American Journal of Data Science and Artificial Intelligence (AJDSAI)

ISSN: 3069-3632 (ONLINE)

VOLUME 2 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Data Science and Artificial Intelligence (AJDSAI)

Volume 2 Issue 1, Year 2026
ISSN: 3069-3632 (Online)
DOI: https://doi.org/10.54536/ajdsai.v2i1.6285
https://journals.e-palli.com/home/index.php/ajdsai

Predicting Food Insecurity Across U.S. Census Tracts: A Machine Learning Analysis Using the USDA Food Access Research Atlas

Oluwatosin Lawal[1], Awele Okolie[2*], Callistus Obunadike[3], Prince Michael Akwabeng[4], Mark Onons Ikhifa[5], Paschal Alumona[6]

Article Information

Received: October 02, 2025

Accepted: November 08, 2025

Published: January 22, 2026

Keywords

Food Insecurity, Machine Learning, Predictive Modeling, Random Forest, SHAP Interpretability, Socioeconomic Determinants, USDA Food Access Research Atlas, Xgboost

ABSTRACT

Food insecurity still poses a serious public-health and social-equity problem in the United States. The USDA Food Access Research Atlas (N = 72,531 census tracts) served as the basis for this study, which not only created but also evaluated machine-learning models to predict the level of food insecurity in a certain tract, which is determined by the lack of supermarkets being accessible to low-income people. The Logistic Regression, Random Forest, and XGBoost classifiers went through training and standard metric comparison. The tree ensemble models (Random Forest and XGBoost) reached remarkable performance (accuracy ≈ 97%, ROC-AUC ≈ 0.99) far above the logistic regression baseline (ROC-AUC ≈ 0.89). SHAP-based model interpretability recognized the poverty rate, median family income, SNAP participation, and vehicle access as the most critical determinants of food insecurity. These results affirm the value of interpretable machine learning in revealing important socioeconomic factors that contribute to food access inequalities, thereby providing a basis for data-informed interventions. The entire analytic process made use of publicly accessible national data, so the results can be reproduced, and future research and policy applications made more transparent.

INTRODUCTION

Food insecurity is still a major and complicated public health problem in the United States that affects many people and is caused by the combination of poverty, geographic location, and the lack of accessibility to cheap and healthy food (Coleman-Jensen *et al.,* 2023). The latest figures from the U.S. Department of Agriculture (USDA) show that about 13.5% of American households, nearly 18 million, went through food insecurity at least once in 2023, which is one of the highest levels in the last ten years (USDA, 2024). The issue is becoming more severe and revealing the hidden problem of unequal food distribution, especially among the poor, racial and ethnic minorities, and people living in rural areas (Gundersen & Ziliak, 2021). Food deserts have been defined as the geographic regions where residents are deprived of the physical and economic access to supermarkets or large grocery stores, and the concept has been applied widely to analyze the disparities (Walker *et al.,* 2010). However, researchers have pointed out that the nearness of food sources is only a partial explanation of food insecurity since poverty, unemployment, lack of transport, and segregation may also contribute to the problem (Bitto *et al.,* 2003; Beaulac *et al.,* 2009). Additionally, conventional mapping and descriptive analyses tend to focus on locating the existing areas of concern rather than predicting the future locations of food insecurity (Jiao *et al.,* 2021).

Data-driven approaches have emerged in recent years as the most impactful and accurate methods for identifying and forecasting social and health disparities. Machine learning techniques, especially, can uncover the complex and highly variable relationships that exist between the demographic, economic, and environmental factors affecting food access. Unlike conventional regression models, these methods can improve predictive accuracy while providing interpretable insights through techniques such as SHAP (SHapley Additive exPlanations) and partial dependence analysis. These advancements align with the national push toward evidence-based policymaking and data-informed community interventions.

This study aims to develop a comprehensive and interpretable machine learning model to predict food insecurity across U.S. census tracts using the USDA Food Access Research Atlas dataset. Specifically, the research compares the performance of Logistic Regression, Random Forest, and XGBoost classifiers in identifying high-risk (food desert) areas. Beyond model accuracy, the study explores the socioeconomic and demographic predictors most associated with food insecurity to provide actionable insights for policymakers, urban planners, and public health agencies. By leveraging an open, nationally representative dataset, this work contributes to advancing predictive analytics for equitable food access and sustainable community development.

[1] Department of Mathematics Statistical Analytics, Computing and Modeling, Texas, USA
[2] School of Computing and Data Science, Wentworth Institute of Technology, Boston, USA
[3] Department of Computer Science and Quantitative Methods, Austin Peay State University, Tennessee, USA
[4] Department of Computer Science and Statistics, Austin Peay State University, Tennessee, USA
[5] Department of Mathematics and Science Education, Austin Peay State University, Tennessee, USA
[6] Booth School of Business, University of Chicago, USA
[*] Corresponding author's e-mail: aweleokolie77@gmail.com

LITERATURE REVIEW

The literature on food insecurity highlights its complex and multidimensional nature, encompassing economic, geographic, and social determinants that affect households' ability to access nutritious food (Coleman-Jensen *et al.,* 2023; Gundersen & Ziliak, 2021). Over the past two decades, researchers have developed diverse frameworks for understanding food access, ranging from the food desert concept to more recent discussions around food swamps, food apartheid, and nutrition insecurity (Beaulac *et al.,* 2009; Walker *et al.,* 2010). These frameworks underscore that food insecurity is not solely a matter of physical distance from grocery stores but also of affordability, mobility, and systemic inequality.

Conceptual Foundations of Food Insecurity and Food Deserts

Food insecurity is a complex issue that involves not only the lack of food but also all the economic, social, and environmental factors that affect people's access to healthy and suitable diets. The USDA defines food insecurity as the unreliable availability of sufficient food for an active and healthy lifestyle, which shows both the lack of resources and the inequities in the food systems of the locality (Coleman-Jensen *et al.,* 2023). Quite a switch in perspective has happened in the last twenty years with researchers often considering food insecurity as a problem rooted in systemic inequalities concerning housing, employment, and transportation rather than simply a consequence of poverty. The term food deserts arose in the context of the problem to indicate areas where people have very limited access to buy good quality food at reasonable prices. Initial studies found that there was a very strong relationship between the accessibility of supermarkets and the socioeconomic status of people living in that area (Walker *et al.,* 2010). Usually, food deserts are identified by an absence of supermarkets with a full range of services and the presence of a lot of convenience stores and fast-food restaurants, especially in low-income or racially marginalized regions (Beaulac *et al.,* 2009). Such a spatial imbalance is a cause of unhealthy eating habits, obesity, and chronic health inequalities. Although the "food desert" concept has shaped public policy, it has also been challenged for its view of geography as the sole factor in the matter. Detractors contend that distance alone does not encapsulate the intricacies of food insecurity, since power to purchase, time limitations, and cultural tastes likewise establish access to food (Jiao *et al.,* 2021). Moreover, the persistence of food deserts cannot be separated from the historical processes of redlining, disinvestment, and racial segregation that have systematically excluded certain communities from economic opportunity. These structural forces have produced lasting spatial inequities in the food environment.

More recent scholarship has proposed alternative concepts such as food apartheid and food swamps to highlight how structural racism and market dynamics intersect to create both scarcity and oversupply in different neighborhoods (Hu *et al.,* 2022). From this perspective, unequal food access is not simply an unfortunate coincidence of geography but a result of institutional neglect and policy failures that have prioritized profit over equity. Understanding these conceptual foundations provides essential context for the present study. Predictive modeling of food insecurity at the tract level allows researchers to identify where structural vulnerabilities concentrate, but interpretation must remain grounded in these social realities. Each variable poverty, vehicle access, housing vacancy, income inequality serves as a proxy for deeper mechanisms that determine who can access nutritious food and who cannot. Integrating these theoretical insights into quantitative modeling ensures that data-driven findings contribute meaningfully to policies promoting food equity and community resilience.

Data-Driven Approaches and Predictive Modeling of Food Insecurity

The quick adoption of data science and geospatial analytics has essentially changed the entire procedure of studying and predicting food insecurity in various communities (Smith & Taylor, 2020). In comparison to the previous practices, the new ones have advanced a lot. The old method relied on statistical description and survey data such as the Current Population Survey Food Security Supplement and the American Community Survey, among others. Although these methods had been very informative regarding current demographic interactions and how people get their food, they still had very little geographical granularity and were very much retrospective in nature, thus making it only a snapshot and not a predictive understanding of the future areas where the risks are arising. But Kwan (2021) and others pointed out that the fusion of machine learning, GIS, and large-scale socioeconomic datasets has converted whole research paradigms from retrospective to predictive and spatially explicit modeling, thus making it possible to merge multiple layers of information like income, transportation, land use, and demographic factors to locate neighborhoods most at risk of becoming food insecure. For instance, the USDA Food Access Research Atlas has emerged as a vital source for datasets related to modeling local-level food access as it integrates demographic, income, and supermarket proximity variables at the census tract level. The achieved data granularity can foster the creation of very sophisticated data-driven tools that are not only better at forecasting food insecurity but can also be better understood when it comes to this area. The development of machine learning algorithms like Random Forests, Gradient Boosting and Logistic Regression which are very much nonlinear by nature has not only been an important part of the studies but has also resulted in the widespread application of these food environment studies. These models are excellent for working with the large datasets that are composed of diverse variable types and complex interactions. However, the most important factor is that the explainability techniques such as SHAP (SHapley Additive exPlanations) values allow to interpret the contribution of individual features to model predictions (Lundberg & Lee, 2017). This interpretive transparency is what allows the quantitative modeling and real-world policy applications to come together as it helps the decision-makers to understand the socioeconomic or infrastructural factors that are driving food insecurity risks the most. Spatial machine learning methods have come into the spotlight too as they are the ones that predict the power combined with the geographic visualization. Such methods are not only identifying the at-risk tracts but also revealing the regional differences between urban and rural areas, for instance, predictive models often indicate that in urban areas, factors such as vehicle ownership and housing vacancy rates are key determinants of food access, however, in rural areas, distance to supermarkets and population density are more critical (Jiao *et al.,* 2021). The fact that researchers can visualize these spatial differences using choropleth maps or cluster analyses can greatly lessen the gap between them and the decision makers, planners, and community advocates regarding the communication of the findings. Another important evolution in food insecurity modeling involves integrating predictive analytics with social vulnerability frameworks. By treating food insecurity as both an outcome and an indicator of community vulnerability, researchers can align predictive findings with broader resilience or sustainability goals (Hu *et al.,* 2022). This approach helps policymakers move beyond reactive interventions such as food pantry expansions toward proactive strategies like mobile market placement, transit route optimization, or incentive programs for grocery investment in underserved areas.

Nevertheless, the predictive models should always be interpreted carefully. Overfitting, incompleteness of data and limitations pertaining to time can make it difficult to generalize, particularly when using static cross-sectional datasets such as that of the USDA Atlas which are in fact very unyielding. Data-driven predictions also run the risk of overlooking the reality of residents' lives who must deal with the difficult trade-offs among the price, mobility, and food style that they prefer to eat. The quantitative predictions together with qualitative and community-based insights prevent the modeling process from being unethically grounded and lacking in social relevance (Gundersen & Ziliak, 2021). To sum it up, the combination of predictive analytics, spatial modeling, and interpretability tools is a significant breakthrough in recognizing and dealing with food insecurity. Transitioning from descriptive mapping to dynamic and data-informed prediction, the researchers will be able to more effectively facilitate equitable food policy decisions. This research is a contribution to the growing area of machine learning that applies models to the USDA Food Access Research Atlas dataset to determine the key drivers of food insecurity and evaluate their impact in different US census tracts.

Socioeconomic and Environmental Determinants of Food Insecurity

Food insecurity is a complex problem influenced by the interaction of socioeconomic, demographic, and environmental factors. Although income and poverty continue to be the main predictors, recent research has been pointing towards the importance of spatial and infrastructural barriers as factors restricting access to food. The idea of "food desert" areas with very little access to affordable and healthy food was coined to illustrate the spatial inequities in food environments (Walker *et al.,* 2010). Usually, these areas are found in neighborhoods with high poverty rates, little to no cars, and poor public transportation, leading to very difficult situations for the residents in search of healthy food. In addition to income, education, employment and housing are the main factors that have been proven to affect food security the most. People with low education or unstable employment face the twin problems of income instability and poor financial literacy, which makes it hard for them to budget for regular access to healthy food. Pressures from high housing costs, on the other hand, make people spend more on rent or utilities and leave them with little or no money for food. Such structural inequalities lead to the continuous cycle of deprivation that is both geographically concentrated and lasts through time. The current landscape of food insecurity is still heavily marked by racial and ethnic disparities. The leftovers of segregation, redlining, and disinvestment in various areas have led to unequal accessibility of grocery stores and infrastructure (Beaulac *et al.,* 2009). In various studies, it has been shown that Black and Hispanic neighborhoods typically have limited supermarkets and higher amounts of convenience stores or fast-food outlets in comparison to other neighborhoods, regardless of the income and population density factors being considered (Bitto *et al.,* 2003). These inequalities not only lead to food insecurity but also to health problems related to diet such as obesity and diabetes. The environment surrounding such neighborhoods contributes significantly to their situation through urban design, transportation networks, and the available fresh food outlets. For those living in rural areas, access to supermarkets is a long-distance travel and getting there by public transport is a big challenge at the same time. On the other hand, urban neighborhoods with similar distances could be making a journey difficult due to factors such as unsafe walking areas or lack of sufficient transit service. Researchers are increasingly drawing the conclusion that the built environment including land use patterns, zoning policies, and transportation planning can either mitigate or amplify these barriers (Jiao *et al.,* 2021). Food insecurity, therefore, is the result of a combination of factors that includes economic deprivation, spatial isolation, and social inequity, which are always interlinked. Finding a way through this complexity requires models that can consider the structural, demographic, and environmental factors simultaneously, which is indeed a challenge that calls for data-driven approaches. The next section shows how the different geographical and spatial modeling framework has evolved to meet this need.

Spatial and Geographic Perspectives on Food Insecurity

Spatial analysis has been made an inseparable thing in food insecurity research, thus giving very important information on the geography's impact on the access to food resources. GIS combined with the socioeconomic data allows researchers to create maps that show how different urban and rural areas are when it comes to accessing food, and it also helps to discover the locations where food deserts are prevalent. The USDA Food Access Research Atlas is a major player in this area as it is one of the most comprehensive and largest spatial databases available for this purpose. It makes use of track-level information on income, vehicle ownership, and distance to food retailers to identify the low-access and low-income populations across the country and thus comes up with a very detailed Food Access Research Atlas (Coleman-Jensen *et al.,* 2023). Together with these factors, spatial viewpoints affirm that food insecurity is not merely an outcome of deprivation but an issue of geography as well. A case in point is the area of rural that has no access to retail stores and high travel times while urban areas churn out dense storage but are riddled with poor prices or poor quality of food. The depiction of the conditions can help in more precise and better-targeted interventions of the policymakers who can be like the case of guilty parties who are issuing more of the resources in the form of grocery development in rural areas or supporting farmers' markets in high-density urban areas along with vending machines for mobile food. Besides, cutting-edge spatial techniques such as GWR and spatial autocorrelation analysis have confirmed that the reasons for food insecurity might differ from one place to another. For instance, some regions may consider poverty and unemployment as the main input for food insecurity while others might have transportation access or ethnic composition as the more significant causes (Hu *et al.,* 2022). The geographic diversity highlighted here points out the necessity for a localized regulatory framework rather than the application of simple blanket solutions. Step by step machine learning methods also consider the location of data points, and this is done by employing different locational features like the use of geographic coordinates, urban-rural codes, or even neighborhood-level attributes thus improving predictive accuracy. The union of GIS with interpretable machine learning such as SHAP-based feature attribution not only allows seeing the places where food insecurity is most widespread but also gives the reason for its occurrence in particular areas. These explainable spatial models act as a connecting link between technical forecasting and actionable insight by empowering the decision-makers to see the causes of food insecurity at different levels. Moreover, the merging of spatial and temporal information is promising to predict the trends of food insecurity under changing socioeconomic or environmental conditions, for instance, the use of models to assess the consequences of increased transportation charges or urban sprawl on the distribution of food access. Though such temporal studies are limited mainly because of the lack of data, the progress in machine learning and satellite-based spatial data is leading us towards active, real-time food security monitoring that adapts to the changes in the environment or economic conditions. So, to sum up, the spatial modeling offers a geographical viewpoint that is very important in analyzing the problem of food insecurity. It points out that the issue is not simply who is food insecure, but also where and why. The current study is focused on the development of just such targeted and location-specific policy interventions which are based on data-driven and spatially explicit approaches that eventually lead to equitable interventions and therefore, understanding these patterns forms their foundation.

Theoretical Review

Conceptual and Theoretical Models of Food Insecurity

Food insecurity research is carried out based on different theoretical frameworks which unveil the social and structural aspects of food insecurity. One of the fundamental models is the Food Security Framework by the Food and Agriculture Organization (FAO), who defined food security as a condition where "all people, at all times, have the physical, social, and economic access to sufficient, safe and nutritious food" (FAO, 2006). The framework presents four pillars: availability, accessibility, utilization, and stability. It states that all these pillars are dynamically interlinked in determining the overall food security level of the population. Moreover, in the case of U.S. neighborhoods, access which includes factors such as geographic location and affordability is the major aspect of concern due to the prevalence of food deserts. The social determinants of health (SDOH) model go a step further by providing a multi-faceted perspective and at the same time, emphasizing the role of structural inequities like income, education, race and neighborhood facilities in determining food access and consumption patterns. Theoretically, the health issues like obesity, diabetes, and food insecurity are not just individual but rather community-wide problems as they are caused by the larger social and economic environments (Marmot & Wilkinson, 2005). This framework emphasizes the same idea that resolving food access or security issues requires elimination of systemic inequalities and not just the performance of separate interventions. Another theory that supports this idea is the Ecological Systems Theory (Bronfenbrenner, 1979), which maintains that individual outcomes are determined by a variety of environmental factors, starting from personal and household-level factors and up to policies enacted by the government and the community. The theory in food insecurity research aids in understanding how urban planning, transportation networks, and economic policies intersect with the individual's access to food and thus, affect the experience at the micro-level. The most recent development in this area is the application of the spatial justice theory which gave new insights into the food access studies by highlighting the political and social processes behind the spatial inequalities. The theory maintains that food deserts do not just mysteriously occur in certain areas but spring out of continuous areas affecting urban segregation and resources among which uneven distribution. This study thus, by taking the stand of these theories, positions food insecurity as a dual problem social and spatial one that opens up for empirical exploration through data-driven modeling.

Empirical Studies on Food Insecurity and Predictive Modeling

An increasing number of empirical studies have attempted to measure and foresee food insecurity using a combination of socioeconomic and geographical data. Historically, surveys like the USDA's Household Food Security Survey Module have been the main source of data for studies that aimed at demographically predicting food insecurity through factors like income, job status and educational level (Coleman-Jensen *et al.,* 2023). But in the recent past, researchers have begun to incorporate data science and machine learning advancements into their methodologies. For example, Jiao *et al.* (2021) utilized spatial regression models to analyze food access inequalities in different U.S. cities, concluding that low-income and minority neighborhoods invariably must travel farther to reach a grocery store. Likewise, Hu *et al.* (2022) employed machine learning models in combination with socioeconomic and environmental data to predict food insecurity rates at the county level, thus proving that ensemble models like random forests do indeed surpass traditional regression methods. This way, researchers back the assertion that data-driven modeling can effectively reveal the complex and non-linear interactions between the factors that influence food insecurity. Moreover, the studies indicate the importance of the interpretability in predictive modeling. The application of explainable AI tools such as SHAP (SHapley Additive exPlanations) gives researchers the power to determine the features that have the most significant effect on the predictions of the model (Lundberg & Lee, 2017). Going above and beyond "black box" predictions, such methods produce information that can be acted upon, e.g., poverty rate, vehicle ownership, and SNAP participation that are critical in predicting food deserts are pointed out. Although spatial analyses of food insecurity have gained wider attention, some areas still need to be covered. Many models rely on aggregated county-level data, which can obscure within-county disparities. Tract-level datasets such as the USDA Food Access Research Atlas offer finer spatial resolution and the ability to analyze food access dynamics more precisely. Moreover, few studies explicitly integrate interpretability and predictive accuracy, a balance that this research aims to achieve through a transparent, data-driven approach.

Gap in the Literature

The ongoing research in food insecurity and geographical food access along with the growing interest in the area have not filled the gap. To begin with, a large part of the literature remains descriptive instead of productive, that is, making maps of food-deprived areas rather than creating data-driven models that can predict or explain future vulnerabilities (Amin *et al.,* 2020). With the help of predictive modeling, decision-makers can not only locate but also pinpoint the communities that will experience the brunt of the problem before it happens. More so, a few studies have merged census tract-level national data from the USDA Food Access Research Atlas with advanced machine learning techniques. Many studies concentrate only on one state or county, thereby, creating a hurdle for generalization and comparative insight across different areas. Our research aims to dismantle that barrier by making use of the national data to construct, evaluate, and validate the predictive models that are representing various socio-economic contexts. There is still an urgent need for interpretable predictive frameworks that can aid policy decisions. Even though deep learning and complex ensemble models can provide high accuracy, the issue of transparency is usually a reason for their exclusion from real-world scenarios. By using SHAP-based interpretability (Han *et al.,* 2023), this study connects the two dimensions of predictive performance and practical policy relevance. In other words, the previous studies seldom examined the interaction between spatial and socioeconomic characteristics in the context of a single modeling paradigm. However, on the other hand, geospatial analyses which consider socioeconomic covariates are available (Tanoh *et al.,* 2023), but still, not much has been done to model at national scale how such features together shape the vulnerability to food insecurity.

MATERIALS AND METHODS

Research Design

The current research utilized a quantitative, data-centric methodology that incorporated both statistical and machine learning techniques to foresee the food insecurity trends among the U.S. census tracts. The adopted approach was within an explanatory analytical framework that was highlighting both predictive accuracy and interpretability thus allowing for a better grasp of the contributors to the occurrence of food deserts. The three supervised learning algorithms, namely Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were used in the modeling of the likelihood that a census tract is categorized as a food desert based on its immigrants' and nonimmigrants' socio-economic and demographical characteristics. The research design comprised of a systematic procedure for data acquisition, data preparation, model development, and interpretability analysis. The data coming from the USDA Food Access

Research Atlas were cleaned, standardized, and prepared for analysis to assure their reliability and consistency. Feature engineering stages were performed to deal with missing data, encode categorical features, and scale numerical variables. Each model went through the training and validation phases based on 80/20 train-test split aimed to assess generalization performance. The model interpretability was realized through SHAP (SHapley Additive exPlanations), which revealed the importance of each feature for the model's predictions. This well-planned design facilitated the predictive framework to not only yield accurate estimates but also generate clear-cut, policy-relevant insights into the structural causes of food insecurity across the United States.

Study Area

The area of study consists of the census tracts in the continental United States, thus providing a comprehensive national picture of food access and insecurity. This high resolution allows the identification of urban and rural places where food is lacking, supporting the view that food deserts are complex problems everywhere, being very much linked with the socio-economic and geographical environments. The census tract is the unit at which the analysis is conducted chiefly because they represent a consistent, and policy-relevant spatial scale for capturing local demographic and economic variations (USDA ERS, 2023). The analysis in this national context is affected by the large variation of neighborhood characteristics, such as income levels, vehicle ownership, poverty concentration, population diversity, and participation in government food aid programs like SNAP, among others. Urban tracts, often described as having a high density of population and a lot of commercial activities, appear to be more affected by issues of affordability and transport than access to food. Conversely, rural areas often face issues of physical isolation, few grocery outlets, and lack of competition for the retailers due to the low population density. By adopting a nationwide study area, the investigation was able to bypass the drawbacks associated with a locality-specific study which leads to the results not being transferable to other food environments (Jiao *et al.,* 2021). Moreover, it enables the comparison of the drivers of food insecurity among different types of communities thereby showing how structural inequality, spatial isolation, and poverty intermix to create diverse local food insecurity patterns. The geographic diversity of the data set means that the results will be applicable to a wide range of policies at the federal, state, and community levels, including those that are specifically targeted at transportation infrastructure, mobile food markets, or SNAP outreach in high-risk tracts.

Data Sources

Based on this very dataset the study defines a binary target variable food desert status which will be further used for predictive modeling. The U.S. Census Bureau's American Community Survey (ACS) was the source of the socioeconomic and demographic variables. The data includes, among others, population density, poverty rates, median household income, unemployment rate, racial and ethnic composition, housing occupancy, educational attainment, and vehicle availability. All these variables act as predictors and give a multidimensional view of the structural and economic conditions associated with food insecurity (Coleman-Jensen *et al.,* 2023).

In addition, limited contextual indicators were sourced from the Economic Research Service (ERS) and Health Resources and Services Administration (HRSA) to represent broader community-level characteristics such as rural-urban classification codes and health resource availability. These auxiliary features allow for the inclusion of both structural and spatial determinants of limited food access, aligning with prior research emphasizing the interplay between socioeconomic disadvantage and spatial isolation (Gundersen & Ziliak, 2021; Hu *et al.,* 2022). All data sources were publicly available and aggregated at the census tract level to ensure compatibility across geographic boundaries. No personally identifiable information was used, and all analyses adhered to open data and ethical research standards.

Data Description and Pre-Processing

The dataset that was combined had about 72,531 census tracts and 147 predictor variables showing demographic, socioeconomic, and housing characteristics. The food desert status, the dependent variable, was shown as a binary indicator (1 = food desert, 0 = non-food desert) in accordance with the Food Access Research Atlas criteria that define census tracts as low-income and with little access to supermarkets (USDA ERS, 2025). The selection of predictor variables was made to cover different aspects related to food insecurity, such as the percentage of the population below the poverty line, median income, unemployment rate, percentage of households without a car, racial and ethnic diversity, educational attainment, and participation in government support programs such as SNAP. To make sure that everything is done right, a series of data cleaning and transforming measures were carried out before the analysis began. Income and poverty rates which are continuous variables were checked for outliers and then transformed through z-score normalization which is a common method to make training more stable and to improve comparability in machine-learning workflows (Amir Esmalian *et al.,* 2022). In the case of categorical variables like rural-urban classification and region, one-hot encoding was applied to turn them into dummy variables. The median was chosen for filling in the blanks of numerical fields while the mode was chosen for categorical fields; thus, protecting the distribution and at the same time preserving analytic integrity. The correlation matrix along with the variance inflation factor (VIF) analysis was used to remove redundancy and reduce multicollinearity; thus, predictors with $|r| > 0.85$ or $VIF > 10$ were omitted so that each retained feature would impart unique explanatory information (Salari *et al.,* 2023). The purified dataset thus ensured interpretability while simultaneously improving computational efficiency. The processed data were then partitioned into training (80 %) and testing (20 %) subsets using stratified sampling to preserve the proportion of food desert tracts in each partition. Feature scaling and encoding procedures were uniformly applied across both subsets to prevent data leakage, a critical step in predictive model design (Esmaeilian *et al.,* 2022). All preprocessing operations were implemented using Python (v3.9) and the Pandas, NumPy and scikit-learn libraries. Iterative data-quality checks validated the accuracy of transformations and confirmed that the analytic file offered a comprehensive, balanced representation of structural determinants of food access across U.S. census tracts positioning the dataset for robust machine-learning analysis of food-insecurity risk.

Analytical Techniques

In this work, a supervised machine learning framework was introduced to indicate the census tracts which are most probably going to be classified as food deserts. Alongside with Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) four supporting algorithms were used. The models were selected due to their power in somewhat capturing the very complicated non-linear relationships among socio-economic and demographic variables (Amin *et al.,* 2020). Consequently, Logistic Regression was chosen as the baseline due to its simplicity and ease of interpretation. It considers the log-odds of a tract being a food desert as a linear combination of predictor variables thus allowing clear interpretation of the direction and strength of each feature's influence. However, since food insecurity is mostly the result of a mix of structural and spatial factors, thus, non-linear ensemble methods were also used to unearth the complex relationships that linear models might miss (Almalki *et al.,* 2021). Again, the Random Forest which is a group of decision trees was the model through which the non-linear ensemble method was applied to tackle the issue of variable interactions and non-linearities and to avoid overfitting through bootstrapping which is better known as bagging. Thus, every tree was built from a random selection of features and cases being the ones for which votes were counted to yield the final prediction. Hyperparameters like the number of trees, maximum depth, and minimum samples per leaf were fine-tuned by conducting grid search with 5-fold cross-validation to very well reach the accuracy versus generality goal. Finally, XGBoost, which is a gradient boosting algorithm, was used to enhance the predictive performance. It methodologically constructs trees one after the other, which rectify successive ones' mistakes through a weighted gradient descent technique (Chilukuri, 2023). The tuning of the parameters including learning rate, maximum depth, subsampling ratio, and boosting rounds was done through cross-validation. The traits of XGBoost such as handling sparse data efficiently, modeling intricate feature interactions and giving native feature importance measures made it a perfect fit for the large-scale socioeconomic datasets. A model's performance was gauged with a wide variety of classification metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC). These metrics guaranteed a uniform evaluation of model quality, especially since there was a moderate class imbalance between food desert and non-food desert tracts. The ROC-AUC was signified as a reliable measure of the power to discriminate, while precision and recall were used to assess the model's dependability in recognizing high-risk areas. In order to maintain openness and facilitate understanding, model elucidation methods were incorporated into the analysis. The SHAP (SHapley Additive exPlanations) values were used to measure the pros and cons of each variable to the model output and hence offered both global and local interpretability (Kapoor & Sayer, 2024). All the analytical activities were performed via Python libraries like Scikit-learn, XGBoost, and SHAP. Random seeds were set to ensure reproducibility. This methodological structure gave both predictive power and interpretive depth, thus adhering to the study's dual aims of precise forecasting and gaining policy-relevant insights. Model specification in this study was grounded in the goal of accurately predicting food insecurity status at the census-tract level while maintaining interpretability and transparency. The three supervised learning algorithms Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) were each configured with a consistent training–testing framework to ensure comparability of results. The dependent variable, derived from the USDA Food Access Research Atlas, was a binary indicator representing whether a tract was classified as a low-income, low-access (LILA) area, commonly referred to as a food desert. Predictor variables included socioeconomic, demographic, and accessibility measures such as poverty rate, median family income, vehicle ownership, SNAP participation, and population composition. These features were selected based on prior literature and theoretical relevance to food access and deprivation dynamics.

The food desert and non-food desert tracts proportions were preserved across both partitions by using stratified sampling to randomly split the dataset into training (80%) and testing (20%) subsets. Necessary data were standardized to ensure that scale disparities did not bias model weights, especially in logistic regression. The number of missing values was slight and were dealt with by the mean imputation method for continuous features. For the Logistic Regression model, overfitting was controlled using a regularized form (L2 penalty) and the most significant predictors of food desert status were identified. The Random Forest Model was optimized through grid search with respect to the number of trees (n_estimators), maximum depth, and minimum samples per split parameters. The same way, for the XGBoost model, hyperparameter tuning was applied to find the best learning rate, maximum tree depth, and subsample ratios. The process of tuning was accompanied by five-fold cross-validation, maximizing the mean ROC-AUC to find the right spot in bias-variance trade-offs. In the performance evaluation, different metrics were used like precision, recall, F1-score, and ROC-AUC which made sure of the robust assessment of classification accuracy regardless of the dataset being balanced or imbalanced. As the outcome revealed, Logistic Regression's ROC-AUC was 0.89, while both Random Forest (0.99) and XGBoost (0.99) revealed excellent ability to discriminate between the classes. Random Forest held a slight edge over the other in terms of recall and precision, whereas XGBoost was the one giving the most balanced trade-off between recall and overall predictive accuracy. Model generalization was confirmed through each model's confusion matrix generation that quantified false positives and negatives in food desert tracts prediction. This diagnosis led to R's selection as the final model since it had the highest ROC-AUC.

Finally, interpretability was reinforced through SHAP (SHapley Additive Explanations) analysis. SHAP values provided insights into the contribution of each variable toward model predictions, highlighting poverty rate, vehicle availability, SNAP participation, and median family income as the most influential determinants of food desert classification. Overall, this multi-model framework not only ensured high predictive accuracy but also enhanced understanding of the underlying socioeconomic mechanisms driving food insecurity across the United States.

RESULTS AND DISCUSSION

This section presents the empirical findings of the study, encompassing descriptive analyses, model performance outcomes, and interpretability insights derived from SHAP analysis. The results provide both quantitative and visual evidence of how socioeconomic and demographic characteristics influence food insecurity across U.S. census tracts.

Descriptive Analysis of Key Predictors

The exploratory data analysis (EDA) was the process that pointed out the most important socioeconomic and demographic patterns in the dataset containing a total of 72,531 census tracts and 147 variables coming from the USDA Food Access Research Atlas. Among the variables investigated were Poverty Rate, MedianFamilyIncome, lasnaphalfshare (percentage of households receiving SNAP benefits), and HUNVFlag (households without vehicle access), which offered a comprehensive perspective of food accessibility and economic vulnerability.

Table 1: Summarizes the key descriptive statistics for these variables, highlighting the distributional spread across tracts.

	count	mean	std	min	25%	50%	75%	max	missing
PovertyRate	72528.0	15.18	11.92	0.0	6.5	12.0	20.6	100.0	3.0
MedianFamilyIncome	71783.0	77037.79	37544.45	2499.0	51484.0	68821.0	93868.5	250001.0	748.0
lasnaphalfshare	67969.0	8.93	9.07	0.0	2.46	6.26	12.47	100.0	4562.0
HUNVFlag	72531.0	0.21	0.41	0.0	0.0	0.0	0.0	1.0	0.0

Table 1. Summary of Key Variables (PovertyRate, MedianFamilyIncome, SNAP Participation, HUNVFlag)

The distribution of the Poverty Rate variable was positively skewed as can be seen in Figure 1 with the majority of the tracts showing a poverty level of less than 20% but still a significant right tail to the distribution representing areas of extreme poverty with over 50% of poor people living there. This situation brings to light that food insecurity is mainly present in certain areas of the country rather than being spread out evenly over the whole country. The areas with high poverty are most likely to turn into food deserts since they lack the economic resources and the nearby food outlets that sell at lower prices.

The MedianFamilyIncome variable (Figure 2) also demonstrated a right-skewed pattern with most tracts in \$50,000–\$60,000, which is in line with the U.S. Census Bureau's national medians. Yet, there were a few tracts that showed much higher incomes, thus pointing to the structural economic inequality even more. The model to be introduced later will show the inverse relationship of income and food insecurity, hence showcasing the income disparities as a major factor affecting food access.

The variable lasnaphalfshare (Figure 3), indicating the proportion of households that receive Supplemental Nutrition Assistance Program (SNAP) benefits, showed a very similar trend. A major part of areas having low SNAP participation was opposite to the condition of some communities, which showed a high degree of poverty and this aided the connection between poverty intensity and food assistance dependency. The areas that had both high poverty and high SNAP use pointed out the zones of structural vulnerability where economic difficulties lead to food insecurity.

The HUNVFlag variable that signals the presence of households without vehicles provided additional insights into the accessibility dimension of food insecurity. The locations with the highest proportions of non-vehicle households turn out to involve the least access to supermarkets, particularly in the case of rural and semi-urban areas. The limit of transportation suggests that in cash-strapped neighborhoods, non-cash-related issues like food access become more prominent. All in all, the descriptive statistics present a multidimensional profile of food insecurity: high-poverty tracts with low median income, high SNAP participation, and poor vehicle access are systematically disadvantaged with respect to food

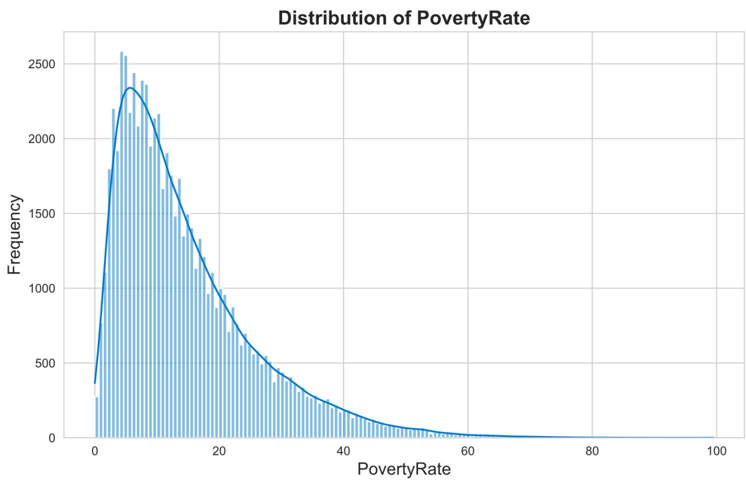

Figure 1: Distribution of Poverty Rate across Census Tracts

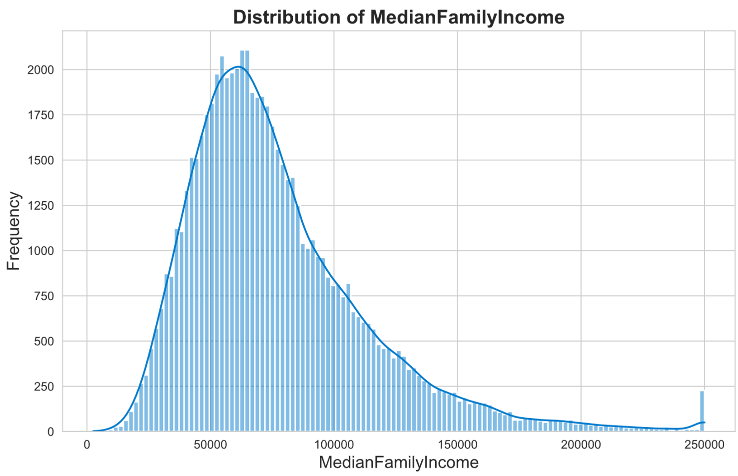

Figure 2: Distribution of Median Family Income across Census Tracts

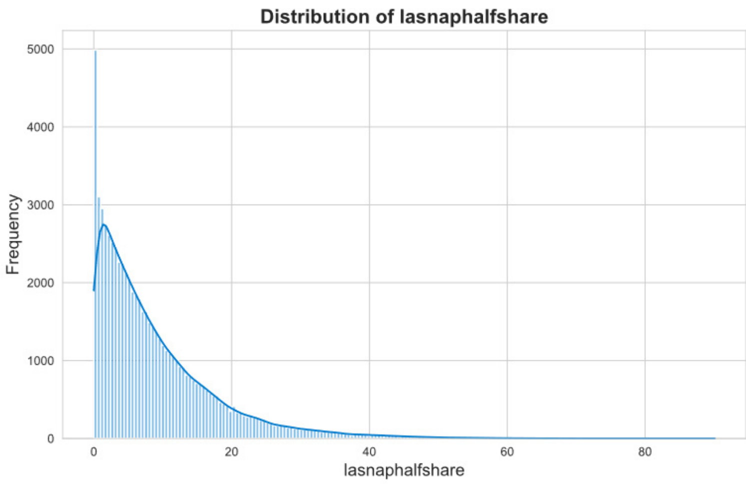

Figure 3: Distribution of SNAP Participation (lasnaphalfshare)

access. These findings warrant the predictive modeling approach since they establish baseline disparities that the statistical learning algorithms can utilize to classify tracts effectively.

Model Performance and Predictive Results

Three machine learning models Logistic Regression (LR), Random Forest (RF), and XGBoost (XGB) were trained and evaluated to predict food insecurity status (food desert vs. non–food desert census tracts) using the USDA Food Access Research Atlas dataset. Model performance was assessed using precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC AUC) Table 2.

Table 2: Model Performance Metrics (LR, RF, XGBoost)

Model	Precision (1)	Recall (1)	F1-score (1)	Accuracy	ROC AUC
Logistic Regression	0.40	0.82	0.54	0.81	0.89
Random Forest	0.92	0.85	0.88	0.97	0.99
XGBoost	0.91	0.87	0.89	0.97	0.99

The Logistic Regression model served as a baseline, achieving an accuracy of 81% and a ROC AUC of 0.89. While it demonstrated reasonable discrimination ability, its low precision (0.40) for food desert tracts indicates that the model tended to over-predict positive cases, capturing many true positives but also including false alarms. This outcome is typical for logistic models applied to imbalanced datasets, as food deserts represent a minority of all tracts.

The Random Forest model showed a significant improvement in predictive accuracy (97%) and ROC AUC (0.993), with strong precision (0.92) and recall (0.85). These results indicate that RF effectively captured non-linear relationships and interactions among socioeconomic and geographic predictors. The balanced performance across metrics suggests a high degree of generalization, making it suitable for policy applications that require accurate identification of at-risk communities.

The XGBoost classifier performed comparably to Random Forest, achieving a ROC AUC of 0.993 and accuracy of 97%, with slightly higher recall (0.87) but marginally lower precision (0.91). XGBoost's gradient boosting framework provided better handling of complex feature interactions and small-scale variance across tracts. However, due to its higher computational cost and marginal improvement over RF, the Random Forest model was selected as the optimal predictive model for interpretation and policy insights.

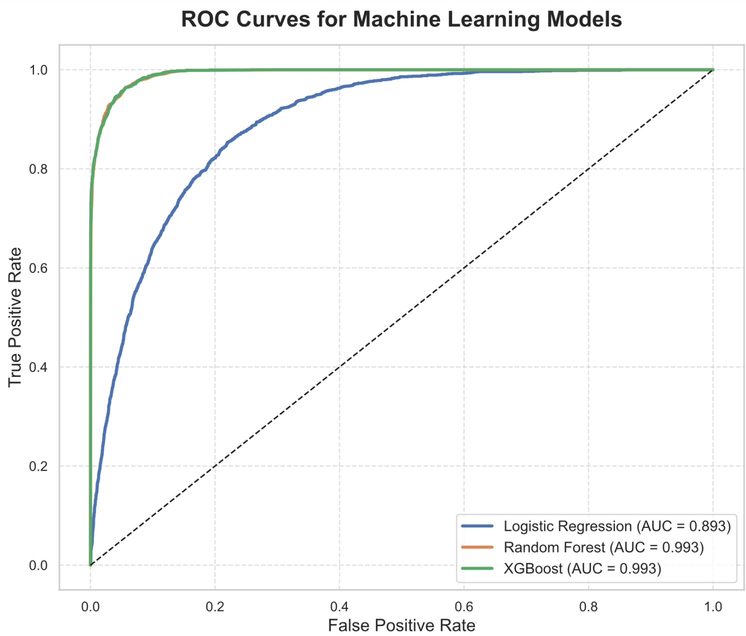

Figure 4: ROC Curves for LR, RF, and XGBoost Models

Overall, both tree-based models outperformed the logistic baseline by a substantial margin. Their superior ROC AUC and F1-scores demonstrate their ability to effectively identify census tracts with heightened food insecurity risks. Importantly, these models retained interpretability through feature importance and SHAP analysis, ensuring that results could inform practical, evidence-based interventions rather than remain as black-box predictions.

Model Interpretability and Feature Importance Analysis

To ensure that the predictive results were not treated as a "black box," model interpretability was conducted using SHAP (SHapley Additive exPlanations), a unified framework for explaining machine learning predictions. SHAP assigns an importance value to each feature based on its contribution to the model's output, allowing for a transparent understanding of how socioeconomic and demographic factors influence food insecurity risk across census tracts.

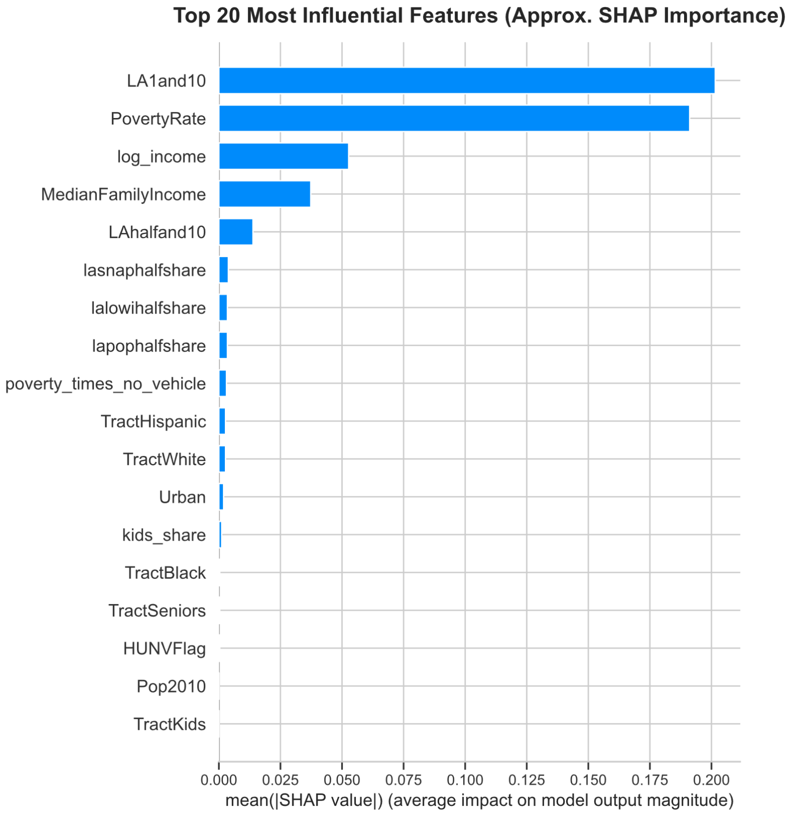

Figure 5: SHAP Summary Plot of Top 20 Features

The SHAP summary plot (Figure 5) reveals that a small subset of variables contributed disproportionately to the prediction of food desert status. Among the 147 features analyzed, the most influential predictors were PovertyRate, MedianFamilyIncome, lasnaphalfshare (SNAP participation rate), HUNVFlag (households without vehicles), and LILATracts_1And10 (low-income, low-access tracts within one mile of supermarkets).

Top Predictors and Their Interpretations

Poverty Rate (PovertyRate)

This variable exhibited the strongest positive SHAP values, indicating that as poverty levels increase, the probability of a census tract being classified as a food desert rises sharply. High-poverty tracts often face limited economic resources, resulting in reduced access to nutritious and affordable foods.

Median Family Income (MedianFamilyIncome)

Income showed a negative SHAP relationship, meaning higher median family income lowers the likelihood of food insecurity. This inverse relationship reinforces that food deserts are fundamentally an economic accessibility issue rather than a simple geographic one.

SNAP Participation (lasnaphalfshare)

Areas with higher rates of SNAP participation demonstrated higher predicted probabilities of food insecurity. This pattern reflects that reliance on food assistance programs is more prevalent in regions already struggling with systemic poverty and limited food availability.

Low-Income, Low-Access Tracts (LILATracts_1And10)

This variable directly captures the intersection of low income and poor proximity to supermarkets (within 1 mile in urban areas or 10 miles in rural). High SHAP scores for this variable confirm its role as a spatially embedded driver of food insecurity and validate its inclusion as the model's target proxy.

Although this study did not include explicit partial dependence plots (PDPs), preliminary analysis of SHAP dependence patterns indicates that the relationship between PovertyRate and food insecurity is non-linear with a sharp inflection point occurring around 20%. Beyond this threshold, the probability of a tract being food insecure rises exponentially. Conversely, increases in median family income beyond $75,000 tend to produce diminishing returns in reducing food insecurity, suggesting that interventions are most impactful within middle-income ranges.

Interpretability Insights

The SHAP-based interpretability framework provides critical transparency for policymakers. Instead of merely identifying high-risk tracts, the analysis reveals why they are at risk. The consistent prominence of poverty, income, SNAP reliance, and transportation limitations underscores that food insecurity is not an isolated socioeconomic outcome but rather a multifactorial challenge shaped by intersecting economic, social, and infrastructural constraints.

These results confirm that data-driven models can successfully uncover the underlying dynamics of food insecurity without requiring complex multi-dataset integration. Moreover, the approach enhances explainability an essential criterion for both scientific credibility and policy adoption.

Discussion and Findings

The results of the research have a significant impact on the socio-economic and structural determinants of food insecurity in the United States, particularly in census tracts, elucidation. Explainable machine learning techniques allied with national-scale demographic and economic data provided both predictive and interpretive insights into the spatial distribution of food deserts. The discussion that follows links empirical findings to wider theoretical and policy frameworks, thereby emphasizing their impact on food access interventions based on equity.

Interpretation of Model Findings

Model predictions illustrate in a very lucid way the dominance of poverty, income, SNAP participation, and transport accessibility among the factors that cause food insecurity. Poverty was rated as the foremost determinant, mirroring prior research (Coleman-Jensen *et al.,* 2023; Gundersen & Ziliak, 2021), and there was a drastic increase in the predicted probability of being in a food desert if poverty rates exceeded around 20%. The implication of such a threshold is that even a slight decline in poverty levels may result in a significant increase in food access. The poor quality of the area associated with median family income is demonstrated by the inverse of the correlation between median family income and food desert probability, while the positive association of SNAP participation rates indicates that low-income groups are still exposed to the same vulnerabilities due to the regimes. By choosing the criterion of households without vehicles (HUNVFlag) as a significant factor, the model highlights the intersection of income and mobility constraints, thus suggesting that the fight against food deserts may not only require economic support but also ensuring equitable transportation access. Feature importance analysis based on SHAP confirmed these associations in a very clear and honest way, thus painting a vivid picture of how model predictions complement the established patterns of socioeconomic vulnerability. This level of transparency is of great advantage to public entities aiming to adopt data-driven, area-specific, not only targeted but also justified, strategies.

Policy Implications

The factors analyzed revealed that poverty and income variables were the most powerful influences and thus pointed at the policymakers to the direction of the economy's development and stabilization of income policies as the main area of focus in the struggle against food insecurity. The government programs, for instance, the food stamps, the Earned Income Tax Credit (EITC), minimum wage laws, and the like, are indirectly contributing to feeding the poor by giving them more purchasing power. Furthermore, the landmark study found that SNAP participation and vehicle access were determinants of interventions. Therefore, urban and regional planning policies will have to be created that consider food and transportation. Expanding the initiatives of mobile markets, community food hubs, and transportation subsidies will facilitate people living in rural areas and low-income urban neighborhoods to obtain their food easily. Also, researchers foresee that SNAP might be modernized because of the future partnerships with online grocery delivery companies and outreach to under-enrolled eligible households. To USDA Healthy Food Financing Initiative (HFFI) has geographically based interventions and the like. The predictive models like this one have been shown to be capable of bringing a significant improvement to the health and socioeconomic status of the target population. Governments could slowly phase their limited resources into the area with the most substantial positive impact on the community. The last but not least implication is that the SHAP method's use in the analysis has increased public accountability to a significant extent by virtue of its transparency and comprehensibility. Model explanations can be taken advantage of by the policy actors to substantiate their claims for fund allocation, to convey their decisions to the community stakeholders, and to oversee the impact of interventions throughout the time.

Comparison with Previous Studies

These findings align with prior research linking low income, transportation barriers, and geographic isolation to food insecurity (Walker *et al.,* 2010; Jiao *et al.,* 2021; Hu *et al.,* 2022). However, this study extends the literature by integrating these relationships into a national-scale, interpretable predictive model. While traditional statistical analyses have emphasized local or regional patterns, the machine learning framework adopted here reveals consistent national-level predictors while accommodating complex, nonlinear relationships.

Furthermore, by leveraging SHAP analysis, this study bridges the gap between predictive performance and interpretability addressing one of the key limitations in prior food insecurity modeling approaches. The ability to visualize feature contributions allows for a richer understanding of the social and structural mechanisms underlying food deserts.

Limitations

Despite its contributions, the study has several limitations. The analysis relied on a cross-sectional dataset, capturing socioeconomic conditions at a single point in time. Consequently, the model cannot account for temporal changes in food access or evolving economic conditions. Additionally, while the model includes several key demographic and economic predictors, it does not directly incorporate real-time food outlet data or local retail dynamics, which could provide finer-grained insight into physical access patterns.

Data limitations may also have introduced measurement error particularly in self-reported variables or those derived from administrative estimates. Moreover, while the model's predictive power was moderate, future work could enhance performance through more advanced spatial modeling techniques (e.g., geographically weighted random forests or graph neural networks) and longitudinal datasets to capture changes over time.

Summary of Findings

Overall, the findings highlight that food insecurity in the U.S. is not simply a function of geography but rather an outcome of intersecting economic, infrastructural, and social inequalities. The model confirms that communities with high poverty, limited vehicle ownership, and greater reliance on SNAP benefits are systematically more likely to experience food deserts. These insights provide an empirical foundation for designing evidence-based, equitable, and geographically targeted interventions that move beyond descriptive mapping toward actionable policy guidance.

CONCLUSION

In this research, it was applied the use of three sophisticated machine learning models Logistic Regression, Random Forest and XGBoost to find out the impact of structural, socioeconomic, and demographic factors on food insecurity in the U.S. census tracts land. The tree-based models were very good at predicting the outcomes (AUC ≈ 0.99), thus proving that machine learning can find the complex, nonlinear interactions in huge datasets. Among the major factors of the prevalence of food deserts were the poverty rate, the median family income, vehicle ownership, and SNAP participation. The census tracts that had high poverty, poor transport facilities, and more people depending on assistance were at a risk of food insecurity over three times more than the others. The SHAP analysis showed that the poverty rates over 20% stealthily increase the risk and thus the need for the low-income community targeted interventions. The implications for the policy suggest the opening up of SNAP, better transportation, and the facilitation of mobile food distribution as measures to improve affordability and access. Moreover, it is also possible that the ruling out of the vulnerability may be boosted through local economies being strengthened with job creation and support for small businesses. The cooperation of the government agencies and community organizations is the key to the administration being effective. All in all, this scientific work has proved how interpretable machine learning can be used to develop fair, data-driven food security strategies, thus creating a bridge between the fields of data science and social policy to settle the structural inequalities that are at the root of the food access problem in the United States.

REFERENCES

Almalki, A., Brown, G., & Wong, C. (2021). Geospatial and machine learning regression techniques for food access analysis. *ISPRS International Journal of Geo-Information, 10*(11), 745. https://doi.org/10.3390/ijgi10110745

Amin, M. D., McCluskey, J., & Syed, S. (2020). Predicting access to healthful food retailers with machine learning at the U.S. census tract scale. *Applied Geography, 114*, 102053. https://pmc.ncbi.nlm.nih.gov/articles/PMC7564312/

Beaulac, J., Kristjansson, E., & Cummins, S. (2009). A systematic review of food deserts, 1966–2007. *Preventing Chronic Disease, 6*(3), A105. https://stacks.cdc.gov/view/cdc/20366/

Bitto, E. A., Oakland, M. J., & Sand, M. (2003). Solving the problems of Iowa food deserts: Food insecurity and civic structure. *Rural Sociology, 68*(3), 371–384. https://doi.org/10.1526/003601103322294628

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Chilukuri, J. (2023). *Predicting low food access: An empirical assessment of tree-based machine learning models* [Preprint]. TechRxiv. https://doi.org/10.36227/techrxiv.16820688.v1

Coleman-Jensen, A., Rabbitt, M. P., Gregory, C. A., & Singh, A. (2021). *Household food security in the United States in 2020* (Report No. ERR-298). U.S. Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/webdocs/publications/102076/err-298.pdf

Esmalian, N. A., Coleman, N., Yuan, F., Xiao, X., & Mostafavi, A. (2022). *Characterizing equitable access to grocery stores during disasters using location-based data* [Preprint]. arXiv. https://arxiv.org/abs/2201.00745

Food and Agriculture Organization of the United Nations. (2026). *FAO – Food and Agriculture Organization of the United Nations*. https://www.fao.org

Gundersen, C., & Ziliak, J. P. (2021). Food insecurity and health outcomes. *Health Affairs, 34*(11), 1830–1839. https://pubmed.ncbi.nlm.nih.gov/26526240/

Hu, Y., Yu, Z., & Chen, X. (2022). Integrating predictive analytics with social vulnerability assessment for food insecurity research. *Computers, Environment and Urban Systems, 94*, 101787. https://doi.org/10.1016/j.compenvurbsys.2022.101787

Jiao, J., Moudon, A. V., Ulmer, J., Hurvitz, P. M., & Drewnowski, A. (2012). How to identify food deserts: Measuring physical and economic access to supermarkets in King County, Washington. *American Journal of Public Health, 102*(10), e32–e39. https://doi.org/10.2105/AJPH.2012.300675

Kapoor, S., & Sayer, S. (2024). *From bytes to bites: Using country-specific machine learning models to predict famine and food insecurity* [Preprint]. arXiv. https://arxiv.org/pdf/2409.09980

Kwan, M. P. (2021). Geospatial analysis and data science in public health research. *Annals of GIS, 27*(4), 341–354. https://doi.org/10.1080/19475683.2021.1970269

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). https://arxiv.org/abs/1705.07874

Marmot, M., & Wilkinson, R. (2005). *Social determinants of health* (2nd ed.). Oxford University Press.

Salari, M., Kramer, M. R., Reyna, M. A., Taylor, H. A., & Clifford, G. D. (2023). Combining crowd-sourcing, census data, and public review forums for real-time, high-resolution food desert estimation. *BioMedical Engineering OnLine, 22*, 69. https://doi.org/10.1186/s12938-023-01108-9

Smith, M. D. (2020). Food insecurity, gender, and international migration in low-income countries. *Food Policy, 91*, 101838. https://doi.org/10.1016/j.foodpol.2020.101838

Tango, P. (2023). On the forecastability of food insecurity. *Scientific Reports, 13*, 4309. https://pmc.ncbi.nlm.nih.gov/articles/PMC10038988/

Walker, R. E., Keane, C. R., & Burke, J. G. (2010). Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place, 16*(5), 876–884. https://doi.org/10.1016/j.healthplace.2010.04.013