ISSN: 2029-4913 / eISSN: 2029-4921

Article in press

TECHNOLOGICAL and ECONOMIC DEVELOPMENT of ECONOMY

https://doi.org/10.3846/tede.2026.28754

MACHINE LEARNING-BASED PREDICTION OF FAMILY RETURN TO POVERTY IN THE CRITICAL POVERTY GROUPS

Qing LI [1]✉, Chuang SUN [1], Damien ECHEVIN [2], Yan Jie LI [1], Ying Xue LI [3]

[1]School of Economics and Management, Xinjiang University, Urumqi, China
[2]CRCHUS, University of Sherbrooke, Sherbrooke, Canada
[3]Jinhe Center for Economic Research, Xi'an Jiaotong University, Xi'an, China

Article History:
- received 7 April 2025
- accepted 11 September 2026
- first published online 13 November 2026

Abstract. To consolidate the achievements of poverty alleviation, identifying the likelihood of return to poverty among the critical poverty groups is essential for proactive intervention. This study aims to predict the return to poverty of family in these groups, based on personal, family, and community environmental characteristics. Employing XGBoost and SHAP decomposition on data collected from the China Family Panel Studies between 2016 and 2020, we find that a combination of individual, family, and community environmental characteristics can effectively predict the likelihood of families in the critical poverty groups returning to poverty. The predictive performance of the family characteristics model is better than that of the personal and community environmental characteristics models. Among families in the critical poverty groups, those with higher expenditures on food, culture, education, and recreation are less likely to return to poverty, while those with lower satisfaction with the surrounding environment are prone to such an outcome. Additionally, most characteristics exhibit a nonlinear relationship with the return to poverty. This study broadens the understanding of family return to poverty from a predictive perspective, and provides empirical evidence to inform proactive supervision and policy intervention.

Keywords: critical poverty groups, family return to poverty, prediction, machine learning, XGBoost, SHAP decomposition.

JEL Classification: C43, C61, D81.

✉Corresponding author. E-mail: *qing.li@xju.edu.cn*

1. Introduction

The eradication of global poverty, one of the primary objectives of the United Nations 2030 Sustainable Development Goals (SDGs), presents a particularly significant challenge in developing countries. In response, governments have implemented various measures. For example, South Africa has established social grants to combat poverty and inequality (Fofana et al., 2024). Similarly, Vietnam uses microcredit and educational loans to alleviate poverty (Tran, 2023). In China, the policy framework includes initiatives such as consumer poverty alleviation and the Precision Poverty Alleviation Program, which are designed to help families escape from poverty and achieve common prosperity (Zeng et al., 2023; Zhou et al., 2023).

These policies, especially the Precision Poverty Alleviation Program, have raised rural incomes at county and family levels and improved people's well-being (Li et al., 2025; Zhou et al., 2023). China ended absolute poverty in 2020, and that provides other countries

Copyright © 2026 The Author(s). Published by Vilnius Gediminas Technical University

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

a useful example for their own anti-poverty work (Wu et al., 2023). But in the process of ending poverty, many poor people have become what we call "the critical poverty groups", and their incomes are just over the poverty line. According to poverty vulnerability theory (Morduch, 1994; Gallardo, 2018), families slightly above the poverty line are not stably non-poor but rather "vulnerably poor." They face a high probability of falling back into poverty when exposed to risks such as illness, unemployment, or environmental shocks. This theory thus highlights the need to focus on critical poverty groups and predict their likelihood of returning to poverty.

Poverty is often concentrated regionally in China as in other places in the world, which has been highlighted by the World Bank. Adverse natural conditions, ecological fragility, inadequate infrastructure, and difficult living environments are examples in line with this. While understanding the geographic dimensions of poverty is essential, it is equally critical to investigate why families risk falling back into poverty after being lifted out. To fill the research gap, this study constructs a prediction model that considers three perspectives: personal characteristics, family characteristics, and community environment characteristics. We employ the XGBoost algorithm for predictive modeling to uncover the underlying relationships between the predictor variables and families' likelihood of returning to poverty. Using data from the 2016 to 2020 China Family Panel Studies (CFPS), we identify the important factors affecting the poverty recurrence of families among critical poverty groups and examine the impact of characteristic variables through Shapley Additive exPlanations (SHAP) dependence plots.

In the study of poverty dynamics, chronic poverty and transient poverty are different. Chronic poverty means families stay poor for a long time. Transient poverty means families fluctuate around the poverty line and may return to poverty after shocks (Baulch & Hoddinott, 2000; Carter & Barrett, 2006). Our focus on "critical poverty groups" in China is about the transient poverty type, because these families are just above the poverty line but face a high risk of falling back.

Unlike previous studies, our main contribution is in methodology and content. In terms of methods, since there is no clear agreement on what drives families back into poverty, we develop an XGBoost model for the critical poverty groups, rather than traditional linear regression approaches. We employ SHAP dependence plots to analyze the factors' contributions and how their influence changes. This methodology enables better understanding of the dynamics involved in the return to poverty. We reveal some findings regarding the anti-return factors and provide empirical explanations for these phenomena.

In terms of research content, most previous studies have focused on the impact of poverty alleviation policies or used one-dimensional analyses of why families return to poverty. In contrast, we study the causes of poverty recurrence in the critical poverty groups using machine learning, with personal, family, and community environmental characteristics included. This work expands the current understanding of the factors driving poverty among the critical poverty groups and may help researchers and practitioners in poverty identification and prevention.

2. Literature review

2.1. Poverty dynamics and vulnerability

In the literature on poverty dynamics, a key distinction is between chronic and transient poverty (Baulch & Hoddinott, 2000; Jalan & Ravallion, 2000). Duclos et al. (2010) defined individual chronic poverty as the average level of deprivation experienced over time by a given individual, and individual transient poverty is defined as the difference between deprivation and chronic poverty. Family changes and poverty experiences both shape how poverty evolves (Biewen, 2009; Finnie & Sweetman, 2003; Jenkins et al., 2003). Chronic poverty is more affected by family size and labor market fluctuations, while transient poverty is more susceptible to the impacts of natural disasters (Bayudan-Dacuycuy & Lim, 2013). Chronic poverty is long-term and structural, while transient poverty arises when families temporarily fall below the poverty line due to various shocks. In China, rural poverty used to be mainly chronic (Wan & Zhang, 2013). Over time, it has shifted from chronic poverty to transient poverty (Ward, 2016).

Escaping poverty is not permanent. And poverty vulnerability theory emphasizes the ex-ante risk of falling into poverty (Morduch, 1994; Gallardo, 2018). Even families now above the poverty line may be "vulnerably poor" if they lack the capacity to face the risks such as illness, unemployment, or environmental shocks, among other unexpected difficulties. Earlier studies have measured ex-ante vulnerability using household assets and examined the contribution of shocks to poverty vulnerability (Echevin, 2013, 2014).

Here we integrate insights from poverty dynamics and poverty vulnerability theory. In other words, we take the return-to-poverty issue as both a manifestation of transient poverty and a reflection of family vulnerability. Bearing this in mind, we employ the machine learning approach to provide a novel way to predict vulnerability *ex ante* and to identify, moreover, the factors that drive transient poverty among critical poverty groups.

2.2. Return to poverty – personal, familial, and community determinants

At the personal level, studies show that marital status and gender are significant predictors. For instance, divorce lowers poverty risk for men but raises it for women (Herbst-Debby et al., 2021). Education is a critical buffer against poverty, especially in breaking the intergenerational transmission of disadvantage, and this effect is strongest in rural areas and among women (Zuo et al., 2023). Higher education reduces the impact of poor backgrounds by improving the noncognitive skills of students from low-income families (Xu, 2018). Health is an important concern for vulnerability. Illness and medical expenses increase family financial burdens and reduce labor capacity (Sun & Xu, 2024; Rigg, 2006). Access to social security mechanisms, such as public pension programs, is shown to reduce inequality and provide a stable safety net (Zhang & Imai, 2024). The internal causes of returning to poverty include the inadequate quality of poverty alleviation efforts and low risk tolerance. It is important to conceptualize vulnerability in this context as a state of defenselessness against deprivation, often characterized by dependency and heightened susceptibility to adverse events.

According to the poverty vulnerability theory, families may return to poverty when their resilience is breached by risk shocks. Family characteristics play an important buffering role

in this process, affecting vulnerability level directly. Existing research predominantly views vulnerability to poverty through the framework of the interplay between the resource constraints and expenditure pressure faced by families. For example, poorer families are often forced to make use of the more polluted water sources which they are able to afford (Yang et al., 2013). According to Lin and Wang (2020), energy-poor families are influenced by price signals and restrict electricity consumption below basic needs. However, rising income levels of wealthy families are stimulating demand. Low-income families, unable to meet appliance expenses, resort to the use of less-efficient fuels and equipment that are more harmful to health and cause more serious environmental pollution, thus aggravating the poverty trap (Poblete-Cazenave & Pachauri, 2021).

Two expenditure-related mechanisms are identified in family return to poverty study. Ongoing inflexible expenses do not allow groups to equate their income and expenses, bringing them back into poverty. The more a family spends on education, the less likely they are to be in relative poverty (Jiao et al., 2025). Examples of such expenditures include daily costs for children's education and family housing (Acar et al., 2016). Long-term payments like investment in children's education are long-lasting and unchanging financial commitment. For families located in the critical poverty groups, these costs make them more vulnerable to falling back into poverty (Acar et al., 2016). Conversely, short-term expenditures (e.g., expensive rituals) can impose crippling unexpected burdens, leading to impoverishment (Suzuki, 2025). Although families in the critical poverty groups are marginally above the poverty line and have a higher material living standard, they are economically very vulnerable and highly sensitive to changing conditions. To minimize these risks created by expenditures, the government has launched several fiscal instruments and support policies including targeted grants and subsidies to create and boost their income and reduce poverty (Jung et al., 2015).

The vulnerable condition of critical poverty groups is determined not only by personal and family features, but also by the community environment. This influence is explained by the "neighborhood effects," whereby community characteristics directly or indirectly affect residents' socioeconomic outcomes (Ludwig et al., 2012, 2013). A growing number of studies have explored how community influences residents' well-being and poverty vulnerability, and they find that this influence operates through several channels. The foundational layer consists of physical infrastructure and public services. Investments in basic facilities for daily life, such as water, sanitation, and sewage systems can reduce multidimensional poverty directly (Barbier & Burgess, 2020; Wu et al., 2024; Shi et al., 2025). On the other hand, environmental degradation, through air pollution for instance, can increase the risk (Zang et al., 2024). Moreover, community safety and social infrastructure are equally critical. The need for strong security facilities as well as adequate policing has a bearing on the safety of the residents. The presence of inadequate community security is linked to an increase in violence, which severely impacts family development (Farrell et al., 2017; Alloush & Bloem, 2022). Furthermore, community spaces facilitate positive social interaction and collective action, enhancing residents' sense of belonging and well-being (Enneking et al., 2025; Ludwig et al., 2012, 2013). Positive neighborhood relations and the presence of higher-educated residents can create beneficial demonstration effects. For example, they can impact the educational aspirations of poorer families (Montalvo-Clavijo et al., 2023). In summary, a community environment can impact the dynamics of poverty through its tangible infrastructure, social networks, and composition.

2.3. Our contribution

A substantial body of literature has explored the factors contributing to poverty and return to poverty, as well as the use of machine learning to predict such outcomes (e.g., Li et al., 2022; Barriga-Cabanillas et al., 2025; Echevin et al., 2025; Wang et al., 2025). Previous research on poverty alleviation in China often concentrated on the evaluation of policies (Li et al., 2025; Zhou et al., 2023). This provides a solid theoretical and methodological foundation; however, several gaps remain. Most of these studies focused on single factors, and this limited the integration of family and community dimensions. Additionally, existing analyses relied mainly on traditional econometric models, which focused on parameter identification and causal inference. The existing studies are valuable for understanding mechanisms, but these models are relatively less suitable for predicting complex, nonlinear, and multidimensional poverty dynamics.

Machine learning adds to these traditional approaches. It focuses on accuracy of prediction and can find more complex relationships. Recent methods such as Double Machine Learning (Chernozhukov et al., 2018) and causal forests (Athey et al., 2019), for instance, attempt to connect prediction and causality although they still rely on relatively strong assumptions. Despite the limitations, machine learning has the benefit of quickly processing multi-dimensional feature variables, in order to find complex drivers of the outcome, i.e., return to poverty.

This study proposes an integrated framework to bridge poverty vulnerability theory and machine learning. Vulnerability theory offers a conceptual basis for understanding the risk, through three dimensions namely exposure, sensitivity, and adaptive capacity, while machine learning provides the means to operationalize these dimensions through data-driven prediction. We predict the return to poverty from multiple factors, explain the complex drivers, and clearly identify the households that are prone to return to poverty.

Vulnerability theory guides the feature selection and interpretation. Personal characteristics (education, health, employment) show sensitivity and adaptive capacity. Family characteristics (assets, expenditures, energy use) reflect resource endowment and resilience. Community characteristics (infrastructure, environment, social capital) represent exposure and collective capacity. Using the XGBoost algorithm combined with the SHAP decomposition, we translate the notion of vulnerability into a measurable probability. The SHAP interpretation further validates the theoretical mechanisms, by identifying which of the aforementioned dimensions most strongly influenced the return to poverty.

In summary, combining vulnerability theory with the machine learning framework, this study provides a theoretically coherent and empirically powerful approach, an actionable tool for early warning and targeted policy design to understand and predict return-to-poverty risk among critical poverty groups.

3. Methodology

3.1. Samples and data

The data come from the China Family Panel Studies (CFPS, 2016–2020). This survey is funded by Peking University and the National Natural Science Foundation of China, and it is carried out by the Institute of Social Science Survey at Peking University. We excluded the individuals

younger than 18 years of age and those whose gender is missing. Missing values are handled with multiple imputation (Shah et al., 2014). Logarithmic transformation is applied to the numerical variables to reduce outlier effects. The final dataset consists of 2,476 residents who were in critical poverty groups in 2018.

3.2. Variables

China announced the eradication of absolute poverty by 2020, but relative poverty persists. The national standard is set at 40% of the median disposable income of all residents (Walker & Yang, 2021). Here, the poverty line is defined as 40% of the median per capita family income (Blackburn, 1990). This relative measure is chosen accounting for regional differences in income levels and living costs. In other words, it identifies families that are prone to the risk of returning to poverty, whether they are in lower-income or higher-income regions. Therefore, it enhances the generalizability across different regions, at least within China.

The lower threshold of critical poverty is this relative poverty line, and the upper limit is set at 1.5 times this line. Individuals with per capita incomes in this range are part of the critical poverty groups (Klevens & Luman, 2001). Following the methodology of Zhou and Wang (2019), we use three waves of CFPS data. The sample for this study comprises residents identified as the critical poverty groups in the 2018 survey, defined by having a per capita family income within the critical poverty range for that year. The outcome variable, "return to poverty," is defined using intertemporal poverty status. More precisely, a family is coded as 1 if it was poor in 2016 (income below the 2016 relative poverty line), was a critical poverty group in 2018, and fell back into poverty in 2020 (income below the 2020 relative poverty line). If not, it is coded as 0. This definition strictly follows the trajectory: poverty → critical poverty → return to poverty.

The selection for explanatory variables is based on poverty vulnerability theory and neighborhood effect theory (Morduch, 1994; Zhou & Wang, 2019; Farrell et al., 2017). These characteristic variables are the family conditions in 2020, divided into three dimensions: personal, family, and community environmental characteristics. First, personal characteristics include gender, age, educational attainment, marital status, health status, type of insurance, and employment status. These indicators reflect exposure to risks (e.g., through illness, unemployment) and coping capacity (e.g., through education, insurance), aligning with poverty vulnerability theory. Second, family characteristics encompass expenditure, housing, fuel, and drinking water, which indicate the strength of family resource endowment. Family loans and occurrences of major accidents reflect the family's risk exposure, while government subsidies and social assistance represent external support that enhances family resilience. Additionally, family size reflects both labor supply and consumption pressure. These complex family characteristics contribute to sources of poverty vulnerability. Finally, the core of neighborhood effect theory posits that community environments influence family behavior, either directly or indirectly, thereby affecting factors such as living costs and risk exposure. We incorporate six dimensions of current community satisfaction: public environmental facilities, surroundings, neighborhood security, neighborly relations, neighborly assistance, and attachment to the neighborhood. Among these dimensions, the surrounding security environment is

directly correlated with risk exposure, as improved security reduces risk. Quantifying community satisfaction captures both the actual state of the neighborhood environment, in line with neighborhood effects theory, and the risk exposure of family within the community as described by the poverty vulnerability theory. Detailed variable definitions and descriptive statistics are presented in Table A1 in the Appendix.

3.3. Machine learning model construction

In recent years, machine learning algorithms have demonstrated excellent predictive performance (Shrestha et al., 2021) and have been applied to poverty research (Corral et al., 2025; Aiken et al., 2023). The key to consolidating the achievements of poverty alleviation lies in efficiently identifying the causes of poverty recurrence, which are complex, multidimensional, non-linear, and dynamic, as well as in effectively conducting proactive interventions. The determinants of returning to poverty among the critical poverty groups are complex and likely involve non-linear interactions. Furthermore, our dataset exhibits a severe class imbalance, with the number of families returning to poverty being substantially lower than those that do not. To address these challenges – complexity and imbalance – we employ the XGBoost algorithm, a powerful gradient boosting method based on decision trees (Li & Managi, 2025).

XGBoost maintains a good trade-off between prediction accuracy and computational efficiency. Its advantage and efficacy in predicting poverty dynamics have been recognized by prior research (Li et al., 2022). The empirical strategy here is divided into two main parts. The first part involves the construction of an XGBoost prediction model. The sample dataset was randomly split into a training set and a test set at a 4:1 ratio (Liu et al., 2023). We employed the R package "xgboost" to implement the XGBoost algorithm. To enhance model performance and prevent overfitting, we optimized the hyperparameters through a grid search, which can find the optimal combination of parameters within the search space. This process was combined with cross-validation to improve the model's generalization capability and was implemented using the R package "mlr3". The hyperparameter tuning workflow is illustrated in Figure 1.

First, the machine learning task and learner are defined. Next, the search space for the key hyperparameters was established, including *nrounds*, *max_depth*, *colsample_bytree*, *subsample*, and *min_child_weight* (Zhang & Zhao, 2026). To balance search efficiency and accuracy, the range of each hyperparameter was divided into five equally spaced values, and the search was limited to 50 evaluations.

K-fold cross-validation involves splitting a dataset into k equal-sized folds. In every instance, one fold is chosen as the testing set and the remaining $k - 1$ folds are chosen as the training set. After k iterations, the model's performance is evaluated by aggregating the metrics from each iteration. Through 5-fold cross-validation, we evaluated different parameter combinations by fitting the models iteratively, and identified the optimal parameters. Through the utilization of this method, full use is being made of the sample data. Poverty recurrence is a low probability event. Thus, the model's ability to correctly identify "non-returning-to-poverty" families (i.e., the majority class) is crucial. As a result, specificity is selected as the evaluation metric for tuning since it reflects the model's sensitivity to negative cases and helps reduce the impact of misclassification.

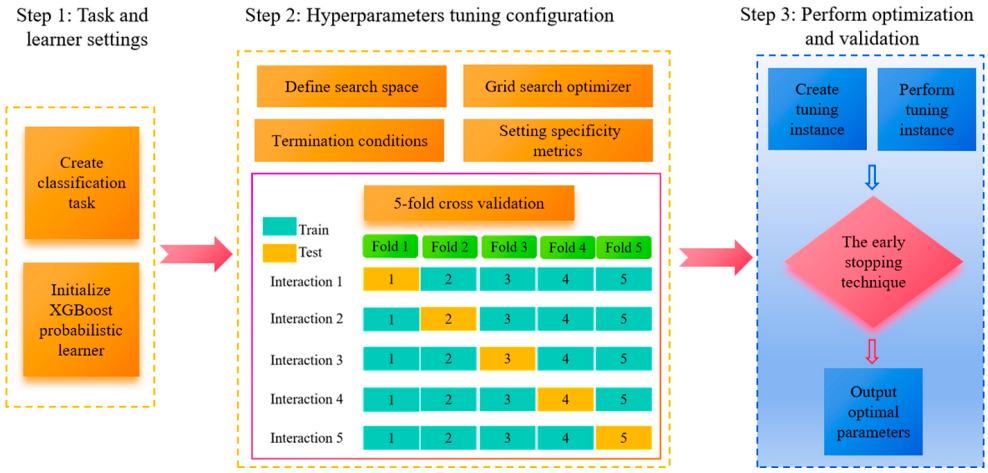


Figure 1. Hyperparameter tuning workflow

The grid search combined with early stopping technique ultimately led to the selection of the best model. The loss function (logloss) for both the training and test sets is monitored during training to prevent overfitting. Early stopping occurs when the loss on the test set no longer decreases but starts to increase, and the iteration at that point is selected. The optimal hyperparameter combination produced using this process is outputted. Refer to Table A2 located in the Appendix for the parameters.

This study examines the three models using the Area Under the Curve (AUC). The score is the probability that each test sample belongs to a positive class, and the AUC value shows the probability that a positive sample is ranked higher than a negative sample based on the model's calculated score value. To ensure a fair comparison when evaluating the contribution of family characteristics to model performance, a fixed threshold of 0.5 was adopted for all models to convert probability outputs into class labels. Subsequently, the accuracy, precision, and recall of models based on personal, community environmental, and family characteristics were compared. To understand these evaluation metrics, we compare the predicted and actual outcomes of family return to poverty by introducing the following key concepts.

TP (True Positive) refers to when a family is predicted to return to poverty, and the true status is indeed returning to poverty.

TN (True Negative) refers to when a family is predicted not to return to poverty, and the true status is also not returning to poverty.

FP (False Positive) refers to when a family is predicted to return to poverty, but the true status is that they have not returned to poverty.

FN (False Negative) refers to when a family is predicted not to return to poverty, but the true status is that they are returning to poverty.

Based on the above concepts, the accuracy rate is defined as the proportion of correctly categorized families. It can be expressed as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{1}$$

Precision is the ratio of correctly identified categorized positive families to the total number of predicted positive families. It is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

Recall is the ratio of correctly identified positive families to the total number of actual positive families. The formula is calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

F1-score represents the harmonic mean of precision and recall. It is calculated as follows:

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

Next, we employ SHAP decomposition. Unlike traditional methods that emphasize only "feature importance", SHAP decomposition provides a more nuanced analysis. It not only ranks the relative importance of features but also utilizes SHAP dependence plots to reveal the average impact of features on the predicted risk. In the SHAP dependence plots, the vertical axis represents the SHAP values of the feature. A positive SHAP value indicates that the feature increases the predicted risk of returning to poverty (i.e., heightens vulnerability), whereas a negative value suggests a protective effect that reduces vulnerability. The horizontal axis displays the actual value of the feature. By analyzing this relationship, we can effectively assess both the direction and magnitude of each feature's influence on the model's output.

Our SHAP analysis was implemented using the R package "shapviz".

$$y_m = y_0 + \sum_{n=1}^{N} g\left(x_{mn}\right), \tag{5}$$

where x_{mn} denotes the n characteristic variable of the m sample, y_m represents the predicted value for the m sample, and y_0 refers to the baseline value, $g\left(x_{mn}\right)$ denotes the Shapley value of the characteristic x_{mn}. The Shapley value for a feature is calculated as follows:

$$g\left(x_{mn}\right) = \sum_{S} \frac{\left|S\right|!\left(\left|F\right| - \left|S\right| - 1\right)!}{F!} \times \left(f_x\left(S \cup \left\{x_n\right\}\right) - f_x\left(S\right)\right),$$
$$S \subseteq \left\{x_1, x_2, \cdots, x_N\right\} / x_n, \tag{6}$$

where F denotes the entire feature set, S denotes the feature subset excluding x_n, $f_x\left(S\right)$ denotes the predicted value of the XGBoost model under subset S, $f_x\left(S \cup \left\{x_n\right\}\right) - f_x\left(S\right)$ denotes the marginal contribution of the characteristic x_n, and $\frac{\left|S\right|!\left(\left|F\right| - \left|S\right| - 1\right)!}{F!}$ denotes the equal treatment of each class subset. $\left|g\left(x_{mn}\right)\right|$ is proportional to y_m, with positive or negative values indicating the enhancement or weakening of the feature x_{mn} variable on the prediction outcome. The introduction of SHAP facilitates a deeper exploration of the predictive trends of feature variables concerning family return to poverty among the critical poverty groups, rather than merely assessing model performance.

4. Empirical results

4.1. Analysis of model output results

The training results concerning personal characteristics, community environmental characteristics, and family characteristics for the sample are derived from XGBoost. Figure 2 illustrates the model's prediction results. The model based on personal characteristics alone demonstrates low performance, with an AUC of only 0.535. In contrast, the community environmental characteristics model achieves an AUC of 0.752, representing an improvement of over 10% in key metrics compared to the personal characteristics model. This suggests that community factors capture residents' subjective attitudes toward public facilities, living surroundings, and support from neighbors, serve as effective proxies for objective living environments, and significantly enhance the model's predictive capability for poverty recurrence.

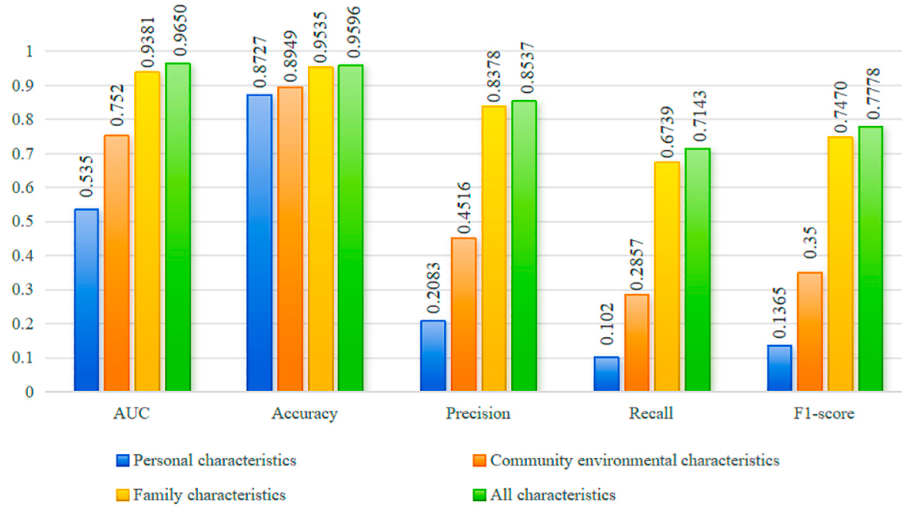


Figure 2. Bar chart comparing XGBoost model evaluation metrics across different characteristics categories

The most powerful predictors are found at the family level. The family characteristics model yields an AUC exceeding 0.90, and incorporating all characteristics performs even better, indicating that both the family characteristics model and the all characteristics model have strong classification abilities. The strong predictive power of family-related factors is economically intuitive: families spend the majority of their time within their families, and family-related factors significantly influence residents' daily lives, including the type of water, fuel, and food consumed daily. By integrating all characteristic variables, the overall performance of the model for predicting families' return to poverty is enhanced.

4.2. Characteristic importance analysis

To explore the contribution of features to the model predictions, we employ SHAP dependence plots based on the XGBoost model developed to analyze a family's return to poverty. The Shapley value of each feature was calculated, as shown in Figure 3. Features are ranked

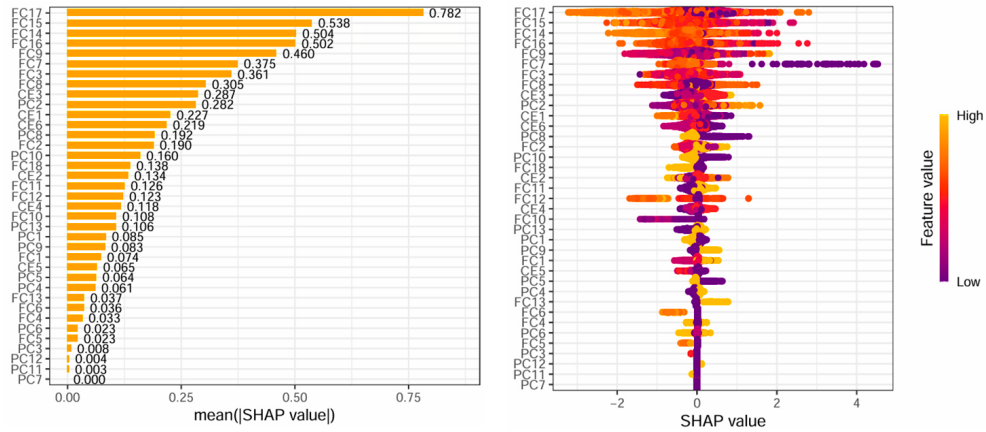


Figure 3. Importance ranking of feature variables and beeswarm plot of model (XGBoost) outputs

by the mean absolute SHAP values, with the higher-ranked features indicating relative importance. To illustrate the contribution of the sample points, we present a beeswarm plot derived from the SHAP summary plot, where each line represents a feature variable, and each point corresponds to a sample point. The position along the y-axis in the figure is determined by the ranking of feature variable importance, while the x-axis reflects the Shapley value.

The positive and negative contributions of the sign variable correspond to the different colors. A more yellow color indicates greater contribution, while a more purple signifies less contribution.

As illustrated in Figure 3, family characteristics exert the most significant influence, with several variables above 0.3: Expenditures on culture, education and recreation (FC17), expenditures on family equipment and daily necessities (FC15), food expenditures (FC14), expenditures on transportation and communication (FC16), current value of the house property (FC9), family size (FC7), monthly electricity costs (FC3), and area of current housing (FC8). This influence arises from the direct impact of family behavior, primarily through four channels:

Firstly, families enhance human capital through educational investment (FC17), thereby interrupting intergenerational poverty transmission (Chen et al., 2024) and consolidating poverty alleviation outcomes. Secondly, housing (FC9 and FC8) represents "family wealth stock." As significant fixed assets, properties can mitigate sudden risks (such as illness or unemployment) through mortgaging or rental income, while vehicles enhance mobility and reduce living costs. Family assets support their long-term resilience to risks, with FC9 being the most critical factor. Thirdly, the family expenditure structure, scale, and abnormal spending directly reflect basic living conditions. Essential expenditures like food (FC14) and family equipment/daily necessities (FC15) signal short-term survival pressures. Fourthly, essential living arrangements, such as electricity expenditure levels (FC3), indicate risk exposure. Families with reasonable and stable electricity spending typically possess greater "economic buffer capacity," demonstrating stronger stability in poverty alleviation.

The neighborhood security environment (CE3) among the community environment characteristics plays a more significant role, as the public security environment directly reflects the

level of risk exposure faced by residents within their communities. The contribution value of age (PC2) within personal characteristics is also notably important, as families are vulnerable to returning to poverty at various life stages: inexperience in youth, the pressures of midlife crisis in middle age, and elevated health risks in old age. These contributions of factors exceed 0.28.

In addition, there are a total of 22 characteristic variables across the three dimensions with contribution values greater than 0.1, while three variables, such as membership (PC12), contribute less than 0.005.

The SHAP outcomes show that family variables are most important, which means the resources and living conditions of families play a major role in resisting poverty. Factors related to the community and the personal also play a meaningful role. There is a reflection of contextual risk and personal characteristics on these factors. Generally, the model captures a multidimensional structure of vulnerability consistent with theoretical expectations.

4.3. Characteristic prediction model analysis

4.3.1. Personal characteristics and family return to poverty

This section provides an analysis of how personal characteristics shape the trend of family return to poverty among critical poverty groups, as illustrated by SHAP dependence plots in Figure 4. The analysis employs a factor-by-factor profiling method, based on the importance ranking of personal characteristics in the comprehensive characteristics model, and the SHAP dependence plots of personal characteristics presented in Figure 4.

The nonlinear trend in the SHAP dependence plot is first discussed. PC2 (Age) is the most important personal characteristic. Its SHAP dependence plot shows a fluctuating, nonlinear relationship with poverty recurrence risk. Those younger in age (with SHAP values > 0) are more vulnerable due to three career challenges. The first of these is that mandatory internships for graduates (such as law and medicine) tend to delay entry into a formal job. Second, to remain classified as graduates, some graduates engage in informal employment by not signing formal contracts and social insurance. This results in erratic earnings and limited access to social safety nets, greatly increasing their risk exposure. Thirdly, mismatch between vocational training and job requirements makes the skills of many graduates useless. Therefore, these deficiencies compounded by lack of formal experience result in these low-wage employment enhancing the risk of poverty. As people reach the age of thirty, their job situation gets fixed, thereby preventing them from falling into poverty.

Nonetheless, an additional peak in risk arises around the age of 35 (SHAP > 0). This effect is mainly attributed to the interaction between skill obsolescence and cost pressure. According to evidence from a large technology firm's internal analysis, the total labor cost of employees born after 1995 is 28–32% lower than that of employees aged 35 and above, even after factoring in a three-year training cycle. A more youthful employee crowd can absorb new technologies better, displacing older employees (Acemoglu & Restrepo, 2020). The hiring criterion "under 35" prevalent in job ads has resulted in statistical discrimination. Unemployed people aged over 35 find it nearly impossible to get re-employed. It erodes their resilience to the economic shocks significantly. In contrast, the people aged from 40 to 58 show negative

SHAP values. Accumulated wealth and job stability at this stage of their career contribute to this lower risk period. Around age 58, however, the protective effect reverses as SHAP values become positive and begin increasing. The rise in risk is possibly driven by age-related health problems and the associated financial burden of medical expenses (Baird, 2016).

PC8 (Health status) is a powerful predictor of poverty recurrence. Individuals in poor health are at high risk (SHAP > 0) according to the plot. Poor health may decrease earning capacity, and drive up catastrophic health expenditures, creating a direct pathway into poverty (Sun & Xu, 2024).

Furthermore, we discuss the linear trends. When the SHAP value for PC10 (Working status) is positive, the corresponding samples are mostly located in category 0 (unemployed). This indicates that unemployment significantly increases the risk of returning to poverty (Li et al., 2021).

PC13 (Urban-rural classification) reveals a distinct dimension to poverty recurrence. The risk is mainly concentrated in rural areas (category 2), where families face a systematically higher likelihood of falling back into poverty.

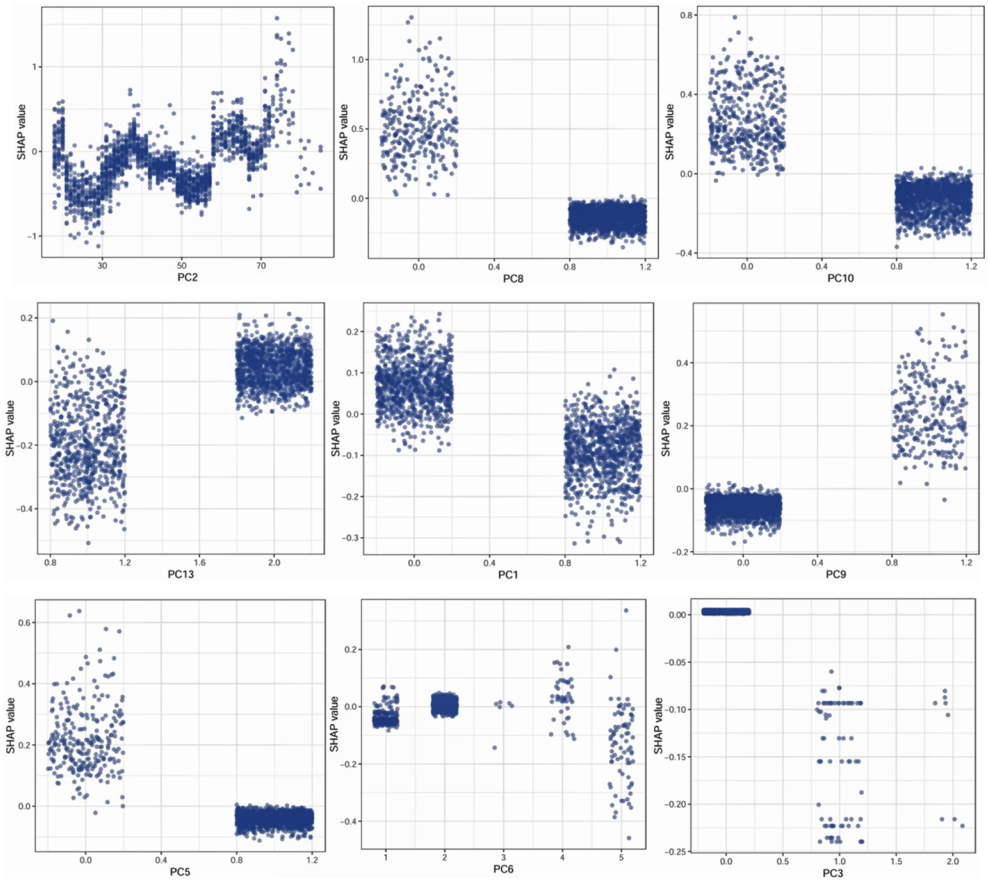


Figure 4. SHAP dependence plots for personal characteristics

PC1 (Gender): Samples with positive SHAP values are predominantly female (category 0), signifying that women in the critical poverty groups are systematically at a higher risk of returning to poverty.

PC9 (Chronic disease) diagnosis increases the chance of falling back to poverty as it has a strong association with positive SHAP value. The long-term medication and care can drain finances over a period of time. In addition, the risk of deterioration of the condition may incur catastrophic health care costs. When people face these two challenges simultaneously, they have much less financial resilience, which makes it more likely that they will fall back into poverty.

PC5 (Medical insurance): Samples with very high risk (SHAP > 0) are largely uninsured. Medical insurance is a risk-sharing mechanism for treatment costs. If those without insurance have to face the full shock of medical bills, it will substantially increase their family risk of returning to poverty.

PC6 (Marital status): Unmarried individuals (category 1) show a reduced poverty risk (SHAP < 0). In contrast, married people (category 2) have a higher risk (SHAP > 0). Marriage may bring heavier financial burdens, such as housing and transportation debt, and also ongoing costs for children's education and daily consumption. These fixed costs put pressure on the family's budget, increasing its vulnerability to income shocks and the risk of falling back into poverty.

PC3 (Education level): Negative SHAP (conveys lower risk) consistently occurs in people with either a college or graduate degree. Individuals who possess some degrees of education in their families tend to use this human capital to access higher-quality employment and social resources, creating a buffer against poverty. In contrast, those with less education lack this buffer and remain more susceptible to economic shocks (Xu, 2018).

Overall, these SHAP patterns confirm the sensitivity dimension of vulnerability theory: health shocks, unemployment, and limited education heighten exposure to risks and reduce their ability to maintain well-being under stress.

Since the mean absolute SHAP values for the four characteristics (PC4, PC7, PC11, PC12) are relatively small and show no obvious trend, we will not focus on profiling these characteristics for the time being.

4.3.2. Family characteristics and family return to poverty

In the model incorporating all characteristics, family characteristics are important in predicting the family return to poverty in the critical poverty groups. This is analysed in depth based on the order of importance and in conjunction with the dependence plots of family characteristics (Figure 5).

We first discuss nonlinear trends, beginning with three characteristics associated with housing assets:

First, FC9 (Current value of the house property) acts as a proxy for family wealth and shows the significant contribution to the return to poverty. The sample points cluster into two segments at a threshold of $x = 0.63$ (approximately ¥33,000). Below this threshold, the relationship follows a U-shaped trajectory. Within this price bracket, most families are simple families. Cheaper dwellings may be remote or of bad quality. This generates high "convenience

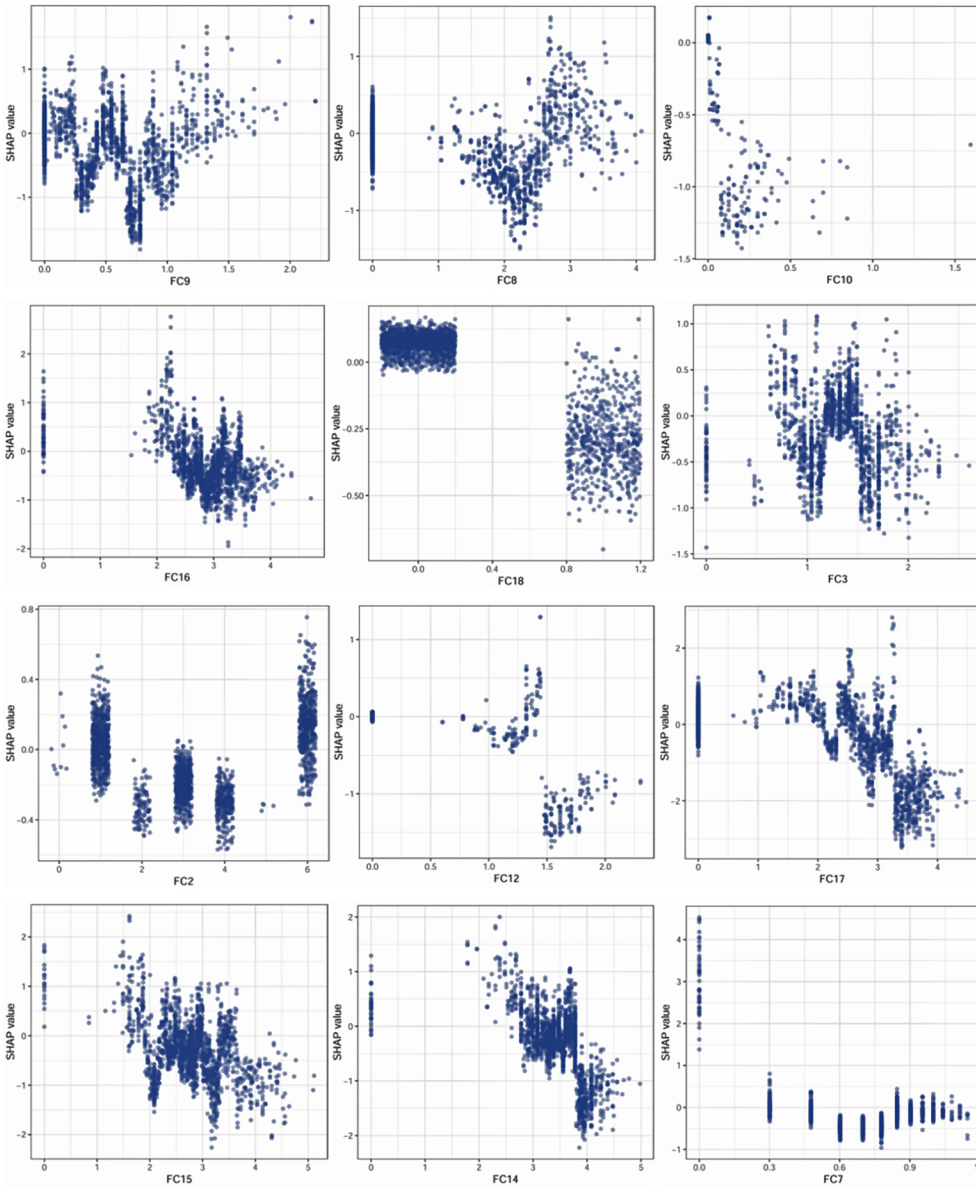


Figure 5. SHAP dependence plots of family characteristics

costs" that can trigger a return to poverty. Homes that cost a reasonable amount may give ordinary families a huge mortgage. This will create a serious risk of indebtedness. Above ¥33,000, a scattered upward trend indicates that higher property values are associated with larger family debts. However, differences in 'income stability' across families lead to significant fluctuations in poverty risk.

Second, FC8 (Area of current housing) indicates that many sample points for this variable have SHAP values greater than zero especially for smaller areas. When the area becomes

larger, most points fall below zero. But after a certain threshold, more points have positive SHAP values. Living in small dwellings is often an indicator of poor living conditions with a lack of economic foundation. Those living in small dwellings are highly vulnerable to sudden shocks. Medium housing satisfies basic needs, has a sufficient economic capacity, thereby reducing poverty risk. Conversely, very large dwellings may impose a heavy cost that crowds out other expenses.

Third, FC10 (Expenditures on mortgage repayment) indicates that greater loan amounts are associated with lower poverty risk. A smaller mortgage might suggest a more sizeable down payment that reduces liquid savings. This may impair the family's ability to cope with major unforeseen expenses like major medical costs, and ultimately diminish its risk-bearing capacity.

In addition, motor vehicles can reflect family wealth. Since FC16 (Transportation and communication expenditures) and FC18 (Car ownership) are strongly correlated, the two characteristics are analysed together. The SHAP values for FC16 show a generally U-shaped trend across sample points. When FC16 is low, positively valued samples are more common, as revealed by cross-analysis with FC18. These high-risk samples are predominantly from families without cars. These families are often located in "transportation-deprived areas." This group lacks private transport, which incurs high commuting costs, probably crowding out priority spending on essential goods such as food and healthcare.

The model identifies family energy conditions as key signals of underlying vulnerability. FC3 (Monthly electricity costs): Negligible electricity expenditures suggest energy poverty, a manifestation of broader financial constraints. This implies these families are probably going without basic appliances, likely undermining their quality of life and buffer against shocks. FC2 (Cooking fuel) indicates the position of the family on the energy ladder in terms of both sociocultural level and economic status. Relying on polluting solid fuels indicates economic and spatial marginalization. The order of risk is clear from high (solid fuels) to low (clean fuels); cooking fuel use, accordingly, is a robust proxy for the bundle of constraints faced by a family.

Low values of FC12 (Total government grants) are associated with SHAP values that are greater than zero. This is suggestive of insufficient support capacity and weak risk resilience. Once values become 1.5, the SHAP values are mostly less than zero. This suggests that beyond this subsidy threshold, government assistance effectively aids families in weathering risks.

The following discussion is concerned with the linear trend of FC17 (Expenditures on culture, education, and recreation). FC17 is the most effective predictor of poverty recurrence among critical poverty groups, and it shows a negative correlation with the outcome. Families who spend less on this have positive SHAP values, meaning they are more likely to return to poverty. Many families may not have this surplus income available for investment in education, culture, and recreation. Consequently, they miss out on the poverty-preventing benefits that such expenditures typically provide. When FC17 reaches about 2.5 (¥317), many sample points display negative SHAP values, suggesting that higher expenditures help prevent poverty and reduce the risk of falling back.

FC15 (Expenditures on family equipment and daily necessities). There is a negative correlation between SHAP values and FC15. The scatter plot shows that when the feature value

is near 2 (approximately ¥100), the SHAP values change from positive to negative. Low expenditure on these items is a reliable indicator of a constrained quality of life, which predicts a higher vulnerability to falling back into poverty.

FC14 (Food expenditures). The SHAP dependence plot exhibits a negative correlation where a lower logged per capita food spending value of approximately between 1.7 (¥50) and 2.7 (¥500) relates to positive SHAP values. A drop in food spending is a significant positive predictor of return to poverty. In economic terms, food is a necessity. Less spending on food is a strong direct indicator of deprivation and tense family finances. It objectively indicates that a family is living at the margin, with limited capacity to absorb other shocks, making it more likely to fall back into poverty.

FC7 (Family size) negatively correlates with SHAP values overall. Samples with SHAP values above zero predominantly fall below 0.6 (approximately three persons). They have higher chances of returning to poverty. A smaller family size indicates less availability of labor resources which can reduce income streams and lead to families returning to poverty in the critical poverty groups. The SHAP values go beyond zero for FC7 = 0.9 (seven persons). As family size increases, the rigid expenditure on education, healthcare, housing, and daily consumption also increases, which heightens the risk of returning to poverty.

The high predictive power of family-level features is an expression of adaptive capacity dimension of vulnerability. The assets, energy use, and expenditure composition reveal families' resilience or fragility in dealing with shocks. This supports the theoretical view that limited adaptive capacity drives vulnerability among critical poverty groups.

4.3.3. Community environmental characteristics and family return to poverty

The community environmental characteristics consist of subjective assessments – both positive and negative – that families make about their environment. These assessments are rated on a scale from 1 to 5 (very good, good, fair, poor, very poor). The SHAP dependence plot (Figure 6) was used to interpret this assessment.

The community environmental characteristics analysis seems to have a complex role in poverty recurrence through the lens of the neighborhood effects.

CE3 (Neighborhood security): A deteriorating security environment (CE3) is positively correlated with a higher risk of poverty recurrence. When the perceived security level changes from "fair" toward "poor," the proportion of high-risk sample points (SHAP > 0) increases, and this pattern is in line with earlier studies (Farrell et al., 2017).

Concerning the relationship between CE1 (Public environmental facilities) and poverty risk, contrary to what one would expect, families whose ratings of the facilities were good or very good face positive SHAP values. This may indicate a cost-of-living paradox. More public infrastructure because of China's poverty alleviation strategy leads to higher concomitant costs (e.g., property management fees, mortgages). Families that are financially fragile may find themselves unable to spend on other essentials due to high fixed spending, and that is a new route to precariousness.

CE6 (Affection for the neighborhood): People with strong place attachment (category 1) have a higher risk of returning to poverty. As social networks are deep-rooted, it provides emotional support so that people don't go away for labor migration towards better economic opportunities. Moreover, it also restrains them from acquiring new skills.

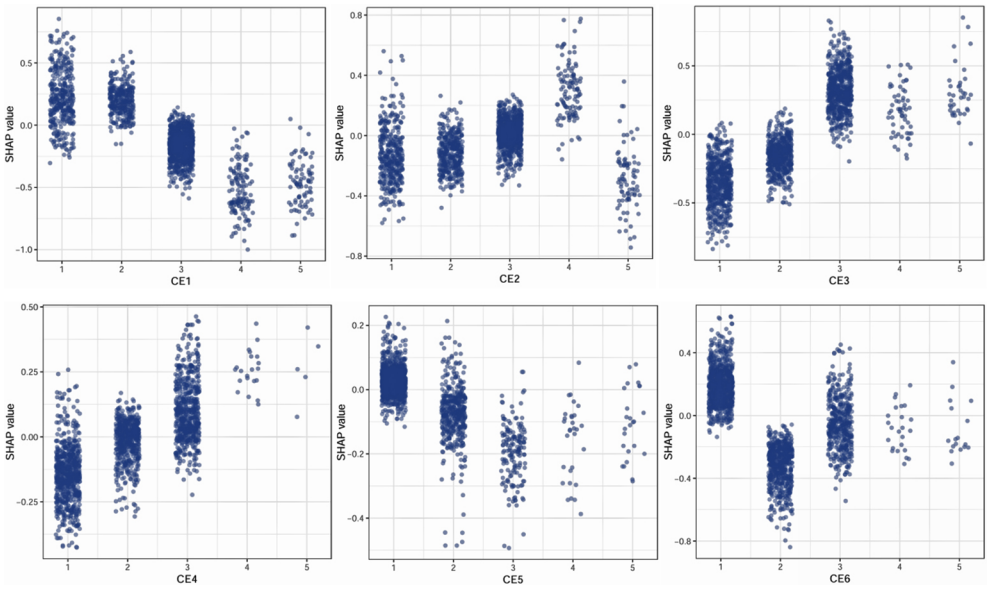


Figure 6. SHAP dependence plots of community environmental characteristics

CE4 (Neighborly relations) has a generally positive correlation between its SHAP values and sample points. Sample points that are characterized by average or poor neighborhood relationships mostly correspond to SHAP > 0. Neighbors with harmonious relations support and share resources with each other, which facilitates the stabilization of living conditions (Enneking et al., 2025). Alienation of neighbors would increase their likelihood of returning to poverty.

CE2 (Surroundings): A negative surrounding environment characterized by issues like noise pollution and garbage is viewed as being a positive predictor of return to poverty. These unsuitable physical conditions may erode the quality of life and severely undermine living stability of families in the critical poverty groups (Wu et al., 2024).

Since the CE5 lacks distinctive features, we will omit it for now.

These findings emphasize the exposure component of vulnerability: communities with weak infrastructure, insecurity, or environmental damage expose their residents to higher external risks. The community environmental conditions shape vulnerability in addition to the individual and family conditions, as confirmed by the SHAP interpretation.

5. Robustness

5.1. Replacement of machine learning algorithms

In a performance comparison, we compared XGBoost with Deep Learning, Random Forest, Naive Bayes, AdaBoost, and Generalized Linear Model (GLM). The first objective is to check whether the performance of the XGBoost algorithm is better than the performance of other machine learning algorithms. The second objective is to check whether the feature variables

can effectively predict the response variable. The evaluation was undertaken under the same training/testing conditions and metrics. The assessment outcomes of the relevant indicators for the integrated feature predictive performance of each algorithm are shown in Figure 7.

The comparative results show that XGBoost predicts the poverty recurrence better when the full features are present. The effectiveness of the model is due to its specific strength in modeling complex, nonlinear relationships. Furthermore, the other algorithms, Random Forest and Deep Learning also perform very well, indicating the chosen feature variables provide great prediction power. Our study demonstrates robustness by these findings being consistent across various models.

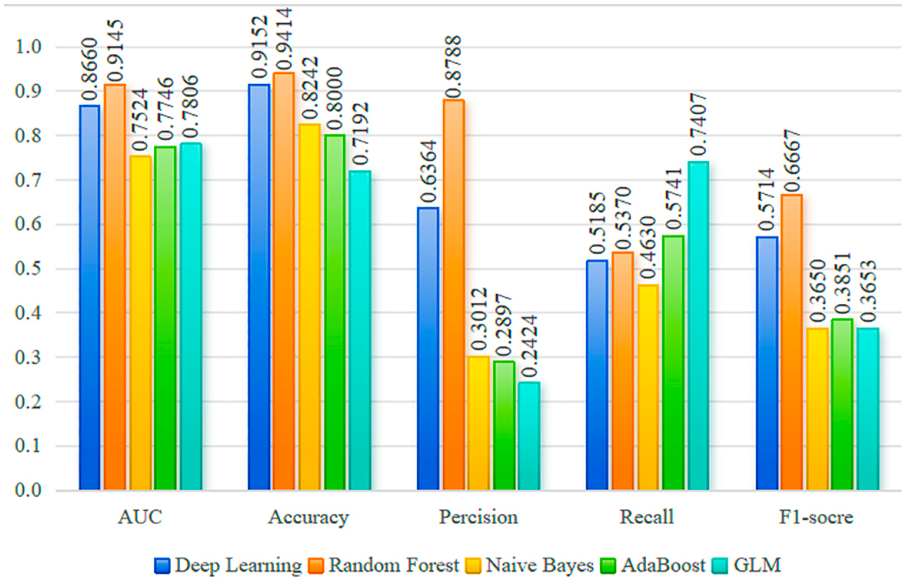


Figure 7. Different algorithms and evaluation metrics

5.2. Undersampling

In the dataset, the families who did not return to poverty outnumber those who did (in the critical poverty groups) by a ratio of 8.8 to 1. Thus, we see significant class imbalance in the response variable. To address this imbalance, an undersampling technique was employed with the aim of achieving a balanced dataset that had a ratio of 1:1 for robustness testing (Echevin et al., 2025). This moderation process assists the model in more efficiently training minority classes and prevents the prediction results from being biased toward the majority class due to imbalanced data.

As shown in undersampling results (Figure 8), the family and comprehensive models achieve the best performance within the balanced dataset. However, the overall performance benchmark set by the original, imbalanced data remains unsurpassed. The undersampling process is likely to cause substantial information loss, resulting in the performance gap. Ultimately, family characteristics are the strongest predictors, regardless of how the sample has been selected.

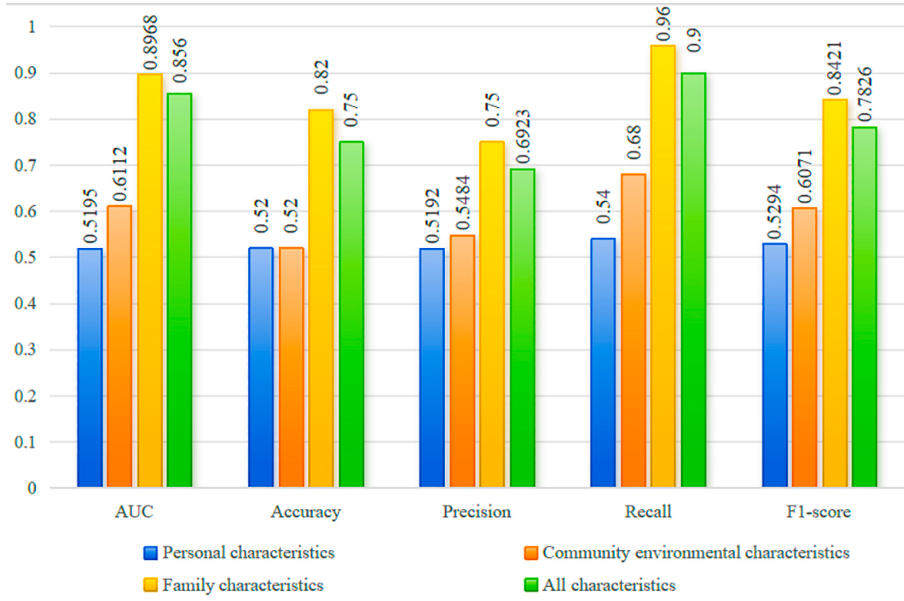


Figure 8. Evaluation metrics for undersampling methods

5.3. Incorporating the variables of the COVID-19 pandemic

The CFPS 2020 survey was conducted from July to December 2020. This survey time closely coincides with the COVID-19 pandemic. This likely influenced poverty recurrence. During this period, quarantine measures at home disrupted employment. They also heightened domestic tensions as families were confined together for extended periods. In the XGBoost, we added variables on pandemic-related employment and domestic conflicts to capture these effects. The model incorporates eight pandemic-specific variables, namely COVID1 (Working hours), COVID2 (Working arrangements), COVID3 (Work income), COVID4 (Monthly income fluctuation), COVID5 (Unemployment probability), along with measures of domestic conflict frequency: COVID6 (Spousal), COVID7 (Child-related), and COVID8 (Parental). The appendix tables give detailed variable descriptions.

After adding the pandemic variables, it performed strongly with an AUC of 0.9623, Accuracy of 0.9515, Precision of 0.8571, Recall of 0.6122, F1-score of 0.7143 similar to the baseline comprehensive model. Nonetheless, as shown in Figure 9, the only pandemic variable appearing in the top 18 predictors is COVID1 (Working hours). The contributions of all other pandemic-related variables were marginal.

The increased poverty vulnerability is shown in the purple section of the SHAP plot with respect to COVID1, which means reduced working hours (COVID1 = 1). However, the widespread adoption of remote work helped families sustain essential income and lessened the overall damage of the pandemic. As a result, the model results are robust to the addition of these pandemic-related variables.

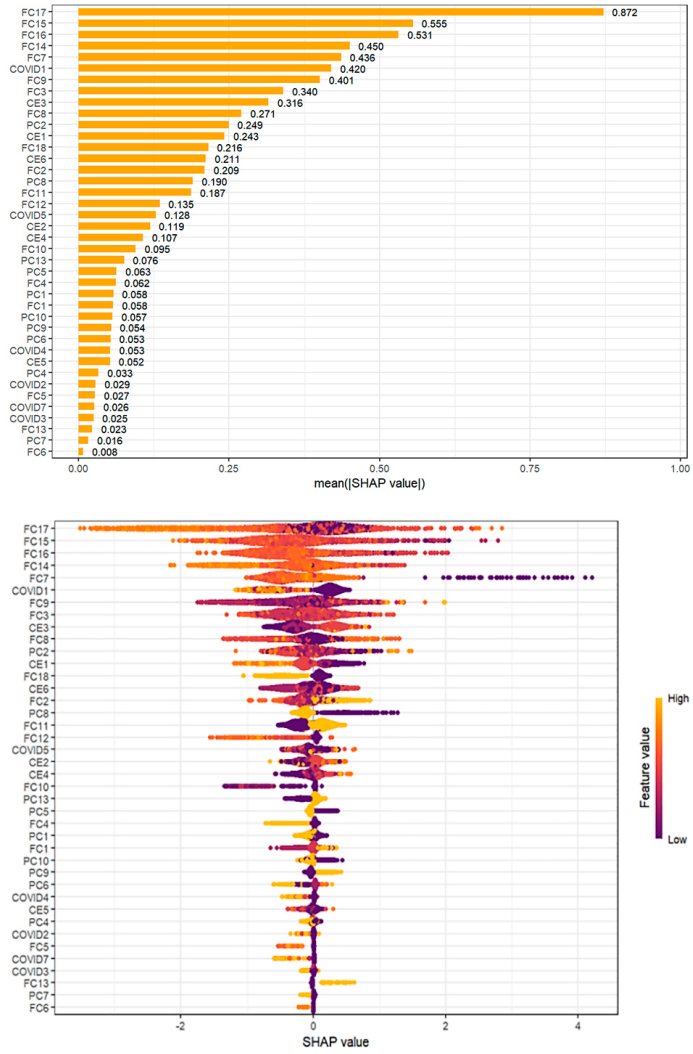


Figure 9. Feature importance ranking and honeycomb plot generated by the comprehensive epidemic model (XGBoost)

5.4. Replacement of core parameters

Machine learning models can be sensitive to hyperparameter settings. This paper modifies two key parameters of XGBoost, namely a learning rate and maximum depth, to observe their impact on the prediction results and evaluation of model's robustness. In this research, the initial learning rate is taken to be 0.2275 and the initial max_depth is taken to 5. A reasonable range of parameter fluctuations with it is defined to test robustness. Three learning rates (0.22, 0.2275, and 0.235) and three max_depth values (4, 5, and 6) are used. This results in nine new parameter combinations (including the original settings). The AUC, accuracy, precision, recall, and F1 score of the prediction results are then evaluated to assess the model's robustness.

From the Table A4 in the Appendix, it can be seen that among the nine combinations of adjusted parameters, the AUC value is generally above 0.9, with some slight downward trends, but all combinations exhibit strong predictive performance. The accuracy remains above 0.9, indicating high precision; the overall precision is above 0.79, reflecting good model performance; and both recall and F1 score have also shown improvements. Therefore, the model can be considered robust.

5.5. Examining the robustness of the poverty-return definition across time

Our baseline model defines a family as having "returned to poverty" (assigned 1) if it experienced the sequence "poverty, non-poverty, poverty" while being part of the critical poverty group; all other cases are assigned 0. Recognizing that the "0" category may encompass complex scenarios, we conducted a series of robustness tests using alternative, temporally specific definitions of poverty return.

Scenario 1: We considered families that may have exited poverty before 2016 but returned in 2016. Specifically, critical poor families following the sequence "poverty in 2012, critical poverty in 2014, and return to poverty in 2016" were assigned 1, with others as 0. The model performance under this definition remained excellent; Scenario 2: We examined cases where a family was in the critical poverty groups in 2016 and 2018 but was poor in 2014 and again in 2020 – a pattern indicating a return to poverty. The results under this definition were consistent with our basic findings; Scenario 3: Additionally, we tested a definition capturing returns to poverty by 2022, i.e., families that were poor in 2016, in the critical poverty groups in 2018 and 2020, and in poverty again by 2022. The results were still stable and consistent with the baseline. Scenario 4: After removing sustained non-poverty samples, the performance experienced a slight decline but remains robust. This indicates that these specific groups have little impact on our study.

In the available temporal definitions, the model shows consistent and robust performance. Please see Appendix Table A5 for specific results.

6. Conclusions and policy implications

6.1. Conclusions

Previous studies often focused on the impact of isolated factors, overlooking how the complex effects of multiple factors influence the risk of returning to poverty among the critical poverty groups. To address this gap, we developed a predictive model using the CFPS data from 2016 to 2020. The SHAP dependence plots were further employed to analyze the trending impact of these key factors.

Our results demonstrate that the XGBoost algorithm utilizing family and community environmental characteristics exhibits significant predictive power for return to poverty among critical poverty groups. The comprehensive model (with all characteristics) shows an improvement over the family-characteristics-only model, with a 2.69% increase in AUC, a 0.61% rise in accuracy, and a 1.59% enhancement in precision, a 4.04% rise in recall. After applying undersampling to balance the classes, the recall rate improves. The model developed in this study demonstrates high prediction accuracy and strong data processing capacity.

Overall predictions indicate that family characteristics exert a stronger influence on return to poverty than either personal or community environmental characteristics. At the family level, the model predicts a lower tendency for families in critical poverty groups to return to poverty when the following conditions are present: greater use of modern cooking fuels, higher food expenditure, larger family size, increased spending on culture-education-recreation and family equipment-daily necessities, and car ownership. In terms of personal characteristics, a higher risk of returning to poverty is predicted among individuals with poor health, chronic diseases, unemployment, lack of insurance, lower educational attainment, and being unmarried. With respect to community environmental characteristics, the model indicates that better public security, improved surrounding environments and better neighborhood relationships are associated with a lower probability of families in critical poverty groups returning to poverty.

Additionally, the analysis reveals predominantly nonlinear associations between most characteristics – whether personal, family, or community-related – and return to poverty among families in critical poverty groups. This finding aligns with the theoretical and empirical literature. The SHAP dependence plots further illustrate that only a small number of traits exhibit clear monotonic relationships with the return to poverty. For most variables, the relationship is more complex, supporting the view that the return to poverty results from complex interplay of multiple factors.

6.2. Policy implications

6.2.1. Government level

The government could establish an integrated poverty recurrence prevention system, comprising three pillars: a machine-learning-driven early-warning platform, a data-informed targeted assistance, and a skills-credit linkage plan.

1. Early warning & monitoring platform

A privacy-protected early warning system for high-risk groups can be developed using machine learning, and it helps identify families and root causes of poverty recurrence more accurately. This could collect data via integrated online-offline channels to address current data gaps and dynamically monitor critical groups. First, the system assesses household resources like expenditure, housing, water access to determine their risk resilience. Then it examines risk shock factors such as health, insurance, and employment. For example, the platform may apply warning rules based on employment status, health insurance coverage, and property value to enable timely risk detection, cause analysis as well as proactive interventions.

2. Data-informed targeted assistance

Low household consumption expenditure is a strong predictor of poverty recurrence. The government could provide subsidies to households whose consumption expenditure falls below specific thresholds for categories such as FC17 (education, culture, and recreation), FC15 (equipment and daily necessities), and FC14 (food). These measures help families promote their human capital and social inclusion while ensuring basic subsistence. The health and

employment departments could collaborate to establish a "Health-Employment File" to share chronic disease management information, ensure treatment access, and create flexible work opportunities (e.g., remote work) for individuals with mild chronic conditions.

3. Skills-credit linkage plan

This enhances the families' capability by providing flexible and accessible skills training (e.g., online training and in-person workshops). Its core innovation links credit access to business or employment plans. Approved plans could then receive preferential loans with favorable conditions, transforming passive aid into self-sustaining economic support.

6.2.2. Community level

Communities act as a critical link between the government and families:

1. Integrated environment, security and family assistance

Public safety management and family dynamic tracking can be combined to build a support network for communities. These efforts can be complemented by voluntary home visits. During home visits, teams may assess fuel use, help with applications for clean energy subsidies, and work with community security personnel to prevent potential safety hazards.

2. Health-employment-education stations

By working with local health departments, communities provide free physical examinations and personalized medication guidance for residents with poor health, design targeted training courses and partner with enterprises to connect job seekers and employers. Additionally, during winter and summer vacations, they can arrange college student volunteers to offer tutoring classes for primary and secondary students, alleviating the families' financial pressure.

3. Tech integration for precision support

Guide community workers to leverage the proposed early warning platform to identify at-risk groups, then cross-verify the results with first-hand data from home visits. Encourage regular participation in technical training programs, thereby maximizing the impact of tech-driven poverty alleviation at the community level.

6.2.3. Long-term societal effects and mitigation of unintended consequences

Privacy breaches and algorithmic stigmatization. Collecting and integrating sensitive household data (expenditure, health, employment, etc.) creates risks of data leakage and unauthorized use. To address this issue, laws must be strengthened to support poverty prevention, and data governance should receive more attention. Other measures may include data anonymization, household consent, and usage of early-warning outputs only inside the system.

Welfare dependency and self-motivation. Long-term or poorly designed aid may create a passive mindset, where the families may become dependent on aid. To reduce this risk, a "Model Family" selection process can be established to officially recognize households who are out of poverty through their own efforts. This initiative is intended to cultivate endogenous motivation, reduce dependency, and build a community culture of diligence and mutual support.

7. Limitations and prospects

First, the CFPS questionnaire lacks items on macroeconomic conditions. This makes it difficult to account for the impact of macroeconomic factors. Also, the sample may not be fully representative of the national population, and issues such as sample additions and panel attrition may compromise the integrity of cross-period data matching. Future research should endeavor to incorporate both micro and macroeconomic factors into a unified empirical framework.

Second, applying our findings to other developing countries may have problems. China is one of the largest developing countries, but the relative poverty lines we use are specific to China, so they may not be directly applicable to other places. On the other hand, many factors like household consumption and community environment are common across countries. Therefore, further analysis clearly requires the adoption of local data to propose corresponding governance measures adapted to actual situations.

Third, we did not handle the endogeneity problem using common identification methods, because our research question is in the prediction paradigm, and not in the causal inference paradigm. This does not mean endogeneity is irrelevant. Rather, it reflects a deliberate methodological choice aligned with our research objective: developing an operational tool for predicting poverty recurrence to enable early intervention and efficient resource allocation. Nevertheless, we acknowledge that the associations we identify are predictive, and not necessarily causal. This does not necessarily harm the prediction. For example, if poverty causes the use of non-clean energy (reverse causality) and this relationship is stable in the study population, then the use of non-clean energy remains a valid predictor of poverty even though we haven't identified the causal direction. We admit, nonetheless, that our results cannot directly provide a quantitative outcome of a specific intervention. Future research should continue to bridge this gap, which is indeed a cutting-edge research topic in econometrics.

Funding

This work was supported by the <National Natural Science Foundation of China #1> under Grant [number 82060265]; <Shaanxi Department of Science and Technology #2> under Grant [number 2025YXYC023]; <Postgraduate Education Reform Project of Xinjiang University #3> under Grant [number XJDX2025YJG16].

Author contributions

Qing Li: Writing–review & editing, writing–original draft, validation, supervision, project administration, methodology, funding acquisition, data curation, conceptualization. Chuang Sun: Writing–review & editing, writing–original draft, visualization, validation, software, methodology, formal analysis, data curation, conceptualization. Damien Echevin: Methodology, writing–review & editing. Yan Jie Li: Visualization, robustness analyses. Ying Xue Li: Writing–review & editing, methodology, validation, funding acquisition.

Disclosure statement

The authors declare no conflict of interest.

References

Acar, E. Ö., Günalp, B., & Cilasun, S. M. (2016). An empirical analysis of household education expenditures in Turkey. *International Journal of Educational Development*, *51*, 23–35. https://doi.org/10.1016/j.ijedudev.2016.03.007

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188–2244. https://doi.org/10.1086/705716

Aiken, E. L., Bedoya, G., Blumenstock, J. E., & Coville, A. (2023). Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. *Journal of Development Economics*, *161*, Article 103016. https://doi.org/10.1016/j.jdeveco.2022.103016

Alloush, M., & Bloem, J. R. (2022). Neighborhood violence, poverty, and psychological well-being. *Journal of Development Economics*, *154*, Article 102756. https://doi.org/10.1016/j.jdeveco.2021.102756

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Baird, K. (2016). High out-of-pocket medical spending among the poor and elderly in nine developed countries. *Health Services Research*, *51*(4), 1467–1488. https://doi.org/10.1111/1475-6773.12444

Barbier, E. B., & Burgess, J. C. (2020). Sustainability and development after COVID-19. *World Development*, *135*, Article 105082. https://doi.org/10.1016/j.worlddev.2020.105082

Barriga-Cabanillas, O., Blumenstock, J. E., Lybbert, T. J., & Putman, D. S. (2025). Probing the limits of mobile phone metadata for poverty prediction and impact evaluation. *Journal of Development Economics*, *174*, Article 103462. https://doi.org/10.1016/j.jdeveco.2025.103462

Baulch, B., & Hoddinott, J. (2000). Economic mobility and poverty dynamics in developing countries. *The Journal of Development Studies*, *36*(6), 1–24. https://doi.org/10.1080/00220380008422652

Bayudan-Dacuycuy, C., & Lim, J. A. (2013). Family size, household shocks and chronic and transient poverty in the Philippines. *Journal of Asian Economics*, *29*, 101–112. https://doi.org/10.1016/j.asieco.2013.10.001

Biewen, M. (2009). Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics*, *24*(7), 1095–1116. https://doi.org/10.1002/jae.1081

Blackburn, M. L. (1990). Trends in poverty in the United States, 1967-84. *Review of Income and Wealth*, *36*(1), 53–66. https://doi.org/10.1111/j.1475-4991.1990.tb00261.x

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *Journal of Development Studies*, *42*(2), 178–199. https://doi.org/10.1080/00220380500405261

Chen, H., Yu, J., Zhang, Z., Li, Y., Qin, L., & Qin, M. (2024). How human capital prevents intergenerational poverty transmission in rural China: Evidence based on the Chinese general social survey. *Journal of Rural Studies*, *105*, Article 103220. https://doi.org/10.1016/j.jrurstud.2024.103220

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Corral, P., Henderson, H., & Segovia, S. (2025). Poverty mapping in the age of machine learning. *Journal of Development Economics*, *172*, Article 103377. https://doi.org/10.1016/j.jdeveco.2024.103377

Duclos, J. Y., Araar, A., & Giles, J. (2010). Chronic and transient poverty: Measurement and estimation, with evidence from China. *Journal of Development Economics*, *91*(2), 266–277. https://doi.org/10.1016/j.jdeveco.2009.09.002

Echevin, D. (2013). Measuring vulnerability to asset-poverty in sub-Saharan Africa. *World Development*, *46*, 211–222. https://doi.org/10.1016/j.worlddev.2013.02.001

Echevin, D. (2014). Characterising vulnerability to poverty in rural Haiti: A multilevel decomposition approach. *Journal of Agricultural Economics*, *65*(1), 131–150. https://doi.org/10.1111/1477-9552.12017

Echevin, D., Fotso, G., Bouroubi, Y., Coulombe, H., & Li, Q. (2025). Combining survey and census data for improved poverty prediction using semi-supervised deep learning. *Journal of Development Economics*, *172*, Article 103385. https://doi.org/10.1016/j.jdeveco.2024.103385

Enneking, G., Custers, G., & Engbersen, G. (2025). The rapid rise of social infrastructure: Mapping the concept through a systematic scoping review. *Cities*, *158*, Article 105608. https://doi.org/10.1016/j.cities.2024.105608

Farrell, C. A., Fleegler, E. W., Monuteaux, M. C., Wilson, C. R., Christian, C. W., & Lee, L. K. (2017). Community poverty and child abuse fatalities in the United States. *Pediatrics*, *139*(5), Article e20161616. https://doi.org/10.1542/peds.2016-1616

Finnie, R., & Sweetman, A. (2003). Poverty dynamics: Empirical evidence for Canada. *Canadian Journal of Economics/Revue canadienne d'économique*, *36*(2), 291–325. https://doi.org/10.1111/1540-5982.t01-1-00002

Fofana, I., Mabugu, R. E., Camara, A., & Abidoye, B. (2024). Ending poverty and accelerating growth in South Africa, through the expansion of its social grant system. *Journal of Policy Modeling*, *46*(6), 1090–1102. https://doi.org/10.1016/j.jpolmod.2024.07.003

Gallardo, M. (2018). Identifying vulnerability to poverty: A critical survey. *Journal of Economic Surveys*, *32*(4), 1074–1105. https://doi.org/10.1111/joes.12216

Herbst-Debby, A., Endeweld, M., & Kaplan, A. (2021). Differentiated routes to vulnerability: Marital status, children, gender and poverty. *Advances in Life Course Research*, *49*, Article 100418. https://doi.org/10.1016/j.alcr.2021.100418

Jalan, J., & Ravallion, M. (2000). Is transient poverty different? Evidence for rural China. *Journal of Development Studies*, *36*(6), 82–99. https://doi.org/10.1080/00220380008422655

Jenkins, S. P., Schluter, C., & Wagner, G. (2003). The dynamics of child poverty: Britain and Germany compared. *Journal of Comparative Family Studies*, *34*(3), 337–355. https://doi.org/10.3138/jcfs.34.3.337

Jiao, X., Wang, X., & Zhang, J. (2025). Household expenditures on education and the relative poverty in China. *Finance Research Letters*, *72*, Article 106476. https://doi.org/10.1016/j.frl.2024.106476

Jung, S., Cho, S. H., & Roberts, R. K. (2015). The impact of government funding of poverty reduction programmes. *Papers in Regional Science*, *94*(3), 653–676. https://doi.org/10.1111/pirs.12089

Klevens, R. M., & Luman, E. T. (2001). US children living in and near poverty: Risk of vaccine-preventable diseases. *American Journal of Preventive Medicine*, *20*(4), 41–46. https://doi.org/10.1016/S0749-3797(01)00281-1

Li, C., & Managi, S. (2025). Impacts of community attachment and community livability on environmental activity according to XGBoost and SHAP. *Cities*, *156*, Article 105559. https://doi.org/10.1016/j.cities.2024.105559

Li, Q., Yu, S., Échevin, D., & Fan, M. (2022). Is poverty predictable with machine learning? A study of DHS data from Kyrgyzstan. *Socio-Economic Planning Sciences*, *81*, Article 101195. https://doi.org/10.1016/j.seps.2021.101195

Li, R., Song, H., Zhang, J., & Zhang, J. (2025). The effects of a multifaceted poverty-alleviation program on rural income and household behavior in China. *American Economic Journal*: *Economic Policy*, *17*(2), 319–357. https://doi.org/10.1257/pol.20210485

Li, S. P., Dong, Y. Q., Zhang, L. X., & Liu, C. F. (2021). Off-farm employment and poverty alleviation in rural China. *Journal of Integrative Agriculture*, *20*(4), 943–952. https://doi.org/10.1016/S2095-3119(21)63616-X

Lin, B., & Wang, Y. (2020). Does energy poverty really exist in China? From the perspective of residential electricity consumption. *Energy Policy*, *143*, Article 111557. https://doi.org/10.1016/j.enpol.2020.111557

Liu, F., Long, X., Dong, L., & Fang, M. (2023). What makes you entrepreneurial? Using machine learning to investigate the determinants of entrepreneurship in China. *China Economic Review*, *81*, Article 102029. https://doi.org/10.1016/j.chieco.2023.102029

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, *337*(6101), 1505–1510. https://doi.org/10.1126/science.1224648

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review*, *103*(3), 226–231. https://doi.org/10.1257/aer.103.3.226

Montalvo-Clavijo, C., Castillo-Chavez, C., Perrings, C., & Mubayi, A. (2023). Neighborhood effects, college education, and social mobility. *Socio-Economic Planning Sciences*, *86*, Article 101471. https://doi.org/10.1016/j.seps.2022.101471

Morduch, J. (1994). Poverty and vulnerability. *American Economic Review*, *84*(2), 221–225.

Poblete-Cazenave, M., & Pachauri, S. (2021). A model of energy poverty and access: Estimating household electricity demand and appliance ownership. *Energy Economics*, *98*, Article 105266. https://doi.org/10.1016/j.eneco.2021.105266

Rigg, J. (2006). Land, farming, livelihoods, and poverty: Rethinking the links in the Rural South. *World Development*, *34*(1), 180–202. https://doi.org/10.1016/j.worlddev.2005.07.015

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, *179*(6), 764–774. https://doi.org/10.1093/aje/kwt312

Shi, J., Chen, J., Xian, Z., & Zhao, L. (2025). Public services and vulnerability to poverty. *Research in International Business and Finance*, *75*, Article 102775. https://doi.org/10.1016/j.ribaf.2025.102775

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science*, *32*(3), 856–880. https://doi.org/10.1287/orsc.2020.1382

Sun, Z., & Xu, X. (2024). Can access to financial markets be an important option for rural families to break the return to poverty due to illness in China? *Agriculture*, *14*(2), Article 165. https://doi.org/10.3390/agriculture14020165

Suzuki, M. (2025). Affording expensive ceremonies: Evidence from quinceañeras in Mexico. *Journal of Development Economics*, *174*, Article 103448. https://doi.org/10.1016/j.jdeveco.2024.103448

Tran, T. K. (2023). Does microfinancing, financial inclusion, and educational loans alleviate poverty and inequality: Evidence from Vietnam. *Technological and Economic Development of Economy*, *29*(6), 1687–1707. https://doi.org/10.3846/tede.2023.20348

Walker, R., & Yang, L. (2021). *China's move to measuring relative poverty: Implications for social protection* (ILO Working Paper No. 23). https://webapps.ilo.org/static/english/intserv/working-papers/wp023/index.html

Wan, G., & Zhang, Y. (2013). Chronic and transient poverty in rural China. *Economics Letters*, *119*(3), 284–286. https://doi.org/10.1016/j.econlet.2013.02.036

Wang, H., Huang, W., Yu, Z., & Handi, J. (2025). Prediction and pattern recognition of the large-scale poverty-returning risks: Empirical evidence from machine learning. *International Food and Agribusiness Management Review*, *28*(2), 357–374. https://doi.org/10.22434/ifamr.1096

Ward, P. S. (2016). Transient poverty, poverty dynamics, and vulnerability to poverty: An empirical analysis using a balanced panel from rural China. *World Development*, *78*, 541–553. https://doi.org/10.1016/j.worlddev.2015.10.022

Wu, J., Yang, H., & Ahmed, T. (2023). An assessment of the policy of poverty alleviation in continuous poverty-stricken areas: Evidence from Yunnan Province, China. *Environment, Development and Sustainability*, *25*(9), 9757–9777. https://doi.org/10.1007/s10668-022-02459-4

Wu, Z., Long, H., & Song, H. (2024). The impact of infrastructure investment on multidimensional poverty. Evidence from Chinese rural migrant workers. *Economic Systems*, *48*(3), Article 101239. https://doi.org/10.1016/j.ecosys.2024.101239

Xu, D. (2018). From poverty to prosperity: College education, noncognitive abilities, and first-job earnings. *Chinese Sociological Review*, *50*(1), 53–82. https://doi.org/10.1080/21620555.2017.1376581

Yang, H., Bain, R., Bartram, J., Gundry, S., Pedley, S., & Wright, J. (2013). Water safety and inequality in access to drinking-water between rich and poor households. *Environmental Science & Technology*, *47*(3), 1222–1230. https://doi.org/10.1021/es303345p

Zang, X., Feng, J., & Song, M. (2024). The impact of air pollution on household vulnerability to poverty: An empirical study from household data in China. *Economic Analysis and Policy*, *82*, 1369–1383. https://doi.org/10.1016/j.eap.2024.05.018

Zeng, Q., He, Z., & Zeng, Y. (2023). Public procurement, consumers' preference and poverty alleviation through consumption. *Socio-Economic Planning Sciences*, *87*, Article 101514. https://doi.org/10.1016/j.seps.2023.101514

Zhang, A., & Imai, K. S. (2024). Do public pension programmes reduce elderly poverty in China? *Review of Development Economics*, *28*(1), 3–33. https://doi.org/10.1111/rode.13016

Zhang, J., & Zhao, Z. (2026). Corporate ESG rating prediction based on XGBoost-SHAP interpretable machine learning model. *Expert Systems with Applications*, *295*, Article 128809. https://doi.org/10.1016/j.eswa.2025.128809

Zhou, D., & Wang, M. (2019). The internal logic of poverty-returning: Theory and verification of vulnerable poverty alleviation. *Journal of Finance and Economics*, *45*(11), 126–139. https://doi.org/10.16538/j.cnki.jfe.2019.11.010 (in Chinese)

Zhou, Y., Huang, X., Shen, Y., & Tian, L. (2023). Does targeted poverty alleviation policy lead to happy life? Evidence from rural China. *China Economic Review*, *81*, Article 102037. https://doi.org/10.1016/j.chieco.2023.102037

Zuo, H., Li, S., Ge, Z., & Chen, J. (2023). The impact of education on relative poverty and its intergenerational transmission – Causal identification based on the Compulsory Education Law. *China Economic Review*, *82*, Article 102071. https://doi.org/10.1016/j.chieco.2023.102071

APPENDIX

Table A1. Descriptive statistics of variables

Category	Variable	Descriptive	Mean	Std. dev.
Target variable	Family return to poverty in critical poverty groups	1 if the individual's family returns to poverty in the critical poverty groups, and 0 otherwise	0.102	0.302
Personal characteristics	Gender (PC1)	1 if male, and 0 otherwise	0.489	0.500
	Age (PC2)	Age at the time of data collection	44.832	15.857
	Educational level (PC3)	2 if the respondent has a graduate degree, 1 if the highest degree is a bachelor's, 0 otherwise	0.045	0.226
	Pension insurance (PC4)	Dummy variable, 1 if pension insurance is purchased, 0 otherwise	0.467	0.499
	Medical insurance (PC5)	Dummy variable, 1 if health insurance is purchased, 0 otherwise	0.862	0.345
	Marital status (PC6)	Categorical variables, 1 to 5 are unmarried, married (in marriage), cohabiting, divorced, and widowed	2.022	0.784
	Hospitalization due to illness (PC7)	Dummy variable, 1 if the individual was hospitalized due to illness, 0 otherwise	0.097	0.295
	Health status (PC8)	Dummy variable, 1 if the individual is healthy, 0 otherwise	0.853	0.354
	Chronic diseases (PC9)	Dummy variable, 1 if the individual has a chronic disease, 0 otherwise	0.137	0.344
	Working status (PC10)	Dummy variable, 1 if the individual is working, 0 otherwise	0.766	0.424
	Religion (PC11)	Dummy variable, 1 if the individual is a member of a religion, 0 otherwise	0.021	0.143
	Membership (PC12)	Dummy variable, 1 if the individual is a union member, 0 otherwise	0.033	0.178
	Urban-rural classification (PC13)	Categorical variable, 1 if the individual is categorized as towns, otherwise 2	1.657	0.475
Family characteristics	Water for cooking (FC1)	1 = River/lake, 2 = Well, 3 = Tap, 4 = Bottled/purified/filtered, 5 = Rainwater, 6 = Cellar, 7 = Pond/spring water	2.991	0.971
	Cooking fuel (FC2)	1 = Firewood, 2 = Coal, 3 = Canned/ liquefied gas, 4 = Natural/piped gas, 5 = Solar/biogas, 6 = Electricity	3.147	1.892
	Monthly electricity costs (FC3)	Log-transformed per capita family electricity costs (after adding 1)	1.286	0.385
	Major domestic accidents (FC4)	Dummy variable, 1 if the family has a major incident, 0 otherwise	0.117	0.321
	Expenditures on major accidents (FC5)	Log-transformed per capita family expenditures on major accidents (after adding 1)	0.171	0.743
	Income from major incidents (FC6)	Log-transformed per capita family income from major accidents (after adding 1)	0.141	0.670

End of Table A1

Category	Variable	Descriptive	Mean	Std. dev.
Family characteristics	Family size (FC7)	Log-transformed number of family members (after adding 1)	0.706	0.204
	Area of current housing (FC8)	Log-transformed per capita family current housing area (after adding 1)	1.042	1.231
	Current value of the house property (FC9)	Log-transformed per capita family current value of the house property (after adding 1)	0.521	0.407
	Expenditures on mortgage repayment (FC10)	Log-transformed per capita family expenditures on mortgage repayment (after adding 1)	0.018	0.083
	Government grants (FC11)	Dummy variable, 1 if the family receives a government subsidy, 0 otherwise	0.473	0.499
	Total government grants (FC12)	Log-transformed per capita families total government subsidies (after adding 1)	0.184	0.493
	Social contributions (FC13)	Dummy variable, 1 if the families receive social contributions, 0 otherwise	0.035	0.183
	Food expenditures (FC14)	Log-transformed per capita family expenditures on food (after adding 1)	3.409	0.629
	Expenditures on family equipment and daily necessities (FC15)	Log-transformed per capita family expenditures on equipment and daily necessities (after adding 1)	2.782	0.660
	Expenditures on transportation and communication (FC16)	Log-transformed per capita family expenditures on transportation and communication (after adding 1)	2.836	0.644
	Expenditures on culture, education, and recreation (FC17)	Log-transformed per capita family expenditures on culture, education, and recreation (after adding 1)	2.000	1.417
	Car ownership (FC18)	Dummy variable, 1 if the family owns a car, 0 otherwise	0.294	0.456
Community environmental characteristics	Public environmental facilities (CE1)	Satisfaction with public amenities, rated from 1 = "very good" to 5 = "very poor"	2.589	1.023
	Surroundings (CE2)	The neighborhood in which the family currently lives is rated from 1 = "very good" to 5 = "very bad"	2.513	1.026
	Neighborhood security (CE3)	Perceived neighborhood security, rated from 1 (very good) to 5 (very poor)	2.144	0.970
	Neighborly relations (CE4)	Quality of neighborly relations ranges from 1 = "very good" to 5 = "very poor"	1.976	0.870
	Neighborly assistance (CE5)	Perceived likelihood of neighborly assistance, measured on a scale from 1 (very likely) to 5 (very unlikely)	1.477	0.795
	Affection to the community (CE6)	Level of emotional attachment to community, rated from 1 (very strong) to 5 (very weak)	1.853	0.881

Table A2. Description of parameters used in the model

Parameter name	Parameter interpretation	Search space	Optimal hyperparameters
objective	The learning objective and task. For this binary classification study		binary: logistic
num_rounds	The number of boosting iterations	[50, 300]	163
eta	The learning rate refers to the parameter that controls the step size during model training and is utilized to prevent overfitting	[0.01, 0.3]	0.2275
max_depth	The maximum depth of a tree refers to the longest path from the root node to the farthest leaf node within a single tree. This parameter is crucial as it helps to prevent overfitting by limiting the depth of the tree	[3, 10]	5
colsample_bytree	This refers to the proportion of randomly sampled features (or columns) per tree, which is used to control for overfitting	[0.5, 1]	0.75
subsample	This refers to the proportion of randomly sampled data rows per tree, which is used to prevent overfitting	[0.5, 1]	0.875
min_child_weight	This refers to the hyperparameter used to control the sum of the weights of the minimum leaf nodes during the process of node splitting, which helps in preventing overfitting	[1, 10]	1

Table A3. Description of the COVID-19 variables

COVID-19 related variables	Description
Working hours during the pandemic (COVID1)	Categorical variable: 0–5 indicates not applicable, reduced, unchanged, increased, completely unable to work, and no job at that time, respectively
Working arrangements during the pandemic (COVID2)	Categorical variable: 0–4 indicates not applicable, fully remote work, mostly remote work, occasionally remote work, and no remote work, respectively
Work income during the pandemic (COVID3)	Categorical variable: 0–3 indicates not applicable, reduced, increased, and unchanged, respectively
Monthly income fluctuation (COVID4)	Continuous variable: percentage change in monthly income
Unemployment probability (COVID5)	Continuous variable: percentage probability of unemployment due to company closure or layoffs within the next year
Frequency of spousal conflicts (COVID6)	Categorical variable: 0–3 indicates not applicable, increased, decreased, and unchanged, respectively
Frequency of child-related conflicts (COVID7)	Categorical variable: 0–3 indicates not applicable, increased, decreased, and unchanged, respectively
Frequency of parental conflicts (COVID8)	Categorical variable: 0–3 indicates not applicable, increased, decreased, and unchanged, respectively

The COVID-19 pandemic data utilized in this study is derived from the CFPS general restricted dataset, which we have formally requested and downloaded from the official CFPS website. Due to data confidentiality requirements, we do not provide descriptive statistics here, instead, we include the explanations of the variables.

Table A4. Results after parameterization

Learning_rate	Max_depth	AUC	Accuracy	Precision	Recall	F1-score
Output results for the parameters selected in this paper						
0.2275	5	0.9650	0.9596	0.8537	0.7143	0.7778
Output results after changing parameters						
0.2275	6	0.9677	0.9515	0.7907	0.6939	0.7391
0.2275	4	0.9595	0.9475	0.7949	0.6327	0.7046
0.235	5	0.9655	0.9535	0.8095	0.7143	0.7589
0.22	5	0.9670	0.9556	0.8462	0.6735	0.7500
0.235	6	0.9714	0.9616	0.875	0.7143	0.7865
0.22	6	0.9696	0.9596	0.8718	0.6939	0.7727
0.235	4	0.9518	0.9596	0.8919	0.6735	0.7674
0.22	4	0.9620	0.9636	0.8974	0.7143	0.7955

Table A5. Examining the robustness of the poverty-return definition across time

Evaluation metrics / Results of different situations	AUC	Accuracy	Precision	Recall	F1-score
Output results of the selected basis	0.9650	0.9596	0.8537	0.7143	0.7778
Output results of returning to poverty in 2016	0.8972	0.9282	0.9600	0.4800	0.6400
Output results of extending critical poverty period	0.9191	0.9593	0.8000	0.6667	0.7273
Output results of returning to poverty in 2022	0.9691	0.9540	0.8974	0.6604	0.7609
Output results after eliminating sustained non-poverty samples	0.8450	0.8342	0.7941	0.5294	0.6353