Journal of Computer Science and Digital Technology, 2026, 2(1), 1-20
https://doi.org/10.71204/hefd5f67

Precise Identification of Multi-Regional Relative Poverty: A Two-Stage Knowledge-Distilled Adaptive Framework

Guanghuang Liu [1, *]

[1] School of Computer Science and Technology, Taiyuan Normal University, Shanxi 030619, China

*** Correspondence:**

Guanghuang Liu

lgh09263295@163.com

Received: 1 February 2026 / Accepted: 10 March 2026 / Published online: 11 March 2026

Abstract

As poverty reduction strategies shift comprehensively toward alleviating relative poverty, precisely identifying multidimensional relative poverty populations has become critical for social governance. However, existing data-driven models often face algorithmic bottlenecks—such as spatial heterogeneity, regional sample sparsity, and extreme category imbalance—when applied to complex scenarios characterized by vast territories and significant regional disparities. To address this, this study proposes a two-stage knowledge-distilled adaptive gradient boosting framework (TKDAF). First, in the prior knowledge extraction stage (Stage I), a base structure-regularized gradient boosting tree model is constructed. Combined with SHAP game-theoretic attribution, this stage quantifies and extracts the global objective weights of poverty-inducing features across regions. Second, in the spatial adaptive enhancement stage (Stage II), the S-DAGB (Spatial-Adaptive Distilled Gradient Boosting) core prediction model is introduced. It achieves deep integration of multiple regularization and feature enhancement mechanisms by incorporating: feature space nonlinear reconstruction based on prior knowledge, a dynamic category weighting mechanism based on effective sample size (ENS), and spatial adaptive optimization using the TPE Bayesian algorithm.Empirical results based on the 2020 China Family Panel Studies (CFPS) multidimensional dataset demonstrate that the S-DAGB model not only effectively overcomes the generalization bottleneck of deep tree models in the sample-constrained Northeast region (achieving 93.52% accuracy),but also significantly improves precision in regions with highly heterogeneous features and extreme class imbalance, such as central and western China. This effectively reduces wasteful allocation of poverty alleviation resources caused by false positive errors. This study provides an algorithmic solution that combines high accuracy with interpretability for precise identification of relative poverty in complex data distribution scenarios.

Keywords: Relative Poverty Prediction; Gradient-Boosted Trees; Knowledge Distillation; Spatial Heterogeneity; Extremely Imbalanced Classification; Bayesian Optimization

1. Introduction

As global poverty reduction efforts advance, China's poverty governance has historically shifted its focus from eliminating absolute poverty to alleviating relative poverty (Wang & Zeng, 2018). Unlike absolute poverty, which relies on a single income benchmark, relative poverty exhibits distinct multidimensionality, dynamism, and spatial heterogeneity. It encompasses not only economic deprivation but also the long-term accumulation of multidimensional disadvantages in education, health, living conditions, and assets (Wang et al., 2020; Sun et al., 2019). The Multidimensional Poverty Index (MPI), promoted by organizations such as the United Nations Development Programme (UNDP), provides a crucial framework for characterizing this form of poverty (Alkire & Foster, 2011). Against this backdrop, traditional identification methods relying on single thresholds or simple linear regression struggle to adequately meet the demand for precise policy interventions amid intersecting multidimensional characteristics.In recent years, machine learning algorithms such as Random Forest and Gradient Boosting Trees have demonstrated immense application potential in micro-level household feature extraction and poverty identification due to their exceptional ability to capture nonlinear feature interactions. They have gradually become a leading paradigm in computational social science (Shi et al., 2024; Jean et al., 2016).

However, existing data-driven relative poverty prediction models still face three severe algorithmic bottlenecks when applied to China's complex micro-survey data, characterized by vast geographical scope and significant regional disparities. The first challenge is regional data sparsity and overfitting risk. When predictions are narrowed to specific regions (such as central-western and northeastern China), the available sample size often shrinks dramatically. Complex deep ensemble tree models like XGBoost (Chen et al., 2016) and LightGBM (Ke et al., 2017) are highly susceptible to local noise interference in small-sample scenarios, severely limiting their generalization capabilities (Qian et al., 2022). Second is the extreme positive-negative class imbalance. After entering the phase of routine poverty governance, the poor population typically exhibits a long-tail distribution within the overall sample. Conventional loss functions, when optimizing such extremely imbalanced data, often sacrifice recall for the minority class (poverty samples) to achieve overall accuracy. Traditional resampling methods also tend to distort the original distribution (Chawla et al., 2002), making it difficult to align with the policy goal of "accurate identification."Finally, spatial heterogeneity in feature contributions and the absence of prior knowledge pose significant challenges. Authoritative development economics research indicates that core factors driving poverty exhibit pronounced structural and spatial imbalances across different regions (Zou & Fang, 2011; Ravallion & Chen, 2007). Existing end-to-end models predominantly explore raw feature spaces without prior constraints, not only reducing information extraction efficiency in small samples but also trapping models in a "black box" dilemma, lacking sociologically meaningful interpretability.

To address these challenges, this study proposes a Two-Stage Knowledge-Distilled Adaptive Gradient Boosting Framework (TKDAF).This framework innovatively decouples the prediction process into two stages: "prior knowledge extraction" and "spatial adaptive boosting." In Stage I, a base boosting tree model with strict structural regularization is constructed. The SHAP (Shapley Additive Explanations) game theory model is then employed to extract the attribution distribution of global impoverishment features as "prior knowledge" (Lundberg et al., 2017). In Stage II, the S-DAGB adaptive model is proposed, achieving high-precision, interpretable poverty identification through multi-mechanism fusion. The main contributions of this paper are as follows:

(1) We introduce a knowledge-driven nonlinear feature space reconstruction strategy. Using the SHAP prior weights extracted in Stage I as guiding factors, we reconstruct the input feature space through an adaptive scaling formula. This guides the underlying tree model to prioritize core impoverishment factors during tree splitting, significantly enhancing feature resolution in small-sample scenarios.

(2) We propose a dynamic weighting mechanism based on Effective Number of Samples (ENS). Addressing extreme data imbalance, it redefines the category penalty weight in the loss function by drawing from ENS theory (Cui et al., 2019), establishing a smoothing buffer layer within the sample space to effectively mitigate model bias toward majority-class dominant features.

(3) We designed a spatially adaptive regularization architecture based on TPE Bayesian optimization. By leveraging the Tree-like Parzen Estimator (TPE) (Bergstra et al., 2011) for joint optimization across multidimensional hyperparameter spaces, the model automatically adapts its underlying sampling engine and regularization terms according to regional data scale and heterogeneity, constructing a robust defense against overfitting.

2. Related Research and Model Framework

2.1. Relative Poverty Prediction and Multidimensional Feature Analysis

Since Alkire and Foster introduced the A-F dual-threshold measurement method, poverty metrics have expanded from a single economic dimension to encompass multidimensional relative poverty across health, education, and living standards (Alkire & Foster, 2011). Focusing on China's context, scholars have observed that regional disparities in resource endowments and economic development gradients result in strongly spatially unbalanced and multidimensionally dynamic trajectories of relative poverty (Zou & Fang, 2011; Ravallion & Chen, 2007). For instance, relative deprivation in the developed eastern regions manifests primarily as shortages of high-value assets and quality educational resources, whereas western and northeastern regions are more constrained by deficiencies in infrastructure and basic healthcare (Li et al., 2020). Traditional prediction methods based on a national-level perspective often obscure these structural regional disparities (Wang et al., 2022). Therefore, incorporating multidimensional feature analysis and designing algorithmic mechanisms at the model level that can adaptively

capture latent regional heterogeneity are crucial for enhancing the targeted identification of relative poverty.

2.2. Gradient Boosting Decision Trees (GBDT)

Ensemble learning possesses inherent advantages in handling complex tabular data (Breiman, 2001). Among these, Gradient Boosting Decision Trees (GBDT), a classic paradigm proposed by Friedman (Friedman, 2001), operates by iteratively accumulating multiple weak learners to progressively fit the negative gradient between the previous model's prediction and the true value.In recent years, with the widespread adoption of efficient frameworks like XGBoost (Chen et al., 2016) and LightGBM (Ke et al., 2017), gradient boosting algorithms have demonstrated significant engineering advantages in capturing nonlinear interactions among multidimensional features.However, when applied to micro-household survey data—particularly in sparsely sampled regions—conventional tree models tend to overfit local noise during node splitting due to insufficient information entropy constraints (Qian et al., 2022). Consequently, stringent structural regularization strategies (e.g.,L_2 leaf node penalties or dynamic feature sampling) are required to limit the complexity of individual trees, thereby ensuring generalization capability across regions.

2.3. Bayesian Hyperparameter Optimization Theory

Within complex machine learning ensemble frameworks, hyperparameter configuration directly determines a model's performance ceiling on specific data distributions. Traditional grid search not only incurs exponentially increasing computational costs but also struggles to locate global optima in continuous high-dimensional spaces. To address this challenge, Bayesian optimization theory based on probabilistic surrogate models emerged (Snoek et al., 2012). Among these, the Tree-like Parzen Estimator (TPE) algorithm proposed by Bergstra et al. constructs a probability distribution of the objective function across the hyperparameter space (Cui et al., 2019). By continuously computing the Expected Increment (EI) using historical evaluation results, it achieves an optimal balance between "Exploration" and "Exploitation."This mechanism enables deep tree models to break free from manual experience constraints, autonomously searching for optimal regularization configurations across varying data sparsity levels in different regions.

2.4. Overall Framework of the TKDAF Model

To effectively address challenges in multi-regional data—such as overfitting in small samples, extreme sample imbalance, and spatial heterogeneity—this study proposes a two-stage Knowledge Distillation Adaptive Gradient Boosting framework (TKDAF). This framework breaks the "black-box" paradigm of traditional end-to-end models by decoupling the prediction process into two stages: "knowledge extraction" and "adaptive enhancement"(overall architecture shown in Figure 1).

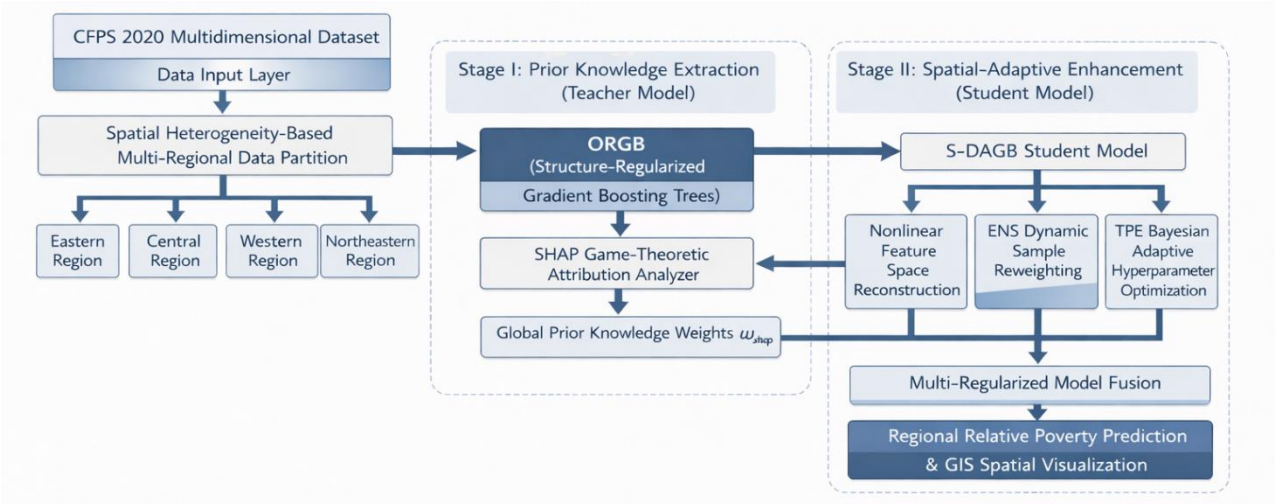


Figure 1. Overall framework architecture of TKDAF

(1) Stage I: Prior Knowledge Extraction Layer. The core objective of this stage is to extract objectively interpretable poverty patterns with sociological significance from geographically independent datasets. The framework first trains a strongly constrained basic objective regularized gradient boosting model (ORGB) on different geographic subsets.Upon model convergence, the SHAP game-theoretic attribution analyzer is applied to precisely quantify the marginal contribution of each multidimensional indicator to poverty status (Zou & Fang, 2011). These SHAP distributions, stripped of local noise, are converted into "global prior knowledge weights" and transferred losslessly to the next stage.

(2) Stage II: Spatially Adaptive Gradient Boosting (S-DAGB). This stage constitutes the core prediction engine—the S-DAGB model. Receiving raw regional data and Stage I prior knowledge, it achieves predictive performance leaps through three mechanisms: First, it performs nonlinear feature space reconstruction, adaptively scaling features using prior knowledge weights to amplify decision resolution for core impoverishment factors;Second, it introduces the ENS dynamic sample weighting mechanism (Ravallion & Chen, 2007) to reshape the category penalty weights in the loss function, countering gradient bias caused by long-tail imbalance (Lundberg & Lee, 2017). Finally, driven by TPE Bayesian adaptive optimization (Cui et al., 2019), the model automatically adapts to the data characteristics of the current region within a deep regularization space, outputting the final high-precision regional poverty identification results.

Unlike existing research paradigms in feature engineering, the knowledge distillation mechanism proposed in this study represents a significant methodological innovation. Current literature on interpretable machine learning (e.g., standard SHAP analysis) predominantly focuses on 'post-hoc explanation' —analyzing the black-box logic of trained models. The TKDAF framework innovatively shifts this approach by converting objective poverty attribution into 'pre-hoc prior knowledge.' Through establishing a nonlinear feature space reconstruction mechanism, the extracted global knowledge weights are seamlessly integrated into the underlying splitting logic of the student model (S-DAGB). This paradigm shift from 'passive interpretation' to 'active reconstruction' effectively guides tree models in small-sample regions to overcome noise traps, achieving truly knowledge-guided distillation.

Furthermore, conventional approaches to addressing extreme data imbalance and spatial heterogeneity predominantly rely on physical-level data resampling techniques (e.g., SMOTE algorithms), which may significantly distort the original multidimensional data's true spatial distribution. In contrast, this study introduces an ENS-based loss function reconstruction strategy within a spatially adaptive architecture, establishing a smoothing buffer layer through gradient optimization's mathematical foundation. The TPE engine-driven spatial adaptive framework dynamically activates underlying stochastic control and structural regularization constraints based on regional data sparsity (e.g., the sparse micro-samples in Northeast China versus the massive samples in the East). This non-invasive joint optimization solution fundamentally resolves the algorithmic bottleneck of traditional single-model approaches, which struggle to accommodate multi-regional heterogeneity while preserving the original spatial topology of regional data.

3. Research Methodology

3.1. Stage I: Extracting Prior Knowledge Based on Structured Regularized Gradient Boosting Trees

In the first stage, this study constructs an Objective Regularized Gradient Boosting (ORGB) model with structured regularization constraints to capture global objective patterns of poverty-inducing factors across different regions. Compared to directly feeding raw data into complex networks, using ORGB as a "teacher model" for preliminary data structure exploration effectively filters redundant noise.

To address the interpretability limitations of black-box models, this study incorporates the SHAP (Shapley Additive Explanations) game-theoretic attribution method to conduct in-depth analysis of ORGB predictions across regional datasets. For any sample's feature i , its SHAP attribution value ϕ_i is calculated as follows:

$$\phi_i(f,x)=\sum_{S\subseteq N\setminus\{i\}}\frac{|S|!(|N|-|S|-1)!}{|N|!}[f_x(S\cup\{i\})-f_x(S)]$$

Where N denotes the entire feature set, S represents the feature subset excluding i, and $f_x(S)$ is the model's predicted value on the subset S .

To obtain globally informative prior knowledge, we calculate the mean of the absolute SHAP values across all samples within a single region and normalize it to derive the global knowledge weight $w_{shap}^{(i)}$ for the i th feature in that region:

$$w_{shap}^{(i)}=\frac{\frac{1}{M}\sum_{j=1}^{M}|\phi_i^{(j)}|}{\sum_{k=1}^{K}\left(\frac{1}{M}\sum_{j=1}^{M}|\phi_k^{(j)}|\right)}$$

Where M denotes the total number of samples in the region, and K represents the total feature dimension. This weight precisely quantifies the actual contribution of each dimensional metric to poverty status in the physical world and is losslessly transferred as prior knowledge to Stage II.

3.2. Stage II: S-DAGB Spatially Adaptive Distilled Gradient Boosting Model

Guided by the prior knowledge extracted in Stage I, this stage introduces the full S-DAGB (Spatial-Adaptive Distilled Gradient Boosting) model. Through three core mechanisms, this model comprehensively addresses the bottlenecks of spatial heterogeneity and overfitting in small samples.

3.2.1. Unevenly Distributed Weight Reallocation Based on Effective Sample Size (ENS)

Relatively poverty data often exhibits extreme positive-negative class imbalance. Traditional resampling methods can distort the true data distribution, while simple inverse-frequency weighting may lead to model overfitting on outliers in the minority class. This study introduces the Effective Number of Samples (ENS) theory to redefine class weights. The ENS is defined as follows:

$$E_n=\frac{1-\beta^n}{1-\beta}$$

Where n represents the actual sample count for a specific class, and $\beta\in[0,1)$ denotes the decay coefficient (hyperparameter). As n increases, the new information provided by marginal samples diminishes progressively. Based on this, we make the weight α_c of each class in the loss function inversely proportional to its effective sample count and apply a normalization mapping:

$$W_c=\frac{1/E_{n_c}}{\sum(1/E_{n_i})}\times C$$

where C denotes the total number of classes. This mechanism penalizes majority class advantage while establishing a smoothing buffer layer, effectively suppressing gradient explosion in extremely imbalanced scenarios.

3.2.2. Knowledge-Guided Nonlinear Feature Space Reconstruction

To enhance the S-DAGB model's focus on key impoverishment factors during tree splitting, we propose a feature space reconstruction strategy based on Stage I prior knowledge. Through an adaptive scaling mechanism, the numerical distribution range of high-contribution features undergoes nonlinear stretching, as expressed by the following formula:

$$X_i'=X_i\times(1+w_{shap}^{(i)}\times\alpha)$$

where X_i represents the original input feature, $w_{shap}^{(i)}$ denotes the global prior weight passed from Stage I, and α is the feature scaling factor. This formula ensures stronger correlated features possess higher resolution during spatial partitioning, effectively preventing ineffective splits on noisy features in small-sample tree models.

3.2.3. TPE-Based Multi-Region Hyperparameter Adaptive Optimization and Regularization

Given the substantial differences in sample size and feature heterogeneity across Eastern, Central, Western, and Northeastern regions, a single parameter configuration cannot achieve global optimization. This study employs a Bayesian optimization algorithm based on the Tree-

structured Parzen Estimator (TPE) to jointly optimize the core regularization architecture and adaptive parameters of S-DAGB.

The optimization space encompasses not only the feature amplification factor (α) and ENS decay coefficient (β), but also deeply integrates structured regularization terms, including:

(1) Tree-structured regularization: Maximum tree depth (Depth) and minimum data in leaf nodes (Min Data In Leaf).

(2) Weight Penalty: Leaf node L_2 regularization coefficient (L_2Leaf Reg), strictly controlling model complexity.

(3) Dynamic Feature Sampling: Introduces hierarchical column sampling rate (Colsample By Level) to forcefully break excessive model dependence on any single dominant feature.

Simultaneously, the model incorporates a spatial scale adaptation mechanism at its core: - In regions with sufficient sample size (e.g., eastern areas), Bernoulli sampling is employed alongside subsampling regularization. - In sparsely sampled regions (e.g., northeast and central areas), the model automatically switches to Bayesian sampling mode and introduces the Bagging Temperature parameter to control randomness. This constructs multiple layers of defense against overfitting.

4. Experimental Results and Analysis

4.1. Dataset and Evaluation Metrics

This study conducts empirical analysis based on the 2020 China Family Panel Studies (CFPS) multidimensional dataset. To accurately capture the spatial heterogeneity of multidimensional relative poverty and validate the model's generalization capability under complex imbalanced distributions, the cleaned national sample is strictly divided into four independent subsamples according to geographical regions: Eastern Region (3,797 samples), Central Region (2,481 samples),Western Region (2,964 samples), and Northeastern Region (1,548 samples). The characteristics of each dataset are shown in Table 1 (using the Northeastern Region as an example).

It is particularly noteworthy that this dataset exhibits extreme spatial imbalance and scale heterogeneity. The available sample size for the Northeast region is exceptionally sparse, amounting to only 40.8% of the Eastern region's sample size.This severe "small-sample" distribution poses significant challenges to the overfitting resilience of complex machine learning models (such as ensemble tree algorithms) and provides an excellent experimental setting to validate the effectiveness of the S-DAGB adaptive regularization framework proposed in this paper. After data preprocessing and feature selection via analysis of variance (ANOVA), core dimensional features including household annual income, children's education, assets, and health status were retained for model input.

Table 1. Descriptive statistics of relative poverty data by region

Region	Variable	Meaning	count	mean	std	min	0.25	0.50	0.75	max
Northeast	x1	Adult Education	1548	0.76	0.43	0.0	1.0	1.0	1.0	1.0
	x2	Children Education	1548	0.14	0.34	0.0	0.0	0.0	0.0	1.0
	x3	Adult Health	1548	2.61	1.35	1.0	1.0	3.0	3.0	5.0
	x4	Children Health	1548	3.86	0.53	1.0	4.0	4.0	4.0	4.0
	x5	Medical Insurance	1548	0.17	0.37	0.0	0.0	0.0	0.0	1.0
	x6	Cooking Fuel	1548	3.71	1.82	1.0	3.0	4.0	6.0	6.0
	x7	Drinking Water	1548	2.81	0.50	1.0	3.0	3.0	3.0	4.0
	x8	Housing	1548	2.62	0.77	1.0	3.0	3.0	3.0	3.0
	x9	Assets	1548	36803.64	80691.22	0.0	2.0×10^3	9.0×10^3	3.43×10^4	1.3×10^6
	x10	Annual household income	1548	62627.23	85312.97	0.0	2.6×10^4	5.0×10^4	8.0×10^4	2.6×10^6
	y0	Real number type ActualAnnual income	1548	26743.24	69852.63	0.0	1.0×10^4	2.0×10^4	3.33×10^4	2.6×10^6
	y1	Category-based actualAnnual income	1548	0.79	0.40	0.0	1.0	1.0	1.0	1.0

To comprehensively evaluate the recognition performance of the TKDAF framework and its comparison models, this study employs five core metrics: Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). Their calculation formulas are as follows:

(1) Accuracy: The percentage of correctly classified samples relative to the total number of samples, measuring the overall predictive accuracy of the model.

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}$$

(2) Precision: The ratio of samples predicted as positive by the model that are actually positive, focusing on false positives.

$$Precision=\frac{TP}{TP+FP}$$

(3) Recall: The percentage of true positives correctly predicted, primarily used to measure false negatives.

$$Recall=\frac{TP}{TP+FN}$$

(4) F1 score: The harmonic mean of precision and recall, used to evaluate the model's performance in positive class prediction.

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall}$$

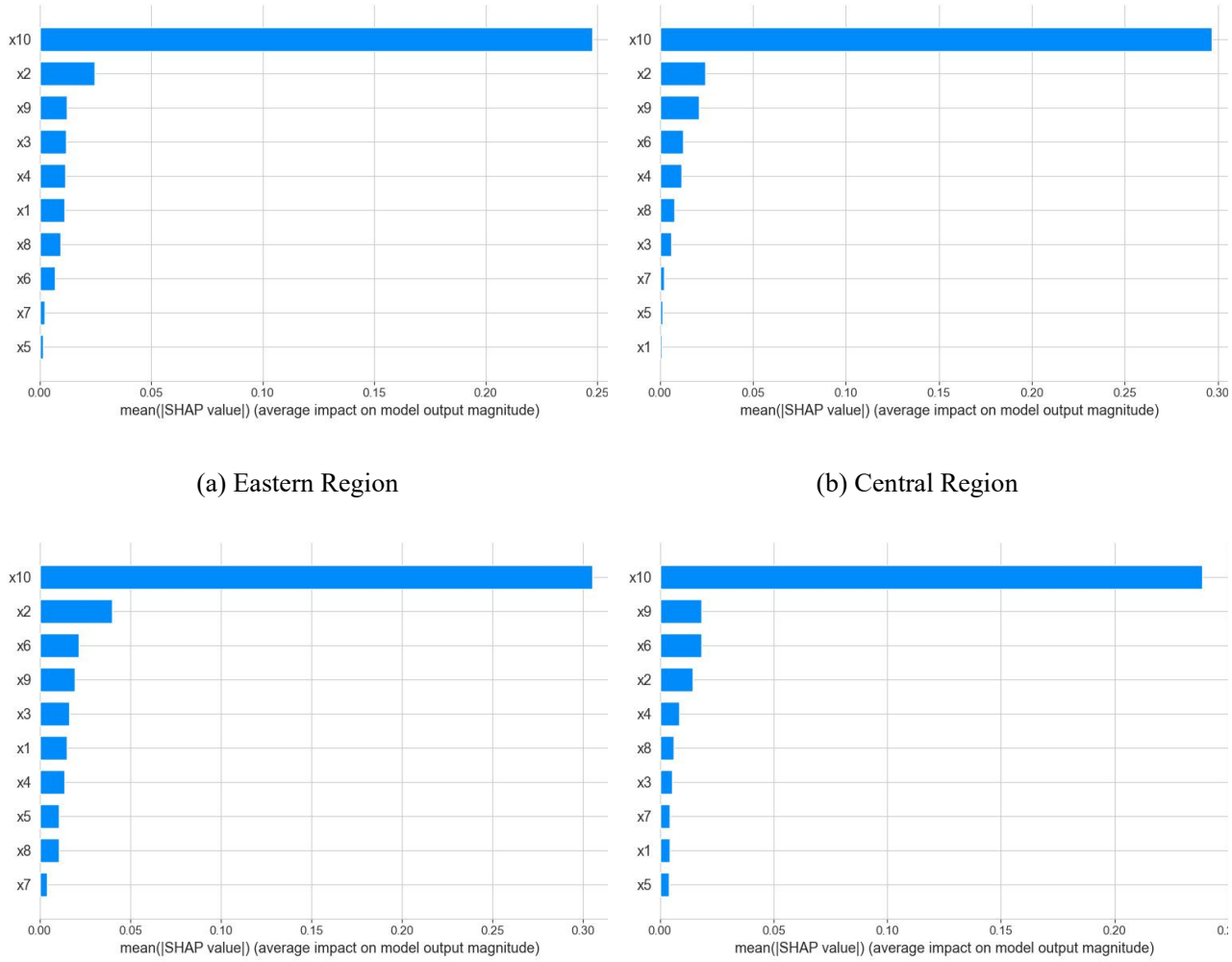
(5) AUC value: The area under the receiver operating characteristic (ROC) curve, measured as the area between the curve and the x-axis. It reflects the model's ability to distinguish between positive and negative classes. A value closer to 1 indicates better performance and greater model accuracy.

WhereTP represents correctly predicted poor samples,TN denotes correctly predicted non-poor samples,FP indicates false positive samples, andFN signifies false negative samples.

To ensure the objectivity of model evaluation and the reproducibility of results, this study adopted strict data partitioning and cross-validation strategies during the experimental design phase. For the independent datasets from four major regions, stratified sampling was used to divide the data into training and testing sets in a 80:20 ratio, ensuring that the initial distribution of positive and negative samples remained consistent before and after partitioning. Additionally, considering the extreme imbalance in data from the central and western regions, as well as the small sample size in the northeastern region, this study implemented stratified 10-fold cross-validation during both model training and TPE hyperparameter optimization. This strategy effectively mitigated the potential randomness in evaluations caused by single data splits, further enhancing the generalization and robustness of the S-DAGB model in complex multi-regional scenarios.

4.2. Stage I: SHAP-Based Extraction of Multidimensional Feature Prior Knowledge

In the first stage of the TKDAF framework, this study first independently trained a base gradient boosting model with structural regularization (Teacher Model) across four geographic subsets. The SHAP game theory model was then introduced to attribute and decompose the predictive contributions of each dimensional feature.The extracted Mean |SHAP value| intuitively reflects the objective weight of each feature in the global space. The SHAP contribution rankings across the four regions are shown in Figure 2.



(a) Eastern Region

(b) Central Region

(c) Western Region

(d) North-Eastern Region

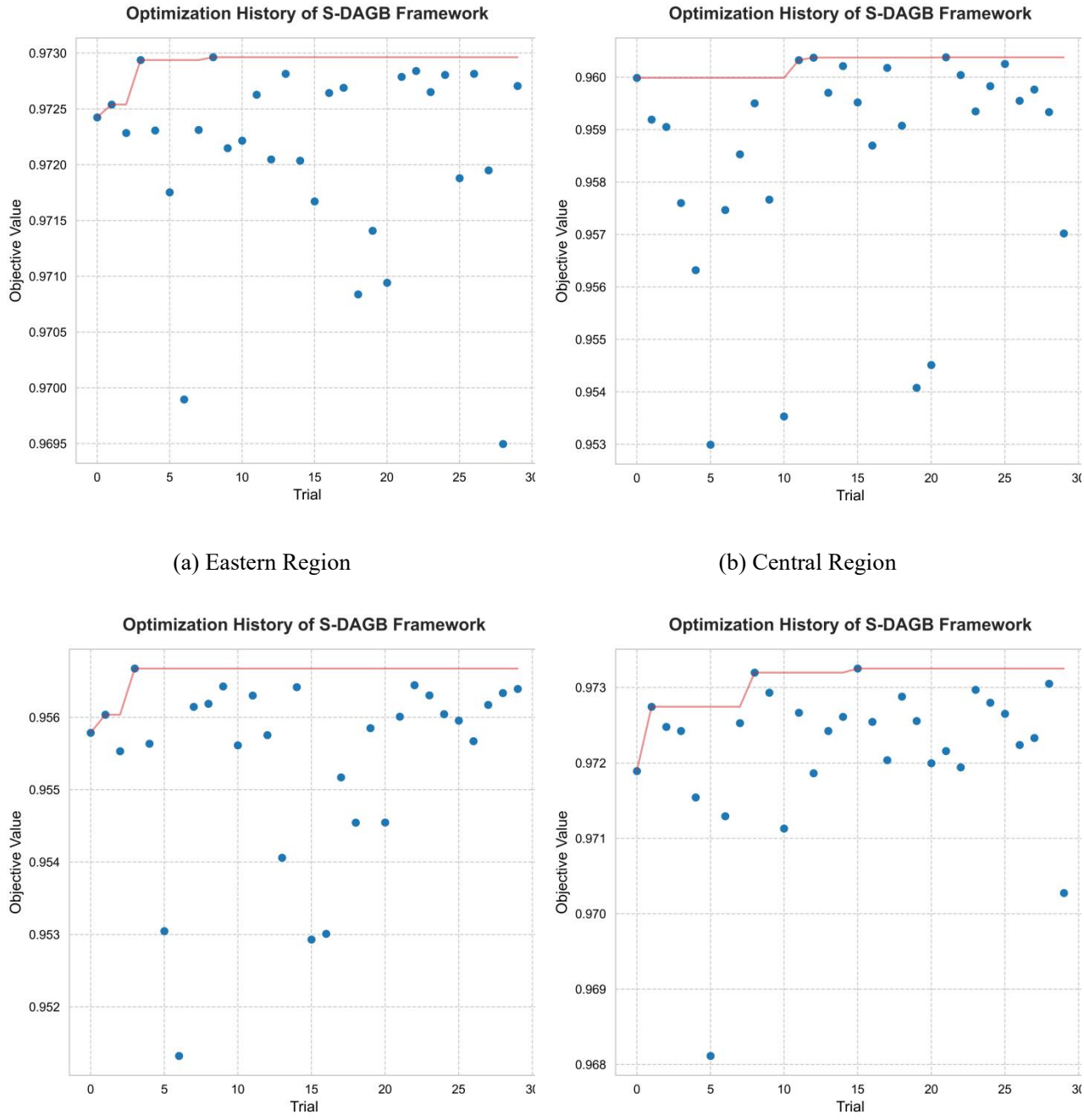
Figure 2. Comparison of mean absolute SHAP values across regions

The quantitative results in Figure 2 clearly reveal the spatial heterogeneity of the impoverishment logic:Although annual household income (x10) holds absolute dominance across all regions, developmental indicators like children's education (x2) and assets (x9) contribute significantly in eastern and central regions. Conversely, basic survival indicators such as cooking fuel (x6) and housing (x8) exhibit markedly elevated weights in western and northeastern regions.

Contrary to traditional sociological interpretations, the core purpose of extracting these distributions is to convert them into machine-readable "prior knowledge." By applying L_1 normalization to the absolute SHAP mean values corresponding to each bar chart, we generated specific prior knowledge weight vectors for the four regions: w_{shap} .After crossing the Stage I boundary, this vector is losslessly injected into the Stage II S-DAGB student model as a quantitative guidance factor for executing "feature space nonlinear reconstruction." This forces the underlying tree model to prioritize impoverishing features with high true physical significance during small-sample tree splitting, effectively shielding against local data noise.

4.3. Stage II: Bayesian Space Adaptive Optimization and Importance Analysis

To ensure the generalization capability of the S-DAGB framework across different regional datasets, this study introduces a Bayesian optimization algorithm based on the Tree-like Parzen Estimator (TPE) in Stage II. This algorithm performs joint optimization within a regularization space composed of feature reconstruction coefficients (Scale Multiplier), effective sample weighting decay rates (Beta), and multi-tree structure parameters.Figures 3 and 4 respectively illustrate the optimization history and hyperparameter importance distribution across four major geographic regions.



(a) Eastern Region

(b) Central Region

(c) Western Region

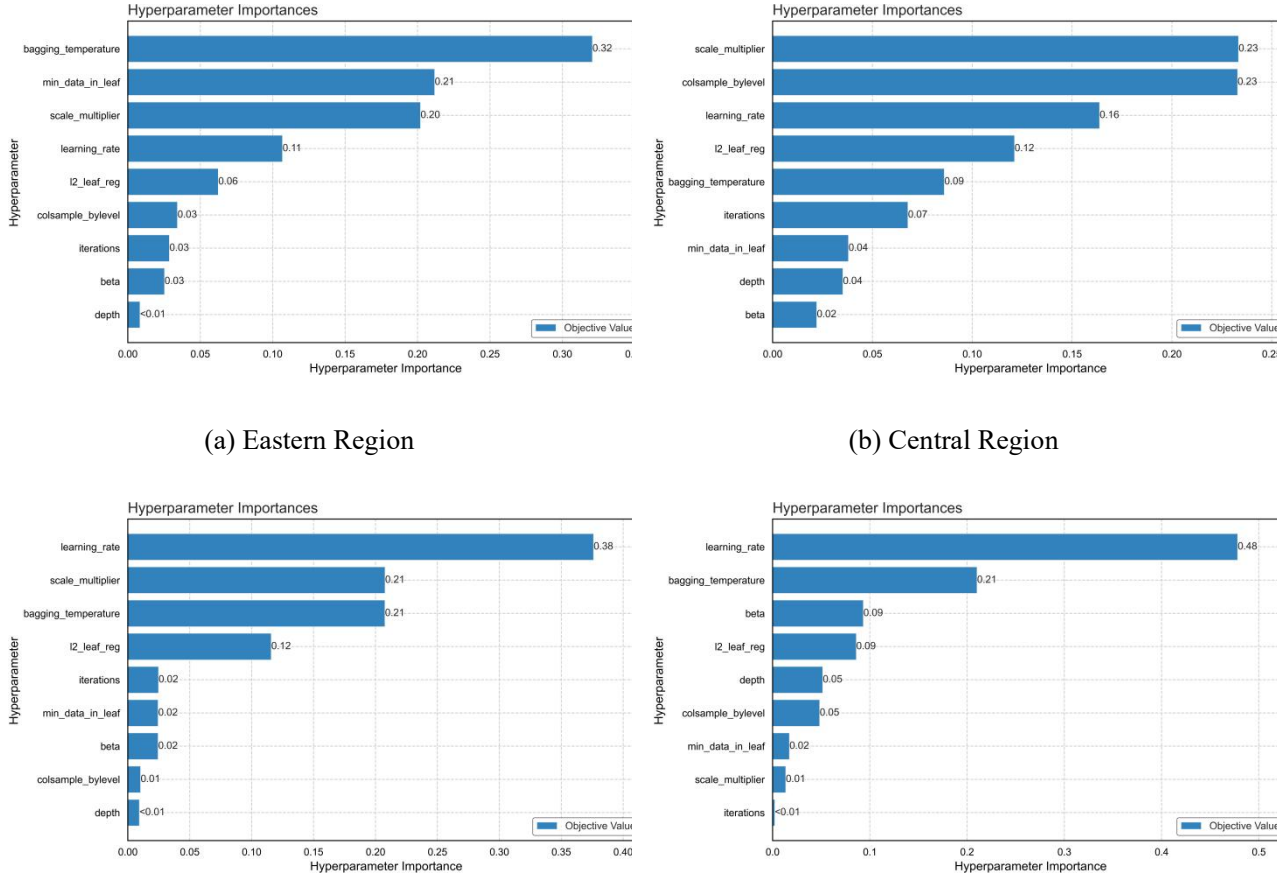
(d) North-Eastern Region

Figure 3. Comparison of optimization history across regions

(Note: Blue dots represent the objective value of a single trial, and the red line represents the best value envelope)

As shown in the optimization history trajectory of Figure 3, the TPE algorithm demonstrates exceptional search efficiency across all regions. The scatter points represent the objective function values from individual trials, while the red line denotes the envelope of the best values explored at each stage.It can be observed that after an initial exploration phase (approximately the first 5-10 trials), the red line exhibits a significant upward leap and converges rapidly to a suboptimal or optimal plateau region. This trajectory intuitively demonstrates that the TPE mechanism effectively reduces the computational overhead of traditional grid search while preventing the model from getting stuck in local optima.

Further examination of the hyperparameter importance bar chart in Figure 4 reveals that the S-DAGB model demonstrates a pronounced adaptive adjustment mechanism in response to data heterogeneity across different regions:



(a) Eastern Region

(b) Central Region

(c) Western Region

(d) North-Eastern Region

Figure 4. Comparison of hyperparameter importance across regions

(1) Validation of the Effectiveness of the A-priori Knowledge Reconstruction Mechanism: As shown by the bar charts for the eastern, central, and western regions in Figure 4, the core parameter scale_multiplier—which controls the strength of SHAP a-priori knowledge injection—exhibited exceptionally high decision weights, ranking third (0.20), first (0.23), and second (0.21)

in importance respectively.This intuitively indicates that in regions with relatively abundant sample information or significant feature distribution differences, the optimization engine adaptively tends to rely on the objective prior weights provided by Stage I. By amplifying the resolution of core impoverishment factors, it enhances overall prediction performance.

(2) Strong Regularization Constraints in Small-Sample Regions: Particularly noteworthy is the Northeast region, which has the most limited sample size (only 1,548 cases). As shown in Figure 4 (Northeast Region), the model does not rely solely on feature reconstruction. Instead, it assigns extremely high weights to the underlying randomness control parameters: bagging_temperature (0.21), dynamic sample weight decay rate beta (0.09), and leaf node penalty term l2_leaf_reg (0.09).This phenomenon clearly demonstrates that when confronted with sparse, small-sample data, the TPE optimization engine intelligently activates and enhances the underlying structural constraints of S-DAGB. This creates a smoothing buffer layer that effectively mitigates the overfitting risks commonly encountered in small-sample scenarios.

4.4. Multi-Region Performance Comparison and Analysis of S-DAGB

To comprehensively evaluate the predictive efficacy of the overall TKDAF framework, this study conducted comparative experiments between the full S-DAGB model and baseline models including Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and LightGBM. Table 2 details the performance of each model across five evaluation metrics in four major regions, while Figure 5 plots the corresponding ROC curves.Experimental results demonstrate that S-DAGB exhibits superior overall performance in multi-region relative poverty prediction tasks:

(1) Model Generalization Analysis Under Small-Sample Constraints (Using Eastern and Northeast Regions as Examples). In the Eastern region with relatively rich sample features and the Northeast region with the most limited sample size, S-DAGB achieved optimal performance across multiple core metrics.Notably, in Northeast China, traditional tree models (e.g., CART decision trees) exhibited varying degrees of performance degradation on small datasets due to insufficient information constraints (CART accuracy reached only 89.47%).In contrast, S-DAGB leveraged prior knowledge injected in Stage I and combined it with regularization strategies selected via TPE. This approach not only maintained high stability but also elevated accuracy to 93.52% and achieved an F1-Score of 95.81%. This demonstrates the framework's effectiveness in mitigating the limitations of deep tree models' generalization capabilities under small-sample conditions.

(2) Precision Optimization Under Severely Imbalanced Data Distribution (Using Central and Western Regions as Examples). In regions like the central and western areas, characterized by high feature heterogeneity and extreme imbalance between positive and negative samples, the baseline teacher model ORGB demonstrates competitive performance in AUC. However, S-DAGB achieves more significant improvements in Precision and F1-Score. For instance, in the central region, S-DAGB achieves a precision of 93.91% (exceeding ORGB's 93.18%); in the western region, precision rises to 91.22%. In practical poverty alleviation applications, traditional prediction models often prioritize high recall, which can lead to elevated false positive rates and

result in inefficient allocation of poverty relief resources.By introducing an Effective Sample Weighting (ENS) mechanism, S-DAGB constructs a smoothing buffer layer at the loss function level, effectively suppressing model bias toward majority classes. This mechanism significantly reduces redundant false positives while maintaining high recall, better aligning with the policy objective of "accurate identification."

Table 2. Comparison Table of Regional Model Performance

Region	Model	Accuracy	Precision	Recall	F1	AUC
Eastern	LR	87.73%	88.64%	97.26%	93.02%	96.51%
	KNN	92.02%	94.10%	96.34%	95.18%	95.44%
	SVM	91.36%	94.83%	94.66%	94.71%	96.56%
	RF	92.02%	94.29%	96.11%	95.17%	96.39%
	CART	89.52%	93.98%	93.19%	93.56%	84.72%
	XGBoost	91.76%	94.33%	95.74%	95.01%	96.55%
	LightGBM	92.15%	94.59%	95.93%	95.25%	97.19%
	ORGB	93.39%	94.86%	97.31%	96.03%	96.59%
	S-DAGB	93.42%	95.17%	97.38%	96.30%	97.23%
Central	LR	88.02%	90.31%	94.97%	92.45%	94.88%
	KNN	89.92%	92.56%	94.39%	93.45%	92.65%
	SVM	89.68%	92.60%	93.96%	93.27%	95.22%
	RF	90.36%	92.77%	94.76%	93.73%	94.63%
	CART	86.13%	91.33%	90.42%	90.84%	81.61%
	XGBoost	89.96%	92.73%	94.25%	93.45%	94.94%
	LightGBM	89.96%	92.58%	94.40%	93.45%	95.49%
	ORGB	91.29%	93.18%	95.57%	94.35%	97.31%
	S-DAGB	91.45%	93.91%	94.90%	94.40%	97.01%
Western	LR	78.45%	80.04%	93.46%	86.03%	92.36%
	KNN	86.13%	88.97%	91.00%	89.96%	89.57%
	SVM	85.86%	87.65%	92.39%	89.93%	93.33%

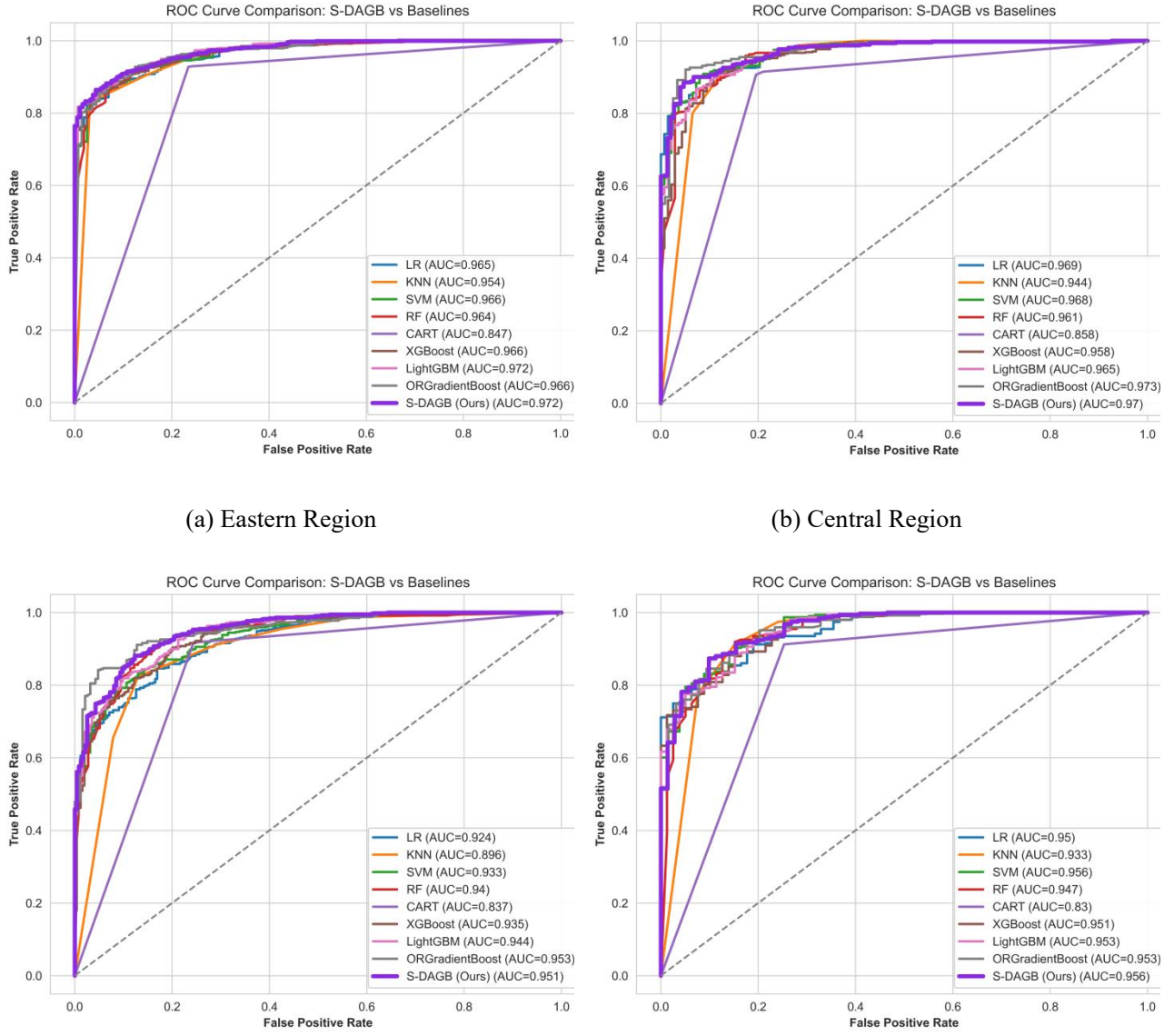
Region	Model	Accuracy	Precision	Recall	F1	AUC
	RF	88.12%	91.00%	91.69%	91.33%	94.02%
	CART	84.01%	89.12%	87.29%	88.16%	83.67%
	XGBoost	87.75%	90.48%	91.73%	91.07%	93.32%
	LightGBM	87.96%	90.82%	91.66%	91.20%	94.39%
	ORGB	89.07%	90.82%	93.49%	92.11%	95.26%
	S-DAGB	89.68%	91.22%	95.25%	92.24%	95.14%
Northeast	LR	87.08%	89.37%	95.94%	93.38%	95.01%
	KNN	92.31%	93.19%	97.48%	94.13%	93.31%
	SVM	92.50%	92.55%	98.54%	94.43%	95.59%
	RF	91.92%	93.51%	96.59%	92.31%	94.72%
	CART	89.47%	93.24%	93.58%	95.01%	83.03%
	XGBoost	90.63%	93.49%	94.80%	95.21%	95.13%
	LightGBM	91.09%	93.60%	95.30%	95.27%	95.26%
	ORGB	93.34%	93.83%	97.70%	95.44%	95.28%
	S-DAGB	93.52%	93.92%	97.78%	95.81%	95.64%

Comparing ROC curves across regions (Figure 5) further illustrates S-DAGB's predictive characteristics. While its curve does not absolutely envelop the baseline teacher model (ORGB) across all thresholds in extremely imbalanced areas like central and western China, it demonstrates two practical advantages:

(1) Remarkable curve smoothness and threshold robustness. Traditional ensemble tree models (e.g., CART and native XGBoost) often exhibit extreme probability distributions in ROC curves due to overfitting local noise when handling sparse or imbalanced data, resulting in "staircase-like" or "jagged" abrupt changes.S-DAGB effectively calibrates the continuity of output probabilities through the TPE depth regularization constraint introduced in Stage II and the ENS weight smoothing mechanism. This smooth ascent trajectory demonstrates the model's exceptional robustness to decision threshold fine-tuning, effectively mitigating drastic fluctuations in predictive performance.

(2) Robust performance in the low false positive rate (FPR) range. Along the left edge of the ROC curve (i.e., the locally high specificity range), S-DAGB's purple curve maintains a

consistently stable and steep upward trend.From the perspective of practical poverty governance, performance in this region is critical: it signifies that S-DAGB can robustly identify genuinely relatively impoverished individuals even under stringent conditions of strict false positive control (i.e., limiting non-impoverished households from occupying policy resources).This strategy of sacrificing a minimal portion of global AUC to achieve decision smoothness and local precision further validates the advanced capabilities of this two-stage knowledge distillation architecture in complex real-world scenarios.



(a) Eastern Region

(b) Central Region

(c) Western Region

(d) North-Eastern Region

Figure 5. ROC Curves of Different Machine Learning Models across the Four Regions

5. Conclusions and Outlook

Based on the 2020 China Family Panel Studies (CFPS) data, this study proposes the Two-stage Knowledge Distillation Adaptive Gradient Boosting Framework (TKDAF) and conducts

empirical research addressing spatial heterogeneity, small sample sizes, and extreme class imbalance in multi-regional relative poverty prediction. Key findings include:

(1) The effectiveness of extracting and reconstructing spatial heterogeneity prior knowledge was validated. By combining the base model with SHAP attribution, this study quantitatively extracted the marginal contributions of poverty-inducing features across regions. Introducing these prior weights into the second stage for feature space reconstruction effectively guided the model to focus on core poverty dimensions while reducing interference from redundant features. (2) Enhanced the generalization capability of deep tree models in small-sample scenarios. Addressing overfitting prone in sparsely sampled regions like Northeast China, the S-DAGB model leveraged TPE adaptive optimization and strong underlying regularization constraints (e.g., leaf node penalties) to effectively suppress overfitting on small datasets while maintaining high classification accuracy. (3) Achieved precision optimization under extremely imbalanced data distributions. Addressing the severe imbalance between positive and negative samples in central and western regions, the model reshapes the loss function by introducing an Effective Sample Number (ENS) dynamic weighting mechanism. Empirical results demonstrate that this mechanism significantly reduces false positive rates while maintaining high recall, better aligning with the practical demand for precise allocation of poverty alleviation resources.

Although this framework demonstrates strong predictive performance in empirical tasks, the following research directions remain open due to objective constraints: (1) Incorporating spatio-temporal dynamics: This study primarily relies on cross-sectional data, failing to fully capture the dynamic vulnerability of relative poverty. Future work could integrate multi-period panel data with spatio-temporal sequence prediction models to better track the temporal evolution of household poverty status. (2) Deep integration of multi-source heterogeneous data: Current input features are primarily confined to micro-level survey data. Subsequent research could explore integrating macro-level geographic information (e.g., nighttime light imagery, regional road networks) with multi-modal data to achieve more granular grid-based poverty identification. (3) Deepening the underlying knowledge distillation mechanism: The prior knowledge transfer in this paper remains at the shallow level of feature weight guidance. Future work could explore deeper knowledge distillation techniques like latent state alignment or logit matching to further expand performance boundaries while controlling model inference complexity.

Author Contributions:

Guanghuang Liu drafted the manuscript, implemented the coding, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

The China Family Panel Studies (CFPS) data utilised in this research were collected and managed by the Institute of Social Science Surveys (ISSS), Peking University. This restricted-access dataset may be obtained by researchers submitting a formal application to ISSS. Details regarding data requests can be found on their official websites:http://www.isss.pku.edu.cn/cfps/.

Conflict of Interest:

The authors declare no conflict of interest.

References

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. Journal of Public Economics, 95(7-8), 476-487.

Bergstra, J., Bardenet, R., Bengio, Y., et al. (2011). Algorithms for hyper-parameter optimization. In Advances in Neural Information Processing Systems (pp. 2546-2554).

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Chawla, N. V., Bowyer, K. W., Hall, L. O., et al. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

Cui, Y., Jia, M., Lin, T. Y., et al. (2019). Class-balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 9268-9277).

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232.

Jean, N., Burke, M., Xie, M., et al. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790-794.

Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (pp. 3146-3154).

Li, X., Zhou, Y., & Chen, Y. (2020). Theory and methods for regional multidimensional poverty measurement. Acta Geographica Sinica, 75(4), 753-768.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

Qian, Y., Wang, C., & Wang, J. (2022). Mutual information and decision tree algorithm for eliminating random consistency. Journal of Shanxi University (Natural Science Edition), 45(5), 1206-1215.

Ravallion, M., & Chen, S. (2007). China’s (uneven) progress against poverty. Journal of Development Economics, 82(1), 1-42.

Shi, Y., Ding, T., Qi, X., et al. (2024). An explainable model for relative poverty identification and early warning. Journal of Shanxi University (Natural Science Edition), 47(1), 155-165.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In Advances in Neural Information Processing Systems (pp. 2951-2959).

Sun, J., & Xia, T. (2019). The evolution of China’s poverty alleviation strategy and post-2020 relative poverty governance. Chinese Rural Economy, (10), 98-111.

Wang, B., Luo, Q., Chen, G., et al. (2022). Differences and dynamics of multidimensional poverty in rural China from multiple perspectives analysis. Journal of Geographical Sciences, 32(8), 1383-1404.

Wang, S., & Zeng, X. (2018). Preliminary exploration of post-2020 poverty issues. Journal of Hohai University (Philosophy and Social Sciences Edition), 20(2), 7-13.

Wang, X., & Feng, H. (2020). China’s multidimensional relative poverty standards post-2020: International experience and policy orientations. Chinese Rural Economy, (3), 2-21.

Zou, W., & Fang, Y. (2011). A dynamic multidimensional study on poverty in China. Economic Research Journal, (12), 42-55.