously discussed. Non-resilience is straightforwardly indicated by persistence in or transitions to food insecurity (scenarios 2 and 3 described above, columns (3) and (4) of Table 2). However, the definition of resilience warrants discussion. A *strict interpretation* would identify resilience solely with continuous food security (scenario 1 described above, column (1) of Table 2), whereas a *broad interpretation* would encompass both consistent food security and recovery from food insecurity to food security



**FIGURE 2** Different interpretations of resilience. A *strict interpretation* of resilience identifies resilience solely with continuous food security. A *broad interpretation* of resilience identifies resilience encompassing both consistent food security and recovery from food insecurity to food security.

(scenarios 1 and 4 described above, columns (1) and (2) of Table 2). A graphical representation of these interpretations is depicted in Figure 2. Our machine learning models will analyze resilience under both interpretations to provide a more comprehensive and holistic understanding of resilience prediction.<sup>7</sup>

Regarding the potential predictors of resilience, we utilize a comprehensive set of variables whose details and summary statistics can be found in Table 3. The selection of variables was broad and inclusive, guided by data availability—from the phone surveys conducted by the World Bank LSMS team—and relevance to the research question, with the aim of maximizing predictive accuracy. They encompass a wide range of domains, including socioeconomics, demographics, income sources, assets, agricultural and livestock activities, and labor. Thus, providing a comprehensive view of the factors that may influence resilience.<sup>8</sup>

For the interested reader, we present in Figures S1 and S2, of the Supporting Information, a visual examination of the distinctions between resilient and non-resilient households under both interpretations discussed previously. The figures showcase bar graphs representing the deviations of standardized predictors from the overall mean. They also show the groupwise 95% confidence intervals (resilient vs. non-resilient). These visualizations offer a comparison of the predictor variables, enabling readers to grasp the differences between the two groups.

## EMPIRICAL FRAMEWORK

This section describes the machine learning approach and algorithms. We assume the binary indicator of resilience ( $y$ ) as some function ( $f$ ) of household socioeconomic and demographic variables ( $X$ ), agricultural and labor variables ( $Z$ ), and country-level control variables ( $c$ ). The equation can be written as:

$$y = f(X, Z, c), \quad (1)$$

where  $y$  represents the binary variable indicating household resilience (=1 if the household is resilient and 0 otherwise) following the two interpretations described above. A data-driven

**TABLE 3** Summary statistics—Sample based on a broad interpretation of resilience.

Variables	Ethiopia		Malawi		Nigeria		Uganda		All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Asset index	−0.16	1.61	−0.03	1.62	0.88	1.62	−0.58	1.55	0	1.68
Total land size owned (hectares)	0.31	1.67	0.15	0.42	0.69	1.27	0.82	1.26	0.49	1.33
Ownership of dwelling (Yes = 1)	0.53	0.5	0.63	0.48	0.61	0.49	0.84	0.36	0.65	0.48
Access to improved water source (Yes = 1)	0.16	0.37	0.97	0.16	0.18	0.38	0.59	0.49	0.44	0.5
Access to improved toilet (Yes = 1)	0.59	0.49	0.52	0.5	0.68	0.47	0.32	0.47	0.53	0.5
Account from financial institutions (Yes = 1)	0.78	0.41	0.41	0.49	0.63	0.48	0.59	0.49	0.62	0.48
Change in number of males aged 15–64	0.05	0.48	0.01	0.38	0.04	0.31	−0.01	0.27	0.02	0.38
Change in number of females aged 15–64	0.05	0.47	0.01	0.4	0.04	0.35	0	0.3	0.03	0.39
Change in overall HH size	0.06	0.59	0.15	0.88	0.17	0.77	0.01	0.63	0.09	0.71
Ownership of any ruminant (large or small) (Yes = 1)	0.3	0.46	0.19	0.39	0.34	0.47	0.49	0.5	0.33	0.47
Ownership of camelid (Yes = 1)	0.01	0.12	0	0	0	0.03	0	0	0.01	0.07
Ownership of equine (Yes = 1)	0.12	0.33	0	0.04	0.01	0.08	0	0.02	0.04	0.2
Ownership of poultry (Yes = 1)	0.17	0.37	0.36	0.48	0.27	0.44	0.42	0.49	0.29	0.45
Ownership of livestock (Yes = 1)	0.34	0.48	0.44	0.5	0.44	0.5	0.64	0.48	0.46	0.5
Cash crop cultivation (Yes = 1)	0.07	0.26	0.08	0.27	0.17	0.38	0.24	0.43	0.14	0.35
Number of crops cultivated	1.57	3.76	2.81	3.14	2.87	3.2	2.86	2.64	2.44	3.31
Sale of crop (Yes = 1)	0.13	0.33	0.35	0.48	0.4	0.49	0.53	0.5	0.34	0.47
Postharvest crop loss (Yes = 1)	0.03	0.16	0.14	0.35	0.03	0.17	0.01	0.09	0.05	0.21
Use of tractor (Yes = 1)	0.04	0.2	0	0.03	0.09	0.28	0.8	0.4	0.23	0.42

(Continues)

TABLE 3 (Continued)

Variables	Ethiopia		Malawi		Nigeria		Uganda		All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Use of any fertilizer (organic or inorganic) (Yes = 1)	0.17	0.37	0.57	0.5	0.34	0.47	0.12	0.32	0.27	0.45
Use of pesticides, fungicides or herbicides (Yes = 1)	0.07	0.26	0.08	0.27	0.33	0.47	0.14	0.34	0.15	0.36
Use of exchange and/or free labor (Yes = 1)	0.13	0.33	0.02	0.15	0.27	0.45	0	0.03	0.11	0.31
Use of hired labor (Yes = 1)	0.12	0.33	0.02	0.15	0.56	0.5	0.31	0.46	0.25	0.43
Working adults working in wage work (%)	23.6	34.7	18.5	28.8	15.2	27.4	20.5	31.02	19.9	31.2
Working adults working in nonfarm family enterprise (%)	13.2	27.6	26.2	33.7	30.2	34.4	16.9	27.82	20.6	31.4
Ownership of non-farm family enterprise (Yes = 1)	0.3	0.46	0.5	0.5	0.61	0.49	0.02	0.14	0.34	0.47
Rental income (Yes = 1)	0.09	0.29	0.09	0.29	0.06	0.23	0.13	0.34	0.09	0.29
Received remittance or assistance (Yes = 1)	0.21	0.41	0.54	0.5	0.38	0.49	0.4	0.49	0.36	0.48
Resilience— <i>Broad interpretation</i> (Yes = 1)	0.77	0.42	0.4	0.49	0.28	0.45	0.83	0.38	0.6	0.49
Observations	2407		1520		1729		1895		7551	

Note: These summary statistics describe each of the key variables in our analysis and describe the composition of our sample based on our *broad interpretation* of resilience. For summary statistics of the sample based on our *strict interpretation* of resilience, see Table S1. The *Asset Index* is a comprehensive index representing the assets of the household. The *total land size owned* is limited to agricultural land. *Rental income* from shop, store, house, car, truck, other vehicles, land, agricultural tools, and transport of animals. *Working adults working in wage work* includes casual and permanent work.

approach is more appropriate for our case since the exact functional relationship between resilience and its predictors is unknown.

Our motivation for using a data-driven approach stems from the exploratory nature of our analysis. We aim to discern patterns from the observed data, in contrast to a confirmatory analysis where one tests hypotheses derived from a structural or econometric model. The task at hand is twofold. First is feature extraction: we seek to identify the characteristics of households that were resilient during the pandemic compared to those that were not. The second task is prediction: we aim to evaluate how well the household features identified by our models can predict resilience in a blind out-of-sample test set.

For a task like this, machine learning models have been shown to have an advantage in extracting complex relationships in a data-driven manner and a high out-of-sample predictive capacity (Athey & Imbens, 2019; Bajari et al., 2015; Baylis et al., 2021; Mullainathan & Spiess, 2017; Storm et al., 2020). These data-driven models reduce the dependency on researchers' prior assumptions about the functional form. They are also more flexible since the parameters are optimally chosen via a grid search.

The learning aspect of machine learning involves the iterative process of tuning hyperparameters based on the prediction errors observed in repeated subsamples. The full dataset was randomly divided into training (80%) and testing (20%) samples. The training sample was utilized to train the model using tenfold cross-validation (CV). This process involved randomly partitioning the training sample into 10 equal-sized subsamples, with nine subsamples used for training and one subsample for validation. This process was repeated 10 times, each time updating the hyperparameters to minimize prediction errors in the validation subsample. Each CV cycle involves training temporary models on nine subsamples and validating on one subsample, facilitating the identification and selection of the best hyperparameters based on the aggregated results from all 10 validations. Once the optimal hyperparameters were determined through this CV process, a final model was then trained using the entire training dataset (80% of the full data) with these selected hyperparameters. Finally, the trained model was evaluated using the held-out testing sample, which was not used during the training phase. This approach was implemented to ensure the robustness of the model (Vabalas et al., 2019; Zhang & Ling, 2018).

We train five popular models for this analysis, including Logistic Regression, Random Forest (RF), Gradient Boosting Classifier (GBC), eXtreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANNs). These models were selected for their superior predictive capabilities compared to other models such as classification trees or support vector machines (Amin et al., 2021; Athey & Imbens, 2019; Bajari et al., 2015; Dreiseitl & Ohno-Machado, 2002; Villacis et al., 2023). Next, we will provide an overview of each model.

## Logistic Regression

The logit or Logistic Regression model is a widely used statistical method for modeling binary outcomes. It belongs to the class of generalized linear models (GLMs) and is particularly suitable for situations where the response variable takes on one of two categorical values. In the logit model, the probability of the binary outcome is modeled as a function of the predictor variables using the logistic function. The logit model provides estimates of the regression coefficients, representing the change in the log odds of the outcome associated with a one-unit increase in the corresponding predictor, holding other predictors constant. These coefficients can be exponentiated to obtain odds ratios, indicating the multiplicative effect of the predictor on the odds of success.

## Random Forest

Random Forest (RF) is an ensemble machine learning algorithm that combines the predictions of multiple decision trees to improve the accuracy of predictions (Breiman, 2001). The algorithm has become widely used in classification and regression tasks due to its robustness and

ability to handle high-dimensional data (Chernozhukov et al., 2017; Wager & Athey, 2018). The Random Forest algorithm operates through a series of steps to create a collection of decision trees. First, it randomly selects subsets of the training data through bootstrap sampling, which involves randomly selecting samples with replacements. This creates a new training set with some repeated samples and some not included samples.

Next, the algorithm randomly selects a subset of features (predictors) from the full set of features. This randomness in feature selection introduces diversity among the trees and reduces correlation, leading to better overall performance. Each decision tree is built using the selected subset of data and features. At each node of the tree, the algorithm searches for the best-split point among the selected features, considering a specific criterion such as the Gini index<sup>9</sup> for classification or variance for regression. This recursive splitting process continues until a stopping condition is met, which can be defined by a minimum node size or a maximum depth. Multiple decision trees are created through the ensemble creation step. The hyperparameter controls the number of trees in the forest.

For classification problems like the model shown above, the prediction is determined by the majority vote or mode of the predictions from all the trees. One of the key advantages of the Random Forest algorithm is its ability to handle high-dimensional data and missing values. It also provides estimates of feature importance, allowing us to assess the relative importance of different predictors in the model. The Random Forest algorithm is less prone to overfitting compared to individual decision trees. The process of randomization in feature selection and data sampling helps reduce variance and provides more robust predictions (Athey & Imbens, 2019).

## Gradient Boosting and eXtreme Gradient Boosting

GBC and XGBoost are popular machine learning models that use the principle of boosting to improve prediction accuracy by combining multiple simple models, typically decision trees. The main idea behind these methods is to build models sequentially, with each new model focusing on correcting the errors made by the ones before it. This process begins with an initial guess, which could be the average of the target values for a regression problem or a log-odds ratio for classification, setting the stage for further refinement. As the sequence progresses, each new model (e.g., a decision tree) is fitted on pseudo-residuals, using the gradients of the loss function with respect to the model predictions (Hastie et al., 2009). The contribution of each tree to the final model is controlled by a learning rate, a parameter that determines how quickly the model approaches the optimum structure. Regularization techniques such as shrinkage (reducing the step size toward the ultimate model), limiting tree complexity, and random subsampling are used in GBC to prevent overfitting and improve model generalization.

XGBoost builds on the foundation of gradient boosting by introducing several optimizations aimed at improving the efficiency, speed, and scalability of the model through a series of targeted optimizations (Chen & Guestrin, 2016). It introduces regularization parameters, for example, one for tree pruning and another for step size shrinkage, to reduce overfitting by penalizing complexity, thereby improving the model's generalization capability. A key feature of XGBoost is its sparsity-aware algorithm, designed to handle missing data and zero-valued entries efficiently, learning the optimal branching direction for missing values to boost model accuracy. Additionally, XGBoost incorporates advanced options such as monotonic constraints and feature interaction constraints, allowing for tailored model adjustments to suit specific domain requirements. To expedite the identification of optimal split points in trees, it employs

a compressed column-based data structure and a quantile sketch algorithm, which together streamlines the process of finding split candidates, significantly curtailing training duration. The integration of parallel processing further accelerates tree construction, while a suite of features for model evaluation and optimization, including built-in cross-validation and early stopping, enhances the overall effectiveness and precision of the model (Chen & Guestrin, 2016).

The main differences between traditional GBC and XGBoost lie in XGBoost's advancements in the boosting framework. XGBoost goes further by improving regularization and computational efficiency, especially in processing sparse data, and designing an architecture that speeds up computations (Bentéjac et al., 2021). While GBC is relatively simpler, it might struggle with handling large datasets or complex models due to its demand for computational and memory resources; XGBoost is more flexible in these scenarios through its optimized tree construction, effective tree pruning, parallel processing, and superior handling of missing values.

## Artificial Neural Networks (ANNs)

ANNs operate by mimicking the structure and function of the human brain's neural networks, consisting of interconnected artificial neurons or "units" organized into layers. These layers include an input layer, one or more hidden layers, and an output layer. The heart of ANNs lies in their artificial neurons, which simulate the behavior of biological neurons. These neurons receive weighted inputs, which are the outputs from the previous layer's units, and each unit receives inputs from the previous layer, performs computations, and transmits outputs to the next layer. Activation functions, such as sigmoid, rectified linear activation unit (ReLU), or hyperbolic tangent, are applied to the weighted sum of inputs to introduce nonlinearity and transform it into an output. The activation function determines the neuron's response to the input and plays a crucial role in shaping the network's behavior (Jain et al., 1996).

This layered structure allows ANNs to learn hierarchical representations of input data and capture complex relationships. During training, ANNs adjust the weights to minimize a defined objective or loss function through a process that involves forward feeding, where inputs propagate through the network, and outputs are compared to the desired outcomes. The resulting prediction errors are used to iteratively update the weights, aiming to reduce the overall loss. Backpropagation, the primary training algorithm for ANNs, employs gradient descent to calculate the gradients of the loss function with respect to the weights, guiding the weight updates and enabling the network to learn from the training data.

Various techniques are employed to enhance the performance and generalization of ANNs. Regularization methods such as L1 and L2 regularization prevent overfitting by adding penalty terms to the loss function. Dropout, another regularization technique, randomly deactivates a fraction of units during training to improve network robustness, ensuring the network learns more general patterns that are not dependent on the presence of specific neurons.

## Feature extraction with Shapley Additive exPlanations values

Predictive algorithms assess the importance of predictors by observing the increase in prediction error when each predictor is permuted, with a higher error signifying greater importance. However, this measure of importance is model-specific and may not reliably account for predictor interdependencies (Lundberg & Lee, 2017). Shapley values (Shapley, 1953), stemming from

cooperative game theory, provide a nuanced explanation of each predictor's contribution, taking into account interactions with other predictors. These values, known as Shapley Additive exPlanations (SHAP), explain how each predictor influences the deviation of the actual prediction from the mean. Introduced by Lloyd Shapley in 1953 and adapted for machine learning, Shapley values are calculated as the mean marginal contribution of a predictor across all possible predictor combinations, offering a more holistic and interpretable assessment than traditional importance factors (Lundberg & Lee, 2017).

Formally, the Shapley value at an observation  $i$  for a predictor  $x_j$  from the set of all predictors  $K$  is denoted as:

$$\phi_{ij} = \sum_{S \subseteq K \setminus \{j\}} \frac{|S|!(|K| - |S| - 1)!}{|K|!} [\hat{y}_i(x_{S \cup \{j\}}) - \hat{y}_i(x_S)]. \quad (2)$$

These values are calculated by summing over all subsets  $S \subseteq K$  of the set of predictors that exclude the  $j$ -th predictor. The difference  $\hat{y}_i(x_{S \cup \{j\}}) - \hat{y}_i(x_S)$  represents the prediction gap caused by adding the  $j$ -th predictor to model. Each gap is respectively weighted by the number of permutations of predictors that can occur, given by the ratio of factorials in the formula. The Shapley value for a predictor is the aggregate of these weighted prediction gaps (Lundberg & Lee, 2017).

Thus, the dimension of the Shapley value matrix corresponds to the number of observations by the number of predictors, as Shapley values are computed for each predictor with respect to each observation. Therefore, the sum of Shapley values across all predictors for a given observation should equal the difference between the prediction for that specific observation and the average prediction over the dataset (or the model's expected value if a baseline has been defined). This ensures that the contribution of all predictors sums up to the actual prediction for each observation.

Shapley values show the relationship between predictors and the outcome by quantifying the change in the predicted value associated with each predictor. A positive Shapley value implies that the predictor's inclusion increases the predicted outcome relative to the average prediction. Conversely, a negative Shapley value denotes a decrease in the predicted outcome when the predictor is included. A zero Shapley value indicates no change from the overall average prediction. The magnitude of a Shapley value signifies the strength of a predictor's impact. Due to the computational intensity of considering all predictor combinations and orderings, we compute Shapley values exclusively for the model with the best out-of-sample performance.

## Model evaluation

Evaluating the performance of machine learning models is crucial in assessing their effectiveness. Several performance metrics are commonly used to measure the quality of classification models, including Accuracy, Precision, Recall, F1 Score, and Cohen's Kappa (Amin et al., 2021; Villacis et al., 2023).

Accuracy is a widely used metric that measures the overall correctness of predictions made by a classification model. Higher accuracy indicates a higher proportion of correct predictions, and it is calculated as the ratio of correctly classified instances to the total number of instances:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (3)$$

where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively.

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It quantifies the model's ability to avoid false positives. Thus, higher precision indicates a lower false positive rate. The measure of precision is important in applications where false positives are costly or undesirable, and it is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (4)$$

Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances. It quantifies the model's ability to identify positive instances. Thus, higher recall indicates a lower false negative rate. The measure of recall is important in applications where false negatives are costly or undesirable, and it is calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

F1 Score combines precision and recall into a single metric via harmonic mean. It provides a balanced measure between precision and recall, and it is particularly useful when classes are imbalanced. The F1 score is important in applications where both false positives and false negatives need to be minimized, and it is calculated as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (6)$$

Cohen's Kappa (Cohen, 1960) is a statistical measure that assesses the agreement between two raters (in this case, the model prediction and the true category) beyond chance agreement. It considers the observed agreement and the expected agreement due to chance, and it is calculated as follows:

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e},$$

where  $p_o$  and  $p_e$  represent observed and expected probabilities, respectively. Kappa values range from  $-1$  to  $1$ , where  $1$  indicates perfect agreement,  $0$  indicates agreement by chance, and negative values indicate agreement worse than chance (Amin et al., 2021; Warrens, 2015).

## Preprocessing

To ensure the accuracy and reliability of our machine learning models, we took two essential steps in preparing the predictor variables for training. First, we centered and scaled the

variables. This helps to remove any potential biases caused by varying scales of the variables and ensures that all variables contribute equally to the model. By doing this, we prevent any variable from dominating the prediction process simply because it has larger values. Secondly, we checked the correlation among the variables in use. High correlations between variables may affect model stability and interpretation and mislead importance factors. Figures S3 and S4, of the Supporting Information, shows the correlation among the variables we use. None of the variable pairs had a correlation close to one or negative one.

## Addressing class imbalance

When a variable exhibits imbalanced classes, it means that one class contains significantly more samples than the other. For instance, under both, *broader* and *stricter* interpretations of resilience, the average values are 0.60 and 0.51, respectively (see Tables 3 and S1). This implies that there are more resilient households than non-resilient ones. However, such an imbalance can pose challenges as machine learning models may become biased toward the majority class, leading to suboptimal performance for the minority class (Amin et al., 2021). To address this issue, we employ the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002).

The SMOTE function is used to balance the classes by creating synthetic observations of the minority class. This is achieved by selecting examples that are close to the feature space, drawing a line between the examples in the feature space, and drawing a new sample at a point along that line. In the context of our analysis, SMOTE is applied to the training data to balance the classes.<sup>10</sup>

## Hyperparameter tuning

The logit model does not require hyperparameter tuning. In Random Forest, the primary hyperparameters being tuned are the number of decision trees in the forest, the maximum depth of each decision tree, and the minimum number of samples required to split an internal node. To find the optimal combination of these hyperparameters, we define a grid of values for each parameter and systematically evaluate the model's performance across all possible combinations. Similarly, in GBC and XGBoost, the hyperparameters under consideration include the number of boosting rounds, the maximum depth of each individual tree, and the step size shrinkage (learning rate) for each boosting round. For ANNs, we focus on tuning the sizes of the hidden layers in the network, the activation function used in these layers, and the L2 regularization parameter. Overall, the process involves defining grids of parameter values for each model and systematically evaluating the models' performance across various combinations. This allows us to identify the set of hyperparameter values that maximizes the models' predictive capabilities.

## RESULTS

In this section, we present our research findings, showcasing the use of machine learning models for resilience prediction and subsequently identifying key predictors. First, in Table 4,

**TABLE 4** Performance of machine learning models.

Response	Model	Accuracy	Recall	Precision	F1	Kappa
Panel A: Validation performance						
Resilience status: <i>Broad interpretation</i>	GBC	0.744	0.78	0.791	0.79	0.469
	Random Forest	0.733	0.77	0.781	0.78	0.445
	XGBoost	0.723	0.781	0.762	0.77	0.419
	Logistic	0.692	0.705	0.762	0.73	0.37
	ANN	0.682	0.729	0.737	0.73	0.339
Resilience status: <i>Strict interpretation</i>	GBC	0.759	0.776	0.759	0.77	0.518
	Random Forest	0.748	0.745	0.758	0.75	0.496
	XGBoost	0.747	0.757	0.75	0.75	0.494
	ANN	0.721	0.739	0.721	0.73	0.441
	Logistic	0.714	0.735	0.714	0.72	0.427
Panel B: Test performance						
Resilience status: <i>Broad interpretation</i>	GBC	0.775	0.804	0.817	0.81	0.534
Resilience status: <i>Strict interpretation</i>	GBC	0.81	0.843	0.796	0.82	0.618

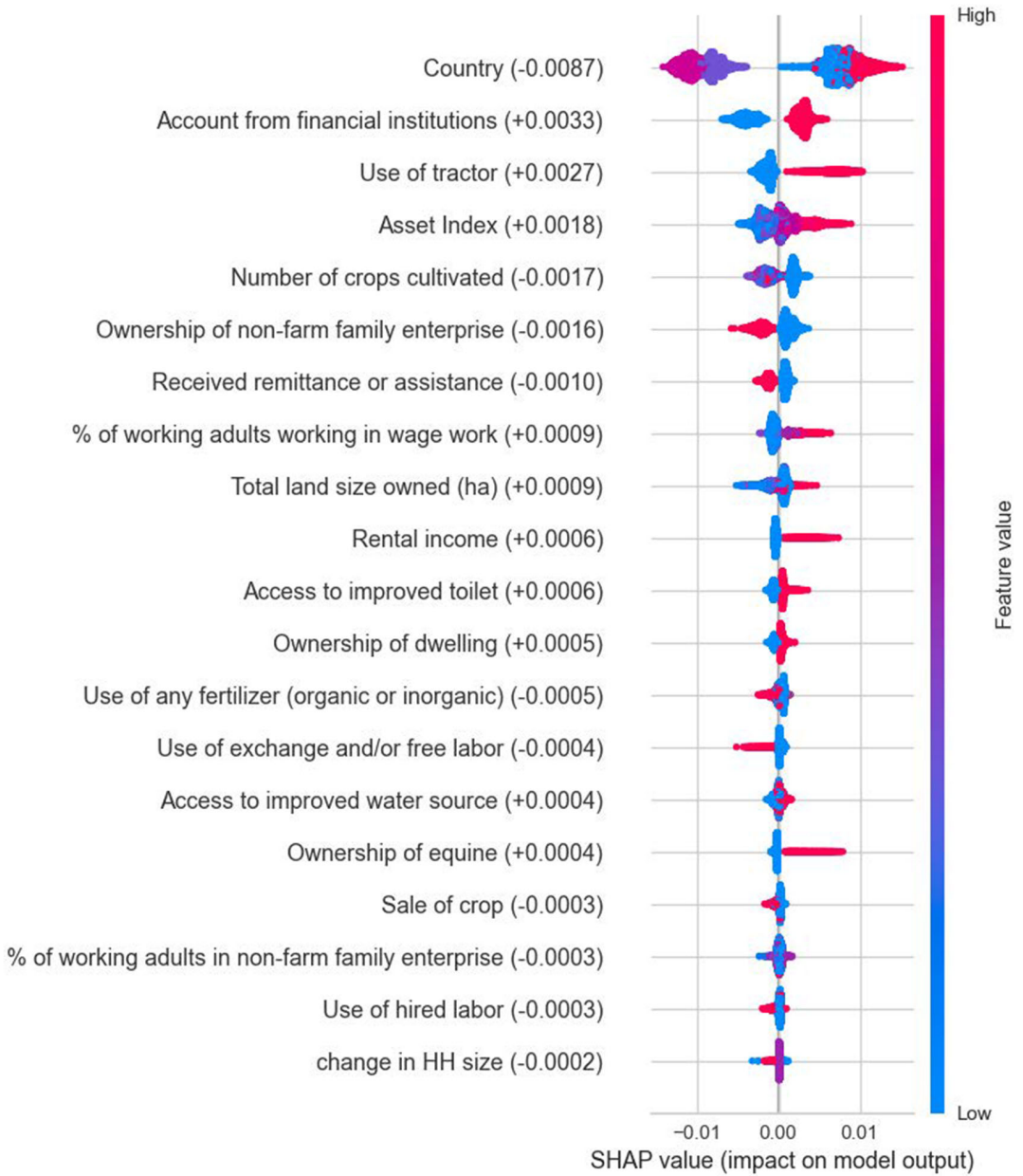
*Note:* Accuracy, precision, recall, F1 score, and Cohen’s Kappa are commonly used performance metrics in machine learning (the higher the better). Accuracy measures overall correctness; Precision focuses on true positives out of all positives predicted; Recall captures the true positive rate; F1 score combines Precision and Recall; Cohen’s Kappa assesses agreement beyond chance. The validation performance is from tenfold cross-validation, and the test performance is based on the 20% out-of-sample data from Ethiopia, Malawi, Nigeria, and Uganda. See Section 2 for more details. Table S5 shows the gains of ML models.

we present results of the performance statistics for the different machine learning models discussed in Section 2 and employed for resilience prediction. Second, Figures 3 and 4 provide the results of feature extraction and identify the key predictors of resilience.

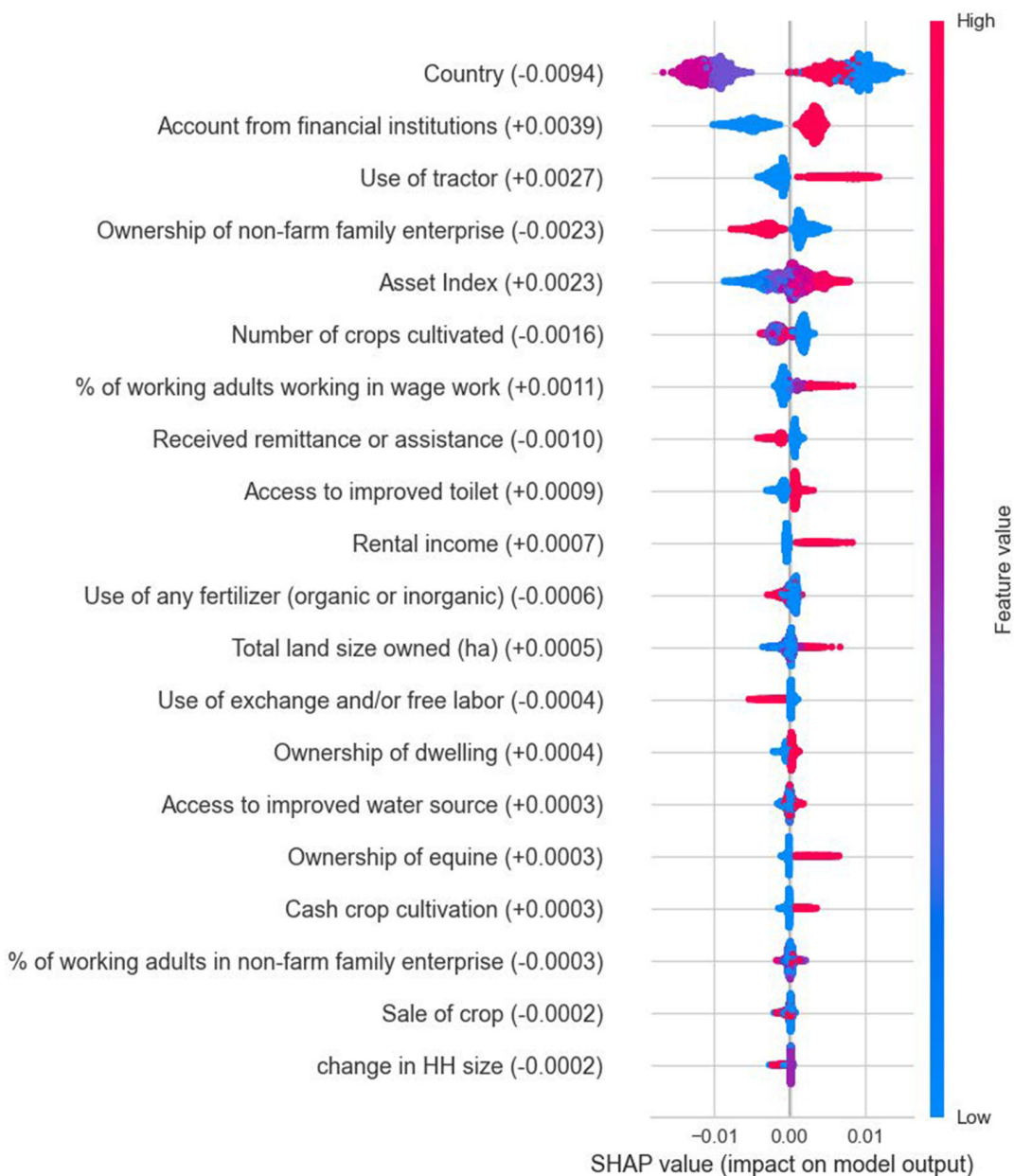
### Predictive performance

This subsection discusses the performance statistics for different machine learning models in predicting resilience among sampled African households. Recall that resilience is measured as a binary variable, where a value of 1 indicates a resilient household, and 0 denotes a lack of resilience. The predictor variables used in our analysis include household income, assets, demographics, agricultural and labor activities, and country-level factor variables. To evaluate the performance of the machine learning models in predicting household resilience, we present the corresponding performance statistics in Table 4. This includes measures of accuracy, precision, recall, F1-score, and Cohen’s Kappa for each model, respectively, for model validation and out-of-sample testing tasks.

Table 4 shows that the GBC stands out as the top-performing model in predicting household resilience, according to the results from a tenfold cross-validation. It exhibits the highest overall



**FIGURE 3** Top 20 predictors of household resilience under the broad definition. The figure contains scatterplots for each predictor, illustrating respective SHAP values. Each dot on the scatterplot represents a SHAP value corresponding to an observation, with the color denoting the feature or predictor's value: Red indicates high values, while blue denotes low values. The horizontal position of a dot reflects the impact of that value on the model's prediction, with positive (negative) values suggesting a higher (lower) likelihood of resilience. The top 20 predictors are presented here based on their mean absolute SHAP values, reported on the y-axis. These mean values are multiplied by the sign of correlation between the predictor and its SHAP values, indicating the direction of association between resilience and the predictor. For detailed results, please refer to Section 3.



**FIGURE 4** Top 20 predictors of household resilience under the strict definition. The figure contains scatterplots for each predictor, illustrating respective SHAP values. Each dot on the scatterplot represents a SHAP value corresponding to an observation, with the color denoting the feature or predictor's value: Red indicates high values, while blue denotes low values. The horizontal position of a dot reflects the impact of that value on the model's prediction, with positive (negative) values suggesting a higher (lower) likelihood of resilience. The top 20 predictors are presented here based on their mean absolute SHAP values, reported on the y-axis. These mean values are multiplied by the sign of correlation between the predictor and its SHAP values, indicating the direction of association between resilience and the predictor. For detailed results, please refer to Section 3.

accuracy, with scores of 0.744 and 0.759 for broad and strict definitions of resilience, respectively. This measure of accuracy indicates GBC's effectiveness in correctly predicting both resilient and non-resilient households over relatively complex models like XGBoost and ANN.

GBC excels in accuracy and demonstrates a well-balanced performance in terms of precision and recall—key metrics where precision is the ratio of true positives to all positive predictions, and recall, or true positive rate, quantifies how many actual positives were identified correctly. This balanced accuracy results in high F1 scores of 0.785 and 0.767 for both resilience interpretations, affirming that F1, as the harmonic mean of precision and recall, is a robust indicator of model reliability.

Cohen's Kappa, which evaluates the agreement between the observed and predicted classifications beyond chance, further solidifies GBC's superiority with scores of 0.469 and 0.518 under broad and strict interpretations of resilience. This suggests that the GBC model is performing significantly better than random chance. Therefore, we select GBC as the final model for the out-of-sample prediction of household resilience.

When testing the out-of-sample performance using the held-out 20% of the data, the GBC maintains its strong performance with scores of 0.775 and 0.81 for *broad* and *strict* resilience definitions, respectively. That is, about four out of five households are correctly predicted to be resilient or non-resilient in the test dataset. GBC also continues to exhibit consistent precision and recall across both interpretations. The model's accuracy over randomness, as indicated by the Kappa statistic, reaches 0.53 and 0.61, respectively, underscoring the model's robustness.<sup>11</sup>

It is important to note that our model predicts the *strict* interpretation of resilience slightly more accurately than the *broad* interpretation of resilience. This is because the *broad* interpretation includes households that transitioned from food insecure to food secure, which introduces some noise into the prediction process. Nevertheless, the similar performance metrics across definitions of resilience, as shown in Table 4, indicate that our model is robust across different interpretations of resilience.

Table S5 shows a comparison of different models based on their gains over a naive baseline (e.g., the mean) and Logistic Regression in predicting resilience. The results indicate that machine learning models generally offer considerable improvements over both a naive baseline (about 50% for GBC) and Logistic Regression (about 16% for GBC), especially in terms of predictive accuracy. While the ML methods deployed in this study provide robust predictive capabilities, it is imperative to consider their limitations in interpretability and the higher demands on computational resources and expertise. The choice of model should be guided not only by predictive accuracy but also by the specific needs and constraints of the application context.

For our task, however, the GBC provides the most balanced and highest performance across various metrics, namely accuracy in terms of identifying resilient households as resilient and non-resilient households as non-resilient. These attributes mark the GBC as the model of choice for the subsequent analysis phase, where we will explore and extract the features that characterize resilient households.

## Feature extraction

Figures 3 and 4 present SHAP summary plots derived from the GBC, illustrating the top 20 predictors of household resilience during the pandemic. Each dot corresponds to the SHAP value for a predictor for a particular observation, with its value shown on the x-axis. Interested readers can refer to Figure S5 to understand how these dots contribute to the overall prediction

through SHAP values. To prevent overlapping, some dots are jittered to demonstrate the concentration of data points.

The color of the dots indicates the value of the predictor, with red representing higher values and blue indicating lower ones. A pattern of red dots to the right of zero suggests that higher values of a predictor have higher SHAP values, that is, the predictor has a positive correlation with the SHAP values (+). Hence, its association with household resilience is positive. Conversely, blue dots to the right of zero would imply a negative (−) correlation between the predictor and its SHAP values. Hence, its association with household resilience is negative. The mean absolute SHAP values are multiplied by the sign of this correlation and placed in parentheses for improved readability. These values are relative and do not have much interpretability in their absolute sense.

The country-level control variable emerges as a significant determinant under both definitions of resilience. This is expected given its encapsulation of country-specific unobserved heterogeneities that can influence household food security outcomes. Indeed, country-level factors are substantial during a global event such as a pandemic, which affects regions in disparate ways based on local policy, healthcare infrastructure, and economic resilience.

We are more interested in household-level predictors. These predictors are similar under both interpretations of resilience (see Figures 3 and 4). For instance, the possession of an account with a financial institution presents the strongest positive association with resilience (+0.0033 and +0.0039, respectively, under the *broad* and *strict* interpretations of resilience). This indicates that households with access to financial institutions were also resilient in the data (Belayeth Hussain et al., 2019; Islam et al., 2016; Sakyi-Nyarko et al., 2022). Similarly, the adoption of agricultural mechanization, as evidenced by the use of tractors, aligns positively with resilience, indicating that the use of a tractor is one of the most identifiable features of resilient households (Amare & Endalew, 2016; Daum & Birner, 2020; Emami et al., 2018).

Following this, the asset index emerges as a strong predictor of household resilience, likely reflecting the capacity to withstand shocks, as discussed in the existing literature (Ansah et al., 2019; Hidrobo et al., 2018; Manlosa et al., 2019). Other important features of resilient households include wage work, access to improved toilets, rental income, land size, ownership of equines, and cash crop cultivation, in that order. These predictors collectively reflect the economic resources, stable income, and ownership characteristics of households, which can be useful in sustaining their food consumption patterns during the pandemic.

On the other hand, higher values of some predictors are associated with lower SHAP values, that is, the association between the predictor and the response variable is negative, as indicated by negative signs in parentheses (Figures 3 and 4). For example, ownership of nonfarm family enterprises, receipt of remittances or assistance, and the use of exchange or free labor have many negative SHAP values with their corresponding dots in red, suggesting that higher values of these predictors are negatively associated with household resilience. Upon closer inspection of these predictors, the operation of a nonfarm family enterprise similarly exhibits a negative association, potentially reflecting the broader economic downturn's impact on such businesses. The practice of cultivating a larger variety of crops is also slightly negatively correlated with resilience. Although speculative, this suggests that crop diversification at the household level, typically a risk mitigation strategy, may have presented challenges during the pandemic due to market disruptions. It is also intuitive that households receiving remittances or assistance were less likely to remain resilient during the pandemic, as these sources of support may have been insufficient or disrupted. Similarly, households utilizing exchange or free labor demonstrate a

low predicted chance of resilience, perhaps because opportunities like these may have been adversely affected by the pandemic.

Among other factors, the use of fertilizer and ownership of a dwelling exhibit negative and positive associations, respectively, with household resilience. However, their respective SHAP values are heavily concentrated around zero, indicating that they may not exert a strong influence on the prediction task.

The feature extraction analysis highlights several key predictors that play a crucial role in predicting resilience within our study region. Notably, factors related to financial inclusion, agricultural mechanization (e.g., utilizing tractors), stable income, and economic resources emerge as significant positive predictors of resilience. Conversely, reliance on external assistance, engagement in nonfarm enterprises, and crop diversification are identified as negative predictors.

## Dimension reduction

Note that the feature selection with SHAP values is based on the observed variables only, under the assumption of appropriate preprocessing, no indication of causal inference, and no high collinearity among predictors (Fryer et al., 2021; Kumar et al., 2020; Lundberg & Lee, 2017; Marcílio & Eler, 2020). The SHAP values from our analysis highlight the relative importance of each predictor in detecting household resilience. As indicated by Figures 3 and 4, country-level characteristics emerge as the most influential factor, with the importance of each subsequent predictor diminishing. The diminishing values of importance allow us to refine our predictor set. For instance, as shown in Figure 3 (*broad* interpretation of resilience), the mean absolute SHAP value declines from 0.0087 for the country variable to 0.0033, and then decreases steadily until “Total land size owned (ha).” Beyond this point, such as with the addition of “rental income,” there is no substantial gain in prediction accuracy, indicating an optimal stopping point for the inclusion of predictors.

The out-of-sample results of this model, with the number of predictors reduced from 29 to only 9, are presented in Table 5. The accuracy stands at 0.772 compared to the previously derived 0.775 (Table 4). Although recall declines slightly, precision increases, keeping the F1 score and Kappa statistic approximately the same. Thus, we can predict the *broad* interpretation of household resilience using only nine predictors while still achieving about 77.5% out-of-sample prediction accuracy.

**TABLE 5** Performance of machine learning models with reduced dimension.

Response	Model	Accuracy	Recall	Precision	F1	Kappa
Test performance						
Resilience status: <i>Broad interpretation</i>	GBC	0.772	0.783	0.826	0.804	0.531
Resilience status: <i>Strict interpretation</i>	GBC	0.79	0.815	0.783	0.798	0.579

*Note:* Accuracy, precision, recall, F1 score, and Cohen’s Kappa are commonly used performance metrics in machine learning (the higher the better). Accuracy measures overall correctness; Precision focuses on true positives out of all positives predicted; Recall captures the true positive rate; F1 score combines Precision and Recall; Cohen’s Kappa assesses agreement beyond chance. The validation performance is from tenfold cross-validation, and the test performance is based on the 20% out-of-sample data from Ethiopia, Malawi, Nigeria, and Uganda. See Section 2 for more details.

A similar exercise for Figure 4 (*strict* interpretation of resilience) shows that the mean absolute SHAP values rapidly decrease up to the variable “access to improved toilet,” then decline more gradually. Overall, Figures 3 and 4 indicate that the main difference between the top predictors of the *broad* and *strict* interpretations of resilience is the ninth predictor: “access to improved toilet” variable is relevant for the *strict* interpretation, while “land ownership” is more relevant for the *broad* interpretation. In Table 5, we also present the out-of-sample performance of the GBC using these nine predictors for the *strict* interpretation of resilience. Although the statistics generally decrease from the full model (as seen in Table 4), the decline is minimal. Accuracy decreases from 81% to 79%. Recall and precision also decrease, affecting the F1 score and lowering the Kappa statistics from 0.618 to 0.579. Nonetheless, the results indicate that these nine variables can predict the strict interpretation of resilience with high accuracy, nearly four out of five times.

In summary, the nine predictors—country-level control, account from financial institutions, use of tractors, asset index, number of crops cultivated, ownership of a nonfarm family enterprise, recipient status of remittance or assistance, percentage of working adults in wage work, and access to an improved toilet—are major indicators of strict resilience. Note that land ownership, which was relevant in predicting broad resilience, is not as relevant in predicting *strict* resilience. A possible explanation is that land ownership may have characterized households transitioning from food insecurity to food security, thus explaining their resilience. Still, it is less telling of the resilience among households who remained food secure throughout.

## DISCUSSION AND POLICY IMPLICATIONS

In this study, we propose a machine learning framework to predict resilience—and therefore well-being dynamics—in regions affected by detrimental shocks stemming from the recent global pandemic. By integrating relevant concepts from the food security literature, our approach contributes to the comprehension of resilience, a concept that has gained significant importance in the field of development economics and has become a focal point for international development and humanitarian agencies in the past decade.

The paper starts with a discussion of how we frame resilience in the face of adverse shocks using a normative approach and drawing upon methodologies for assessing food security proposed by the Food and Agricultural Organization (FAO). By anchoring our definition of resilience as a normative outcome and indexing it to the Sustainable Development Goals (SDGs),<sup>12</sup> we ensure that our approach remains as a pro-poor concept (Barrett et al., 2021; Barrett & Conostas, 2014). Subsequently, our application tests a battery of different machine learning algorithms and explores their capabilities in predicting household resilience status. We find that the machine learning models are able to identify eight out of 10 resilient households.

Our dimension reduction analysis has identified eight major household features—given the country characteristics—that can help distinguish households likely to be resilient during a disease-induced pandemic from those that may not be. In this context, recognizing and emphasizing the significance of these predictors can substantially aid in preemptively identifying households that could be vulnerable during a crisis. Moreover, these factors are critical for households across different countries. While there will inevitably be country-level differences, these factors warrant increased focus at the household level to better cushion future shocks. Specifically, for smallholder farmers from the selected group of African nations in our study, having access to financial institutions emerges as the most crucial predictor of resilience,

followed by the adoption of agricultural mechanization, as evidenced by the use of tractors. Complementing these predictors are risk and income diversification strategies, as evidenced by the number of crops cultivated, the ownership of nonfarm enterprises, and assets ownership.

These results from our feature extraction exercise and identification of key predictors of resilience are in line with a growing literature linking resilience with financial inclusion (Arouri et al., 2015; Belayeth Hussain et al., 2019; Islam et al., 2016; Islam & Maitra, 2012; Jordan, 2015; Khandker et al., 2012; Sakyi-Nyarko et al., 2022), resilience with agricultural mechanization (Amare & Endalew, 2016; Daum & Birner, 2020; Emami et al., 2018), and resilience with household assets (Ansah et al., 2019; Gilligan & Hoddinott, 2007; Guo, 2011; Hidrobo et al., 2018; Little & Ahmad, 2002; Manlosa et al., 2019). However, it is important to note that the characterization of resilient and non-resilient households presented in this study must be interpreted with caution. While our models identify features helpful for prediction, these characteristics should be interpreted in the context of the limitations of our data and methodology.

Leveraging multi-country data, our study extends beyond the scope of prior single-country analyses, enhancing the external validity of our findings. This approach is pivotal in substantiating the role of household assets and access to financial institutions as critical factors in determining household resilience, corroborating existing literature. However, our results also underline the importance of considering country-specific nuances in resilience studies.

The correlations observed through our predictive modeling exercises should not be misconstrued as causality, a distinction of paramount importance when considering the implications of our findings for policy formulation. As such, we aim to provide valuable insights to policymakers that, while informed by our analysis, acknowledge the limitations of our study: (i) regarding causality and (ii) recognizing the methodological constraints inherent in predictive modeling. Based on this context, the policy implications discussed next should be viewed as exploratory, guiding future empirical inquiries rather than prescribing definitive actions.

The empirical associations we identify call for targeted pilot initiatives, focusing on key household-level predictors such as financial access, asset ownership, diversification in risk and income, and the uptake of agricultural mechanization. Rigorous evaluation of these interventions, including randomized control trials (RCTs), will be essential for elucidating their effects on resilience, potentially paving the way to establishing causal links. This necessitates policy measures that foster multidisciplinary collaborations combining quantitative and qualitative assessment methods to understand the mechanisms at play deeply. In addition, policymakers can embrace adaptive frameworks that can evolve based on emerging evidence. This approach would allow for the refinement of strategies in light of ongoing research findings, including those that may establish causality in the future.

Emphasizing continued investment in data collection and analytical capabilities that leverage machine learning and other advanced statistical methods is another important avenue policymakers should prioritize (Villacis & Badruddoza, 2023). This can improve the understanding of resilience dynamics over time, informing more nuanced policy interventions that can be adjusted as more is learned about the causal relationships between household-level factors and resilience.

The significance of country-level control variables in our results highlights the importance of considering country-specific factors when designing and implementing policy interventions. While our findings do not directly measure the effectiveness of specific policies, they suggest that policymakers should account for the unique socioeconomic and political contexts of each country to enhance the relevance and potential impact of their interventions. While specific

household-level predictors are highlighted in our study, policymakers must consider resilience as a multifaceted concept that requires a holistic approach. This means designing policies that not only address the direct predictors identified here but also consider broader social, economic, and environmental systems. Finally, engaging stakeholders and communities from the beginning of any intervention based on our findings is critical. This approach can help to uncover additional insights, foster local ownership of initiatives, and improve the effectiveness of resilience-building efforts.

It is also important to acknowledge that in our specific setting, we implicitly focus on assessing *short-term* resilience given our data points span over a period of approximately 7 months. To overcome this limitation, future research could employ machine learning methods to investigate *long-term* dynamics by utilizing big data covering extended periods of time. Such an exercise would align with Conostas et al. (2014) definition of resilience as “the capacity that ensures adverse stressors and shocks do not have long-lasting adverse development consequences.” However, the availability of data may pose constraints on the feasibility of such endeavors. Lastly, the utilization of a normative approach to resilience in our study—anchored in the context of food insecurity— may have inherent limitations. This normative approach could potentially constrain the measurement of resilience and conflate different phenomena (Barrett et al., 2021). Therefore, it is crucial to consider this caveat for the broader application of our results, as their applicability will depend on the specific objectives and intended goals of policy interventions.

## ACKNOWLEDGMENTS

The senior authorship is shared between Villacis and Badruddoza. The authors thank Chris Barrett, two anonymous reviewers, and Gopinath Munisamy, managing editor at *Applied Economic Perspectives and Policy*, for constructive comments on a previous draft of this manuscript. We are also grateful to participants at the 2024 Southern Agricultural Economics Association (SAEA) and the 2024 Agricultural and Applied Economics Association (AAEA) Annual Meetings for constructive feedback that helped us improve this paper. The views expressed here are those of the authors and do not necessarily reflect those of donors or the authors' institutions. All errors are our own and the usual disclaimers apply.

## ENDNOTES

- <sup>1</sup> Knippenberg et al. (2019) used LASSO and Random Forest algorithms. Garbero and Letta (2022) employed Classification Trees, Bootstrap Aggregating (bagging), Random Forests, k-nearest Neighbor, and Support Vector Machine algorithms. A number of studies show that gradient boosting and Neural Network models are more robust (Amin et al., 2021; Bajari et al., 2015; Jain et al., 1996; Mullainathan & Spiess, 2017).
- <sup>2</sup> Other examples of well-being outcomes include expenditures, consumption, income, assets, poverty, food security indicators (including, but not limited to, dietary diversity indices, coping strategies indices), health indicators (child health, anthropometry, morbidity, mortality), happiness and life satisfaction, equality, marginalization, safety and security, experiences of conflict or violence (Barrett et al., 2021).
- <sup>3</sup> See the [Supporting Information](#) for more information on the high-frequency phone surveys (HFPS) conducted by the World Bank.
- <sup>4</sup> Additional phone survey rounds are available for these and other countries. For more information, see World Bank (2021).
- <sup>5</sup> In addition to considering compromised diet quality and reduced food quantity, the eight questions used to construct the FIES scale also capture psycho-social elements associated with anxiety or uncertainty regarding the ability to procure enough food, a facet that other measures do not (FAO, 2017).

- <sup>6</sup> Table S2 presents the distribution of food-insecure households per country across the two rounds of data collected based on the FIES indicators.
- <sup>7</sup> For our analysis of the *strict interpretation* of resilience, we specifically omit households that moved from food insecurity to food security in the latter period.
- <sup>8</sup> All but three of the predictor variables were collected only during the first round of surveys. The three predictor variables also collected in the second round of surveys are: (i) the number of males aged 15–64, (ii) the number of females aged 15–64, and (iii) overall household size. For our analysis, we combine the information of these variables from the two periods and code them as (i) change in the number of males aged 15–64, (ii) change in the number of females aged 15–64, and (iii) change in the overall household size. See summary statistics in Table 3.
- <sup>9</sup> The Gini Index is calculated for each group of observations as,  $\text{Gini Index} = 1 - (\text{probability of yes})^2 - (\text{probability of no})^2$ , hence gives a measure of how mixed the classes are in that group. A Gini Index of 0 means perfect purity (all instances belong to the same class), and an Index of 0.5 means maximum impurity (instances randomly assigned to classes) (Raileanu & Stoffel, 2004).
- <sup>10</sup> We also ran the models without class balancing and found the results to be similar, with a slightly greater asymmetry in precision and recall (please see Table S3). The high validation performance but low test performance of models indicate that balancing classes was an appropriate step for this data.
- <sup>11</sup> The out-of-sample performance metrics for the remaining (suboptimal) models are placed in Table A4.
- <sup>12</sup> In SDG 2, countries commit to “End hunger, achieve food security and improved nutrition and promote sustainable agriculture” by 2030 (FAO, 2017).

## REFERENCES

- Amare, Dagninet, and Wolelaw Endalew. 2016. “Agricultural Mechanization: Assessment of Mechanization Impact Experiences on the Rural Population and the Implications for Ethiopian Smallholders.” *Engineering and Applied Sciences* 1(2): 39–48.
- Amin, Modhurima Dey, Syed Badruddoza, and Jill J. McCluskey. 2021. “Predicting Access to Healthful Food Retailers with Machine Learning.” *Food Policy* 99: 101985.
- Ansah, Isaac Gershon, Cornelis Gardebroek Kodwo, and Rico Ihle. 2019. “Resilience and Household Food Security: A Review of Concepts, Methodological Approaches and Empirical Evidence.” *Food Security* 11(6): 1187–1203.
- Arouri, Mohamed, Cuong Nguyen, and Adel Ben Youssef. 2015. “Natural Disasters, Household Welfare, and Resilience: Evidence from Rural Vietnam.” *World Development* 70: 59–77.
- Athey, Susan, and Guido W. Imbens. 2019. “Machine Learning Methods that Economists Should Know About.” *Annual Review of Economics* 11: 685–725.
- Bajari, Patrick, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang. 2015. “Machine Learning Methods for Demand Estimation.” *American Economic Review* 105(5): 481–85.
- Balashankar, Ananth, Lakshminarayanan Subramanian, and Samuel P. Fraiberger. 2023. “Predicting Food Crises Using News Streams.” *Science Advances* 9(9): eabm3449.
- Barrett, Christopher B., and Mark A. Constas. 2014. “Toward a Theory of Resilience for International Development Applications.” *Proceedings of the National Academy of Sciences of the United States of America* 111(40): 14625–30.
- Barrett, Christopher B., Kate Ghezzi-Kopel, John Hoddinott, Nima Homami, Elizabeth Tennant, Joanna Upton, and Wu. Tong. 2021. “A Scoping Review of the Development Resilience Literature: Theory, Methods and Evidence.” *World Development* 146: 105612.
- Baylis, K., T. Heckeley, and H. Storm. 2021. “Chapter 83-Machine Learning in Agricultural Economics.” In *Handbook of Agricultural Economics*, Vol 5, edited by C. B. Barrett and D. R. Just, 4551–4612. Oxford, UK: Elsevier.
- Belayeth Hussain, A. H. M., Noraida Endut, Sumonkanti Das, Mohammed Thanvir Ahmed Chowdhury, Nadia Haque, Sumena Sultana, and Khandaker Jafor Ahmed. 2019. “Does Financial Inclusion Increase Financial Resilience? Evidence from Bangladesh.” *Development in Practice* 29(6): 798–807.

- Bentéjac, Candice, Candice Anna, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2021. "A Comparative Analysis of Gradient Boosting Algorithms." *Artificial Intelligence Review* 54: 1937–67.
- Blake, Paul, and Divyamshi Wadhwa. 2020. "Year in Review: The Impact of COVID-19 in 12 Charts." <https://blogs.worldbank.org/voices/2020-year-review-impact-covid-19-12-charts>.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5–32.
- Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–357.
- Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." 785–794.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. 2017. "Double/Debiased/Neyman Machine Learning of Treatment Effects." *American Economic Review* 107(5): 261–65.
- Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20(1): 37–46.
- Constas, Mark, Tim Frankenberger, and John Hoddinott. 2014. *Resilience Measurement Principles: Toward an Agenda for Measurement Design* 1. Resilience Measurement Technical Working Group, Technical Series: Food Security Information Network.
- Daum, Thomas, and Regina Birner. 2020. "Agricultural Mechanization in Africa: Myths, Realities and an Emerging Research Agenda." *Global Food Security* 26: 100393.
- Dreiseitl, Stephan, and Lucila Ohno-Machado. 2002. "Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review." *Journal of Biomedical Informatics* 35(5–6): 352–59.
- Emami, Mohammad, Morteza Almassi, Hossein Bakhoda, and Issa Kalantari. 2018. "Agricultural Mechanization, a Key to Food Security in Developing Countries: Strategy Formulating for Iran." *Agriculture & Food Security* 7: 1–12.
- FAO. 2017. "The Food Insecurity Experience Scale: Measuring Food Insecurity Through People's Experiences." <https://www.fao.org/3/i7835e/i7835e.pdf>
- Foini, Pietro, Michele Tizzoni, Giulia Martini, Daniela Paolotti, and Elisa Omodei. 2023. "On the Forecastability of Food Insecurity." *Scientific Reports* 13(1): 2793.
- Fryer, Daniel, Inga Strümke, and Hien Nguyen. 2021. "Shapley Values for Feature Selection: The Good, the Bad, and the Axioms." *IEEE Access* 9: 144352–60.
- Garbero, Alessandra, and Marco Letta. 2022. "Predicting Household Resilience With Machine Learning: Preliminary Cross-Country Tests." *Empirical Economics* 63(4): 2057–70.
- Gilligan, Daniel O., and John Hoddinott. 2007. "Is There Persistence in the Impact of Emergency Food Aid? Evidence on Consumption, Food Security, and Assets in Rural Ethiopia." *American Journal of Agricultural Economics* 89(2): 225–242.
- Guo, Baorong. 2011. "Household Assets and Food Security: Evidence From the Survey of Program Dynamics." *Journal of Family and Economic Issues* 32: 98–110.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.
- Hidrobo, Melissa, John Hoddinott, Neha Kumar, and Meghan Olivier. 2018. "Social Protection, Food Security, and Asset Formation." *World Development* 101: 88–103.
- Hossain, Marup, Conner Mullally, and M. Niaz Asadullah. 2019. "Alternatives to Caloriebased Indicators of Food Security: An Application of Machine Learning Methods." *Food Policy* 84: 77–91.
- Hsiang, Solomon, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, et al. 2020. "The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 Pandemic." *Nature* 584(7820): 262–67.
- Islam, Asadul, Chandana Maitra, Debayan Pakrashi, and Russell Smyth. 2016. "Microcredit Programme Participation and Household Food Security in Rural Bangladesh." *Journal of Agricultural Economics* 67(2): 448–470.
- Islam, Asadul, and Pushkar Maitra. 2012. "Health Shocks and Consumption Smoothing in Rural Households: Does Microcredit Have a Role to Play?" *Journal of Development Economics* 97(2): 232–243.
- Jain, Anil K., Jianchang Mao, and K. Moidin Mohiuddin. 1996. "Artificial Neural Networks: A Tutorial." *Computer* 29(3): 31–44.

- Jones, Lindsey, Mark A. Conostas, Nathaniel Matthews, and Simone Verkaart. 2021. "Advancing Resilience Measurement." *Nature Sustainability* 4(4): 288–89.
- Jordan, Joanne Catherine. 2015. "Swimming Alone? The Role of Social Capital in Enhancing Local Resilience to Climate Stress: A Case Study From Bangladesh." *Climate and Development* 7(2): 110–123.
- Josephson, Anna, Talip Kilic, and Jeffrey D. Michler. 2021. "Socioeconomic Impacts of COVID-19 in Low-Income Countries." *Nature Human Behaviour* 5(5): 557–565.
- Khandker, Shahidur R., M. A. Baqui Khalily, and Hussain A. Samad. 2012. "Seasonal Hunger and Its Mitigation in North-West Bangladesh." *The Journal of Development Studies* 48(12): 1750–64.
- Knippenberg, Erwin, Nathaniel Jensen, and Mark Conostas. 2019. "Quantifying Household Resilience With High Frequency Data: Temporal Dynamics and Methodological Options." *World Development* 121: 1–15.
- Kumar, I. E., S. Venkatasubramanian, C. Scheidegger, and S. A. Friedler. 2020. "Problems with Shapley-Value-Based Explanations as Feature Importance Measures." In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, edited by H Daume and A Singh, 5447–56. International Machine Learning Society (IMLS).
- Lieslehto, Johannes, Noora Rantanen, Lotta-Maria A. H. Oksanen, Sampo A. Oksanen, Anne Kivimäki, Susanna Paju, Milla Pietiäinen, et al. 2022. "A Machine Learning Approach to Predict Resilience and Sickness Absence in the Healthcare Workforce During the COVID-19 Pandemic." *Scientific Reports* 12(1): 8055.
- Little, Peter D., and Abdel Ghaffar Muhammad Ahmad. 2002. "Building Assets for Sustainable Recovery and Food Security." Broadening Access and Strengthening Input Market Systems.
- Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 30: 4765–74 <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.
- Manlosa, Aisa O., Jan Hanspach, Jannik Schultner, Ine Dorresteijn, and Joern Fischer. 2019. "Livelihood Strategies, Capital Assets, and Food Security in Rural Southwest Ethiopia." *Food Security* 11: 167–181.
- Maraboli, S. 2009. *Life, the Truth, and Being Free*. Port Washington, NY: A Better Today Publishing.
- Marcilio, W. E., and D. M. Eler. 2020. "From Explanations to Feature Selection: Assessing SHAP Values as Feature Selection Mechanism." In *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)* 340–47. IEEE.
- Martini, Giulia, Alberto Bracci, Lorenzo Riches, Sejal Jaiswal, Matteo Corea, Jonathan Rivers, Arif Husain, and Elisa Omodei. 2022. "Machine Learning Can Guide Food Security Efforts When Primary Data Are Not Available." *Nature Food* 3(9): 716–728.
- Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2): 87–106.
- Mullan, M., L. Danielson, B. Lasfargues, N C Morgado, and E Perry. 2018. *Climate-Resilient Infrastructure: Policy Perspectives*. OECD Environment Policy Paper No. 14. Paris: OECD.
- Pinstrup-Andersen, Per. 2009. "Food Security: Definition and Measurement." *Food Security* 1(1): 5–7.
- Raileanu, Laura Elena, and Kilian Stoffel. 2004. "Theoretical Comparison between the Gini Index and Information Gain Criteria." *Annals of Mathematics and Artificial Intelligence* 41: 77–93.
- Roberts, David L., Jeremy S. Rossman, and Ivan Jarić. 2021. "Dating First Cases of COVID-19." *PLoS Pathogens* 17(6): e1009620.
- Rudin-Rush, Lorin, Jeffrey D. Michler, Anna Josephson, and Jeffrey R. Bloem. 2022. "Food Insecurity during the First Year of the COVID-19 Pandemic in Four African Countries." *Food Policy* 111: 102306.
- Sakyi-Nyarko, Carlos, Ahmad Hassan Ahmad, and Christopher J. Green. 2022. "The Gender-Differential Effect of Financial Inclusion on Household Financial Resilience." *The Journal of Development Studies* 58(4): 692–712.
- Shapley, L. S. 1953. "A Value for n-Person Games." In *Contributions to the Theory of Games. Vol. 2 of Annals of Mathematics Studies*, edited by H. W. Kuhn and A. W. Tucker, 307–318. Princeton, NJ: Princeton University Press.
- Storm, Hugo, Kathy Baylis, and Thomas Heckeley. 2020. "Machine Learning in Agricultural and Applied Economics." *European Review of Agricultural Economics* 47(3): 849–892.
- The World Bank. 2021. "COVID-19 High Frequency Phone Survey of Households 2020—World Bank LSMS Harmonized Dataset." <https://microdata.worldbank.org/index.php/catalog/4072/study-description>.

- Vabalas, Andrius, Emma Gowen, Ellen Poliakoff, and Alexander J. Casson. 2019. "Machine Learning Algorithm Validation with a Limited Sample Size." *PLoS One* 14(11): e0224365.
- Villacis, Alexis H., and Syed Badruddoza. 2023. "Using Artificial Intelligence to Predict and Prevent Future Food Insecurity." *Georgetown Journal of International Affairs* 24(2): 191–97.
- Villacis, Alexis H., Syed Badruddoza, Ashok K. Mishra, and Joaquin Mayorga. 2023. "The Role of Recall Periods when Predicting Food Insecurity: A Machine Learning Application in Nigeria." *Global Food Security* 36: 100671.
- Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113(523): 1228–42.
- Walsh-Dilley, M., W. Wolford, and J. McCarthy. 2016. "Rights for Resilience: Food Sovereignty, Power, and Resilience in Development Practice." *Ecology and Society* 21(1): 11. <https://doi.org/10.5751/ES-07981-210111>.
- Warrens, Matthijs J. 2015. "Five Ways to Look at Cohen's Kappa." *Journal of Psychology & Psychotherapy* 5(4): 1.
- Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature Communications* 11(1): 2583.
- Zhang, Ying, and Chen Ling. 2018. "A Strategy to Apply Machine Learning to Small Datasets in Materials Science." *npj Computational Materials* 4(1): 25.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Villacis, Alexis H., Syed Badruddoza, and Ashok K. Mishra. 2024. "A Machine Learning-Based Exploration of Resilience and Food Security." *Applied Economic Perspectives and Policy* 46(4): 1479–1505. <https://doi.org/10.1002/aepp.13475>