

Measuring Agricultural Modernization and Diagnosing Development Bottlenecks in China's Provinces: Evidence from high-quality development and new quality productivity

Changrun Chen¹, Lijuan Zhang^{1,*}

School of Artificial Intelligence, Guangzhou Huashang University, Guangzhou, 110870, China

*Email: 464085923@qq.com

Abstract

Key conclusions are as follows: (1) provincial agricultural modernization exhibits pronounced spatial heterogeneity—coastal and major economic provinces generally outperform inland regions in both high-quality development and new quality productivity; (2) model-based indicator scoring reveals clear “knee/fault-line” structures, motivating parsimonious feature-retention thresholds (0.04 for the high-quality development set and 0.05 for the new quality productivity set) while maintaining predictive accuracy; (3) for the lagging cluster (2022 comprehensive score < 1.2), obstacle-degree diagnostics indicate that binding constraints concentrate in green production efficiency, openness-related factor conditions, shared public services, and the digital rural environment rather than a uniform weakness; (4) the stacking ensemble delivers the best overall predictive performance (lower RMSLE and smaller MSE/MAE than any single learner), supporting its use as a monitoring tool for key modernization dimensions. These findings imply that modernization policies should shift from average upgrading to cluster-specific, bottleneck-oriented interventions.

Keywords: high-quality development; new quality productivity; agricultural modernization; stacking ensemble learning; K-means clustering; obstacle-degree analysis

Received July 10, 2025

Revised September 20, 2025

Accepted December 15, 2026

Available Online December 30, 2026

Measuring Agricultural Modernization and Diagnosing Development Bottlenecks in China's Provinces: Evidence from high-quality development and new quality productivity

1. Introduction

China's agricultural modernization goals have continuously evolved in step with national development strategies and the dominant constraints of each era. From the early "four modernizations" agenda and management-oriented reforms, to the launch of internationalization, then to large-scale mechanized production as a defining pathway of modernization (Huang, 2007), the country has now entered a new stage that places high-quality development at the center and treats new quality productive forces as a foundational driver.

Over the past decade, agricultural modernization in China has been propelled largely by scale expansion and technology diffusion, including mechanization and advances in agricultural biochemical technologies. These efforts have supported stable grain output above 1.3 trillion jin and a broader, more diversified supply of agricultural and livestock products. National policy documents also report substantial mechanization progress -- for example, a comprehensive mechanization rate of crop tillage and harvesting reaching 71% (State Council, 2021). As a result, the long-standing challenge of ensuring basic food security for a large population has been substantially alleviated.

Yet a speed- and output-centered narrative is increasingly insufficient for evaluating modernization today. In practice, modernization quality has often been under-discussed and under-measured, despite persistent constraints such as low production efficiency and weak comparative returns. Available policy reporting indicates that agricultural labor productivity remains markedly below that of non-agricultural sectors, about 25.3% by one national account (State Council, 2023). Meanwhile, modernization has coincided with ecological pressures, including farmland occupation and the excessive use of pesticides. This implies that the core challenge is no longer only "producing enough," but "producing well" under resource and environmental constraints—making quality, efficiency, and sustainability central to modernization assessment. Accordingly, China requires a multi-dimensional evaluation system that moves beyond traditional single-dimensional approaches and aligns with contemporary strategic priorities.

Since 2017, "high-quality development" has been established as a major national objective grounded in the new development philosophy. More recently, the central government has further emphasized accelerating the construction of a modern industrial system and developing new productive forces (State Council, 2024). Together, these agendas signal a dual shift: from growth that prioritizes quantity to development that prioritizes quality and balance, and from factor-driven expansion to innovation-driven productivity upgrading.

Against this backdrop, evaluating provincial agricultural modernization through the combined lenses of high-quality development and new productive quality forces is both timely and policy relevant. This study therefore reframes agricultural modernization measurement around these two strategic orientations, aiming to (i) produce a comparable provincial benchmarking tool, (ii) identify development types among lower-performing regions, and (iii) pinpoint binding constraints that can guide targeted interventions. By doing so, we seek to provide an evidence-based diagnostic foundation for promoting more comprehensive, higher-quality, and more innovation-driven agricultural modernization in China.

Essential innovations relative to existing literature are threefold. First, unlike studies that evaluate agricultural modernization via single-dimensional output expansion or rely on a single strategic lens, we explicitly integrate high-quality development and new quality productivity into one unified provincial assessment framework. Second, rather than applying conventional entropy/TOPSIS-style weighting or linear screening, we employ correlation diagnostics plus multi-learner, tree-based importance estimation to accommodate non-linear relationships and high-dimensional indicators. Third, we connect measurement to actionability by coupling K-means typology with Pareto-style obstacle-degree diagnostics, thereby producing cluster-specific “critical few” bottlenecks and policy priorities, and we further introduce a stacking ensemble to strengthen predictive monitoring of core indicators.

Research problems. This manuscript addresses four questions: RQ1: How can provincial agricultural modernization be measured in a way that simultaneously reflects the policy priorities of high-quality development and the emerging concept of new quality productive forces? RQ2: Given a high-dimensional indicator system, how can redundant indicators be screened and indicator importance be learned under potential non-linearity? RQ3: How can low-performing provinces be classified into interpretable development types and diagnosed in terms of binding bottlenecks rather than broad, non-actionable deficits? RQ4: Can ensemble learning improve the robustness and accuracy of monitoring/predicting key benchmark indicators relative to single learners?

2. Literature Review

To align with this study’s objectives, we organize the related scholarship into three streams: (i) high-quality agricultural development, (ii) the concept and measurement of new quality productivity, and (iii) agricultural modernization and its evaluation frameworks.

2.1 Research on High-Quality Agricultural Development

High-quality agricultural development is widely regarded as a core component of national high-quality development, and its conceptual scope has expanded as research deepened after the policy discourse emerged in 2017. Early studies proposed multi-dimensional understandings, for example, emphasizing agricultural economic growth, outcomes, rural social development, and sustainability (He & Song, 2017). Subsequent work stressed standardization of agricultural products, industrial efficiency, the completeness of operating systems, and international competitiveness (Zhong, 2018). Later, scholars further incorporated institutional construction and efficiency-driven mechanisms, building broader analytical frameworks for high-quality agricultural development (Du & Luo, 2021).

In recent years, research has increasingly moved beyond the generic “new development philosophy” lens and examined high-quality agricultural development through context-specific drivers that reflect contemporary transformation pressures, such as the digital economy and rural revitalization. Evidence suggests that digital inclusive finance is positively associated with high-quality agricultural development (Zhang & Wang, 2021). Related discussions also position digital transformation as a pathway to resolve structural constraints, offering policy-oriented diagnoses and solutions (Wang & Yang, 2023). From a regional perspective, studies have proposed that precision

agriculture and green development can function as pivotal levers for high-quality agricultural development under new economic cycles (Hou et al., 2024).

2.2 Definition and Research on New Quality Productivity

New quality productivity has emerged as a policy-oriented construct that highlights innovation-led upgrading under high-quality development. Policy narratives emphasize that it should accelerate the development of productive forces through innovation and modernization (State Council, 2024).

Although the concept is relatively new, domestic discussions have begun to converge on several core elements. A representative definition conceptualizes new quality productivity as innovation-led, oriented toward strategic emerging and future industries, and capable of balancing digital and green transformation so that potential productivity can be converted into realized productivity (Xu et al., 2023). In this framing, “newness” lies not in abandoning traditional productivity components, but in digitally reconfiguring labor, objects of labor, and means of production to fit the characteristics of the current technological era, thereby improving the materialization and effectiveness of productive forces (Xu et al., 2023).

Extant applications of the new quality productivity lens have expanded across multiple domains. Some studies discuss its necessity for building an “agricultural powerhouse” and for restructuring agricultural development pathways (Lin & Dong, 2024). Others examine enabling conditions in state-owned enterprises or sectoral restructuring (Song & Lü, 2024) and explore industry-specific transitions such as cultural and tourism development (Qi & Li, 2024). Empirically, emerging work has attempted to measure new quality productivity at city or provincial levels using methods such as TOPSIS, entropy weighting, Dagum Gini decomposition, kernel density estimation, Moran’s I, and regression-based analyses (Cao & Chen, 2024; Fu & Cai, 2024; Han et al., 2024).

2.3 Research on Agricultural Modernization Development

Since the reform and opening-up period, China’s agricultural modernization has been characterized by path dependence on national conditions, with the definition and priorities of modernization adjusting to changing macro environments. Under the current development stage, agricultural modernization is increasingly described as a system-level transformation that requires demand-driven orientation, supply-side upgrading, and improved benefit distribution, while further promoting marketization, standardization, specialization, specialization, specialization, specialization, specialization, integration, precision, and intelligence (Development Research Center of the State Council, 2020).

A growing body of work treats high-quality agricultural development as an essential stage within agricultural modernization and applies diverse quantitative methods to measure regional progress and its determinants. For example, studies have examined the relationship between the digital economy and high-quality agricultural development using regression frameworks (Wang et al., 2023), assessed coupling coordination between water-resource carrying capacity and high-quality agricultural development using CRITIC weighting and spatial statistics (Peng & Wu, 2023), and explored spatio-temporal patterns and drivers using entropy-weighting and grey relational methods combined with spatial analytical tools (Xiao & Zhang, 2023; Xu et al., 2024).

Despite these advances, research that explicitly integrates high-quality development with new

quality productivity in a unified agricultural modernization assessment remains limited. Moreover, traditional indicator screening and weighting approaches (e.g., regression-based selection and TOPSIS-style procedures) may be less flexible when faced with high-dimensional indicators and non-linear relationships. This motivates the present study to jointly adopt the two strategic lenses and to introduce machine-learning-based indicator screening and ensemble prediction, thereby constructing a more adaptive and reproducible evaluation-and-diagnosis pipeline for provincial agricultural modernization.

2.4 Summary of research gaps and this study's contributions.

Existing studies provide rich discussions on (i) multi-dimensional high-quality agricultural development, (ii) early measurement attempts for new quality productivity at city/provincial levels, and (iii) agricultural modernization evaluation methods (entropy weighting, TOPSIS, coupling coordination, grey relational analysis, and spatial statistics). However, three shortcomings remain salient. First, integration is limited: few studies jointly operationalize high-quality development and new quality productivity within the same agricultural modernization benchmark, weakening cross-province comparability under the new policy discourse. Second, many indicator-screening/weighting routines are sensitive to linearity assumptions and may struggle with high-dimensional indicators and non-linear interactions, reducing robustness when indicators are strongly coupled (especially in the digital domain). Third, policy translation is often incomplete: even when composite indices are constructed, the linkage from “score differences” to cluster-specific binding constraints and implementable policy sequencing is not fully established. To fill these gaps, this manuscript proposes a replicable pipeline—indicator construction → redundancy screening → multi-learner scoring → province benchmarking → clustering typology → obstacle-degree diagnosis → stacking-based prediction—thereby strengthening both measurement validity and policy actionability.

3. Construction of an Indicator System for High-Quality Agricultural Modernization

Building on prior work on agricultural modernization evaluation (Xin & An, 2019; Qiu & Yang, 2019; Hao & Shen, 2022; Chang et al., 2022; Zhang & Wang, 2023; Zheng, 2023), this study constructs an integrated indicator system that links “high-quality development” with “new quality productivity” to assess provincial agricultural modernization.

In contrast to single-dimensional (speed-oriented) assessments, the proposed system emphasizes both quality and efficiency and is designed to be consistent with China’s contemporary strategic priorities, including the Rural Revitalization Strategic Plan (2018–2022).

3.1 Indicator-selection logic and structure

Following the study’s objective, a total of 31 indicators are organized into two overarching dimensions: (i) High-Quality Development and (ii) New Quality Productivity. Measurement and Obstacle Analys...

The first-dimension captures innovation, green production, coordination, openness, and shared benefits; the second dimension captures digital rural development and agricultural science-and-

technology capability.

3.2 Evaluation Index System

Table 1. Evaluation Index System for Agricultural and Rural Modernization

| Objective layer | Tier-1 indicator | Tier-2 indicator | Specific indicator | Unit | Attribute |
|--------------------------|-------------------|-------------------------------------|---|-----------------------|-----------|
| High-Quality Development | Innovation | Innovation input | Total power of agricultural machinery | 10,000 kW | + |
| High-Quality Development | Innovation | Resource-allocation efficiency | Completed fixed-asset investment by rural households | billion yuan | + |
| High-Quality Development | Green development | Green living | Solar water-heater area | 10,000 m ² | + |
| High-Quality Development | Green development | Green production | Pesticide usage | tons | - |
| High-Quality Development | Green development | Green production | Fertilizer usage | tons | - |
| High-Quality Development | Green development | Green production | Water-treatment area | 1,000 hectares | + |
| High-Quality Development | Green development | Green production | Agricultural plastic-film usage | tons | - |
| High-Quality Development | Coordination | Urban-rural dual economic structure | Per-capita consumption expenditure of rural residents | yuan | + |
| High-Quality Development | Openness | Rural labor force | Urban-rural income ratio (rural = 1) | unitless ratio | - |
| High-Quality Development | Openness | Rural labor force | Producer Price Index of agricultural products | Unitless index | + |
| High-Quality Development | Openness | Rural labor force | Share of rural employment in primary industry | % | + |
| High-Quality Development | Openness | Rural labor force | Number of rural laborers | persons | + |
| High-Quality Development | Openness | Cross-border e-commerce | Number of comprehensive pilot zones for cross-border e-commerce | count | + |

| | | | | | |
|--------------------------|------------------------------|------------------------|---|-----------------------|---|
| High-Quality Development | Sharing | Living standard | Completed housing area | 10,000 m ² | + |
| High-Quality Development | Sharing | Infrastructure | Rural water-supply coverage | % | + |
| High-Quality Development | Sharing | Infrastructure | Rural gas penetration rate | % | + |
| High-Quality Development | Sharing | Infrastructure | Length of paved (hardened) roads | km | + |
| High-Quality Development | Sharing | Output level | Agricultural expenditure budget | billion yuan | + |
| High-Quality Development | Sharing | Output level | Agricultural output value | % (as reported) | + |
| High-Quality Development | Sharing | Social security | Number of rural doctors | persons | + |
| New Quality Productivity | Digital rural development | Digital environment | Disposable-income-to-consumption ratio of rural residents | % | + |
| New Quality Productivity | Digital rural development | Digital environment | Number of Taobao villages | count | + |
| New Quality Productivity | Digital rural development | Digital environment | Peking University Digital Inclusive Finance Index | Unitless index | + |
| New Quality Productivity | Digital rural development | Digital infrastructure | Rural electricity consumption | 100 million kWh | + |
| New Quality Productivity | Digital rural development | Digital infrastructure | Rural broadband access users | 10,000 households | + |
| New Quality Productivity | Agricultural S&T development | Network infrastructure | Number of domain names | 10,000 | + |
| New Quality Productivity | Agricultural S&T development | Network infrastructure | Capacity of mobile telephone exchanges | 10,000 households | + |
| New Quality Productivity | Agricultural S&T development | Network infrastructure | Optical-cable line length | km | + |

3.3 Data sources

The indicator data are drawn from China Rural Statistical Yearbook, China Science and Technology Statistical Yearbook, China Environmental Statistical Yearbook, and China Statistical Yearbook (2020–2022), as well as third-party sources including the Alibaba Research Institute, the

Spatial Planning Research Center of Nanjing University, and Cube Data Society.

4. Data Preprocessing and Indicator Scoring

4.1 Data Cleaning

First, missing values in the compiled indicator dataset (see Appendix) were imputed using the column mean.

Second, given substantial and policy-relevant heterogeneity in agricultural modernization across provinces, outlier detection and winsorization were not performed, so as to avoid mechanically truncating meaningful interregional differences.

Third, all metric indicators were standardized in Python using Z-score normalization (code provided in the Appendix) to harmonize units and scales, thereby improving model stability, training efficiency, and comparability across indicators.

4.2 Feature Engineering

To improve readability and facilitate model training, we re-labeled the two indicator groups via label encoding. Importantly, we unify the notation as follows: high-quality development indicators are denoted by $G1 \dots Gm$, and new quality productivity indicators by $X1 \dots Xn$.

4.3 Correlation Analysis and Selection of Predictive Indicators

Following Pearson correlation analysis (Pearson, 1895), this study generated correlation heatmaps for both indicator groups over 2020–2022, corresponding to Figures 1 and 2.

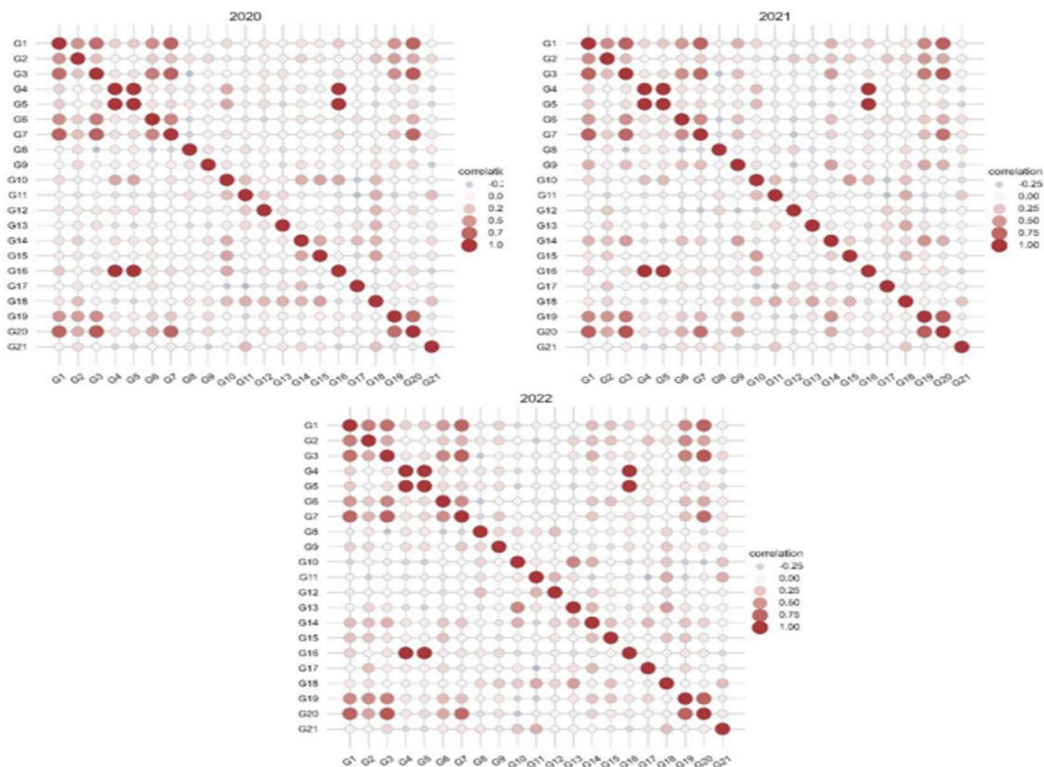


Figure 1: Thermal combination chart of high-quality development indicators in 2020

Across 2020–2022, most high-quality development indicators exhibit weak pairwise correlations, suggesting limited redundancy overall.

However, indicators G4, G5, and G16 remain strongly correlated across years, implying near-duplicate informational content. Accordingly, we retained G4 and excluded G5 and G16 to reduce multicollinearity while preserving interpretability.

Compared with high-quality development, the new quality productivity indicators show materially stronger correlations, consistent with the tighter coupling among digital-infrastructure and e-commerce variables.

Only two highly correlated pairs are salient in 2022 (X4 with X6; X5 with X7). For consistency across years and for model parsimony, we excluded X5 and X6 in all three years and retained X4 and X7.

For predictive benchmarking, we select G1 (agricultural output value) as the core outcome proxy for high-quality development, given its integrative reflection of mechanization, ecological conditions, and production efficiency/benefits.

Similarly, we select X1 (number of Taobao villages) as the key proxy for new quality productivity, because it directly captures rural e-commerce development and aligns with the “technology-driven” emphasis of the new quality productivity concept.

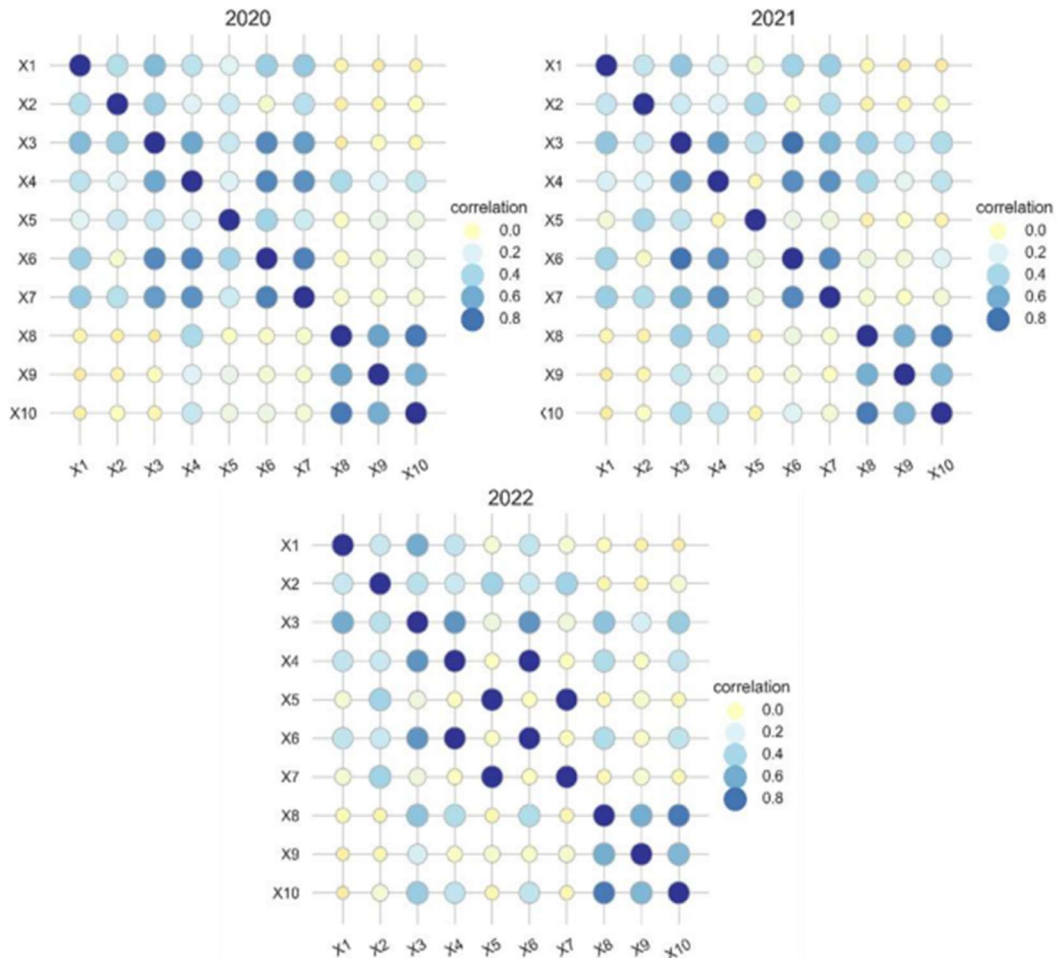


Figure 2: Thermodynamic combination chart of new quality productivity indicators in 2022

4.4 Indicator Scoring Results Based on Multiple Models

Using the selected base learners, this study scores the remaining indicators in each dimension by treating G1 and X1 as benchmark targets. Final indicator scores are computed as the average importance/score across learners, and the results are reported in Figures 3–4.

The thresholds are data-driven rather than arbitrary. For each dimension, we (i) compute learner-specific feature-importance vectors and average them across the seven tree-based learners to obtain a stable importance score per indicator; (ii) sort the averaged scores in descending order and examine the score–rank curve; (iii) identify the “knee” (fault-line) where marginal gains diminish sharply, using a curvature-based elbow rule (the rank at which the second-order difference of the sorted score curve is maximized). These yields cutoffs close to 0.04 (high-quality development) and 0.05 (new quality productivity), consistent with the visible separation in Figures 3–4. We then conduct a predictive–parsimony check by re-training the prediction models under a grid of candidate thresholds around each knee ($\pm 20\%$) and selecting the smallest feature set whose cross-validated RMSLE is within one standard error of the minimum. Under this criterion, 0.04 and 0.05 remain preferred because they reduce dimensionality while preserving out-of-sample accuracy and interpretability.

(1) Scores of High-Quality Development Indicators

A pronounced “fault-line” pattern appears at 0.04, with G2, G7, G9, G19, and G20 lying above this cutoff and exhibiting substantially higher importance for predicting G1.

To enhance predictive efficiency without sacrificing key information, only indicators with scores > 0.04 were retained for subsequent prediction within the high-quality development dimension.

(2) Scores of New Quality Productivity Indicators

Relative to Figure 4, the scoring distribution is more concentrated: X3 emerges as the dominant indicator, and a cutoff at 0.05 yields the exclusion of X8, X9, and X10 (below the fault line) to streamline the feature set for prediction.

(3) Visual Analysis

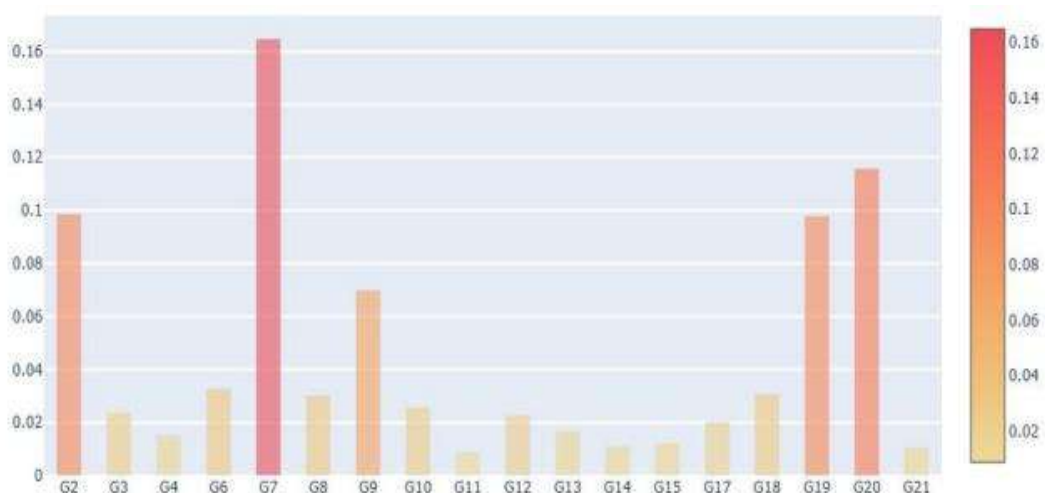


Figure 3: Score situation of high-quality development indicators

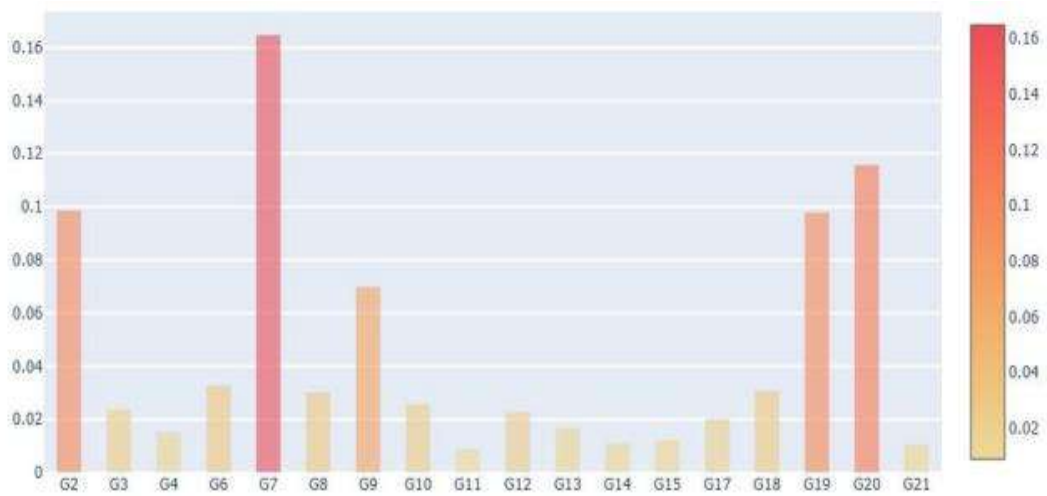


Figure 4: Scores of new quality productivity indicators

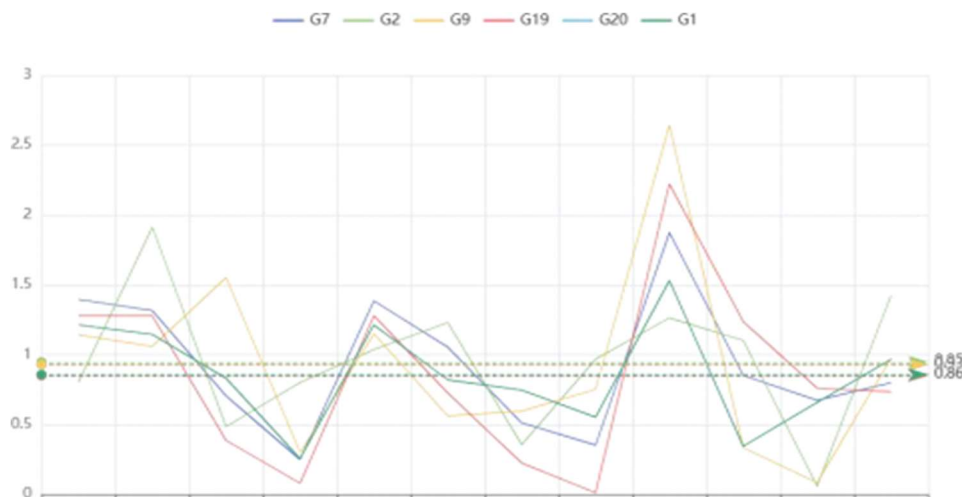


Figure 5: Distribution of key G-category indicators in each province in 2022.

Using 2022 data for eastern coastal provinces (excluding Taiwan due to data unavailability), we visualize the benchmark indicator G1 together with the retained high-importance G-indicators. The plot suggests clear cross-province dispersion and clustering, with Shandong exhibiting the strongest overall performance and Fujian the weakest.

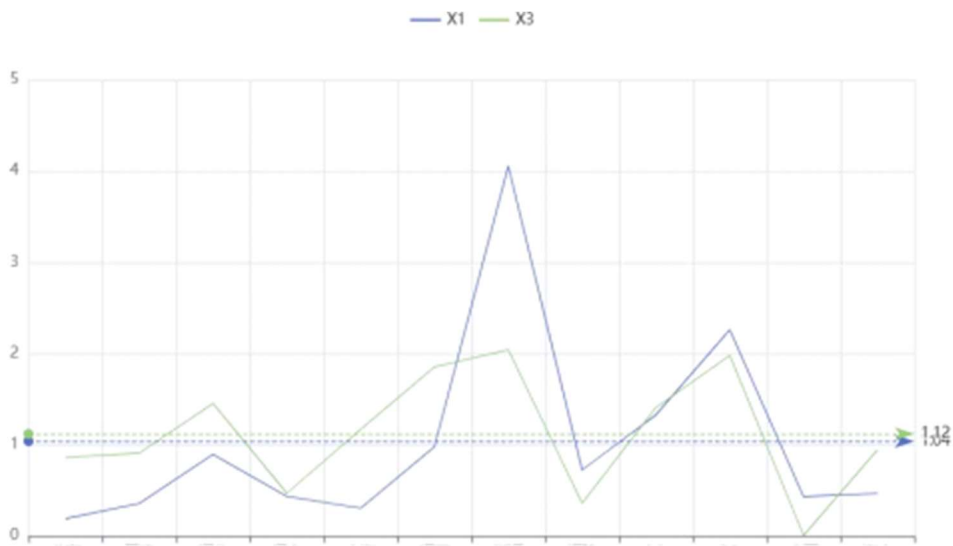


Figure 6: Distribution of key X-category indicators in each province in 2022.

Analogously, the visualization based on X1 and the key X-indicator(s) indicates that Zhejiang performs best and Guangxi performs worst, while cross-indicator dispersion is relatively modest.

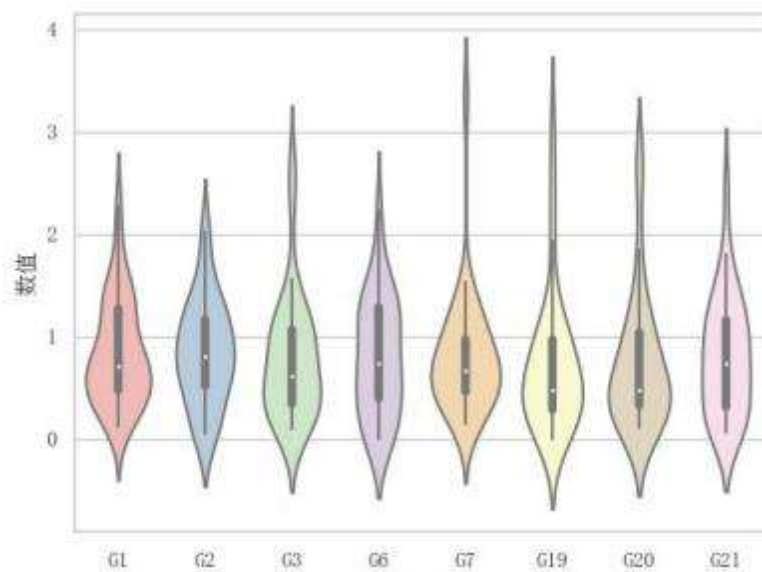


Figure 7: Distribution of various G-class indicators.

Most G-indicators lie within a relatively narrow distribution band (roughly 0 to 1.5), while G7, G19, and G20 show longer tails, indicating greater dispersion and potential outliers—consistent with uneven provincial development in these sub-dimensions.

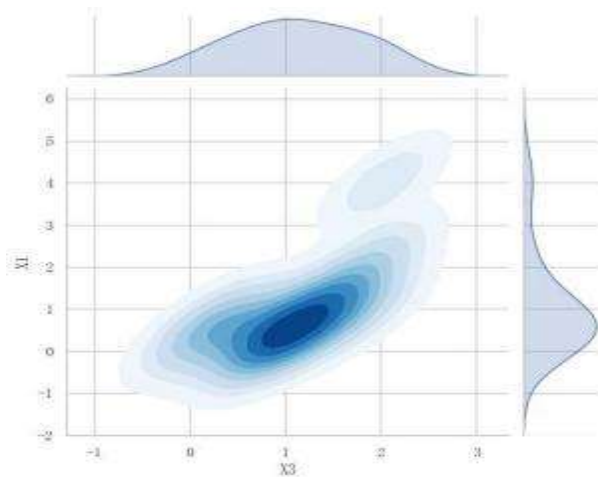


Figure 8: Distribution of the X3 indicator and the X1 indicator

X3 appears relatively balanced across provinces, whereas X1 is concentrated at lower levels for most provinces, with only a small subset exhibiting high performance—suggesting that “Taobao village” development remains spatially uneven.

(4) Provincial Scores (2020–2022)

Provincial scores for high-quality development, new quality productivity, and the comprehensive level were computed by combining indicator values with their learned scores/weights, producing the score profiles shown in Figures 9–12.

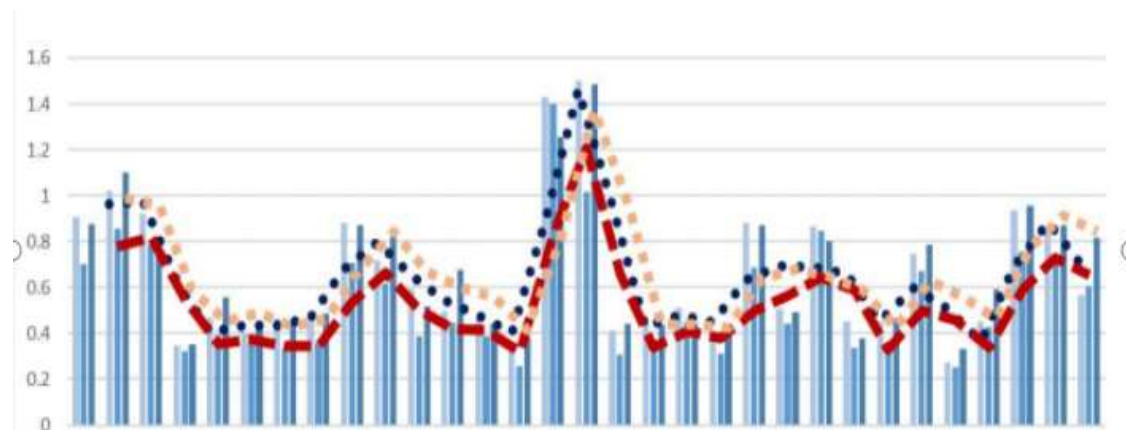


Figure 9: Scores of high-quality developments in each province in each year

High-quality development displays noticeable spatial patterning. The moving-average trend suggests the highest overall level in 2020 and the weakest in 2021, plausibly reflecting widespread disruption to rural production activities during the COVID-19 period.

New quality productivity is highest in major southeast coastal provinces/municipalities (e.g., Jiangsu, Guangdong, Zhejiang, Shanghai) and Beijing, aligning with their strong e-commerce ecosystems and reform-oriented trade intensity.

Temporal variation is also evident: some provinces improve steadily, while others experience sharp declines following the epidemic shock.

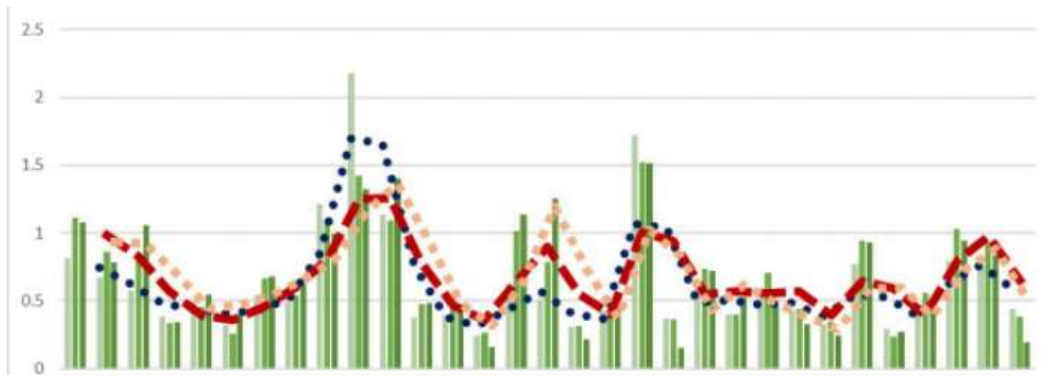


Figure 10: Annual new quality productivity scores of each province

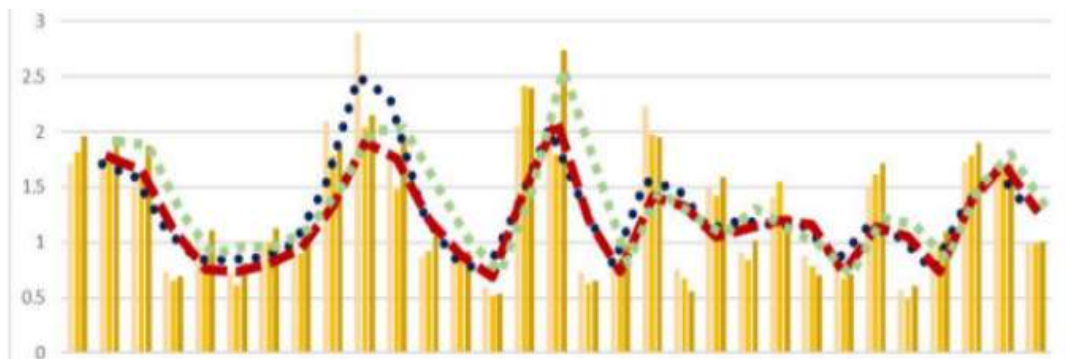


Figure 11: Comprehensive scores of each province in each year

The comprehensive trend line indicates heterogeneous epidemic impacts across provinces; for example, Jiangsu shows a persistent decline, while Henan exhibits a sustained rise in the comprehensive score. To improve clarity and journal-style presentation, Table 2 is rewritten as follows (content unchanged):

Table 2: Comparison of Comprehensive Scores by Province and Year.

| Year | Top 5 provinces/municipalities | Bottom 5 provinces/municipalities |
|------|---|---|
| 2020 | Jiangsu; Guangdong; Shanghai; Shandong; Henan | Shaanxi; Jiangxi; Hubei; Liaoning; Shanxi |
| 2021 | Shandong; Jiangsu; Guangdong; Beijing; Henan | Shaanxi; Jiangxi; Liaoning; Hubei; Shanxi |
| 2022 | Henan; Shandong; Jiangsu; Beijing; Guangdong | Jiangxi; Guangxi; Shaanxi; Hubei; Shanxi |

Across 2020–2022, Jiangsu, Guangdong, Shandong, and Henan appear repeatedly among the top performers, while Shaanxi, Jiangxi, Hubei, and Shanxi repeatedly appear among the bottom performers.

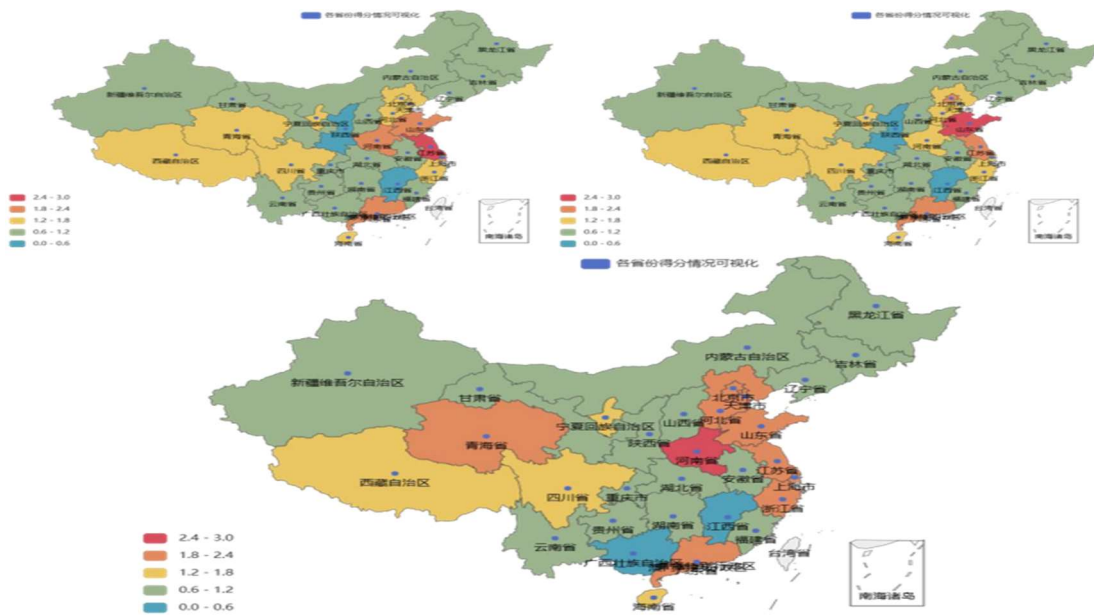


Figure 12: (retain original figure number and title in the manuscript)

The spatial visualization further suggests that provinces in the more economically developed southeastern coastal belt tend to exhibit higher comprehensive scores.

5. Analysis and Suggestions on the Obstacles to the Development of Agricultural Modernization in Provincial Regions

5.1 Province Clustering Based on K-means

Building on the composite evaluation results reported in the previous section, this study further investigates the binding constraints that impede agricultural modernization in relatively low-performing provincial regions. To improve interpretability and policy relevance, we focus on provinces whose composite score in 2022 is below 1.2, because provinces above this threshold are more likely to have entered a comparatively stable stage (or have already formed development models suited to local conditions). Given space constraints, the analysis is conducted for 2022 only.

To identify provinces with similar developmental bottlenecks, K-means clustering was applied to the selected provinces, and the silhouette coefficient was used to determine the optimal number of clusters. The clustering outcomes are summarized in Table 3.

Table 3 Cluster Results (2022, composite score < 1.2)

| Cluster 1 | Cluster 2 | Cluster 3 |
|----------------|-----------|-----------|
| Shanxi | Fujian | Guizhou |
| Inner Mongolia | Jiangxi | Yunnan |
| Liaoning | Hubei | Shaanxi |
| Jilin | Hunan | Gansu |
| Heilongjiang | Guangxi | |
| Anhui | Chongqing | |

Next, Pareto distribution maps were produced for each cluster to examine indicator importance from the perspectives of frequency and cumulative contribution, as shown in Figures 13–15. In the Pareto chart, the left y-axis denotes indicator frequency, while the right y-axis represents cumulative contribution. Following the “critical few” logic, indicators with cumulative contribution below 40% are treated as hindering indicators (i.e., the most salient obstacles), whereas those at 40% or above are treated as development indicators.

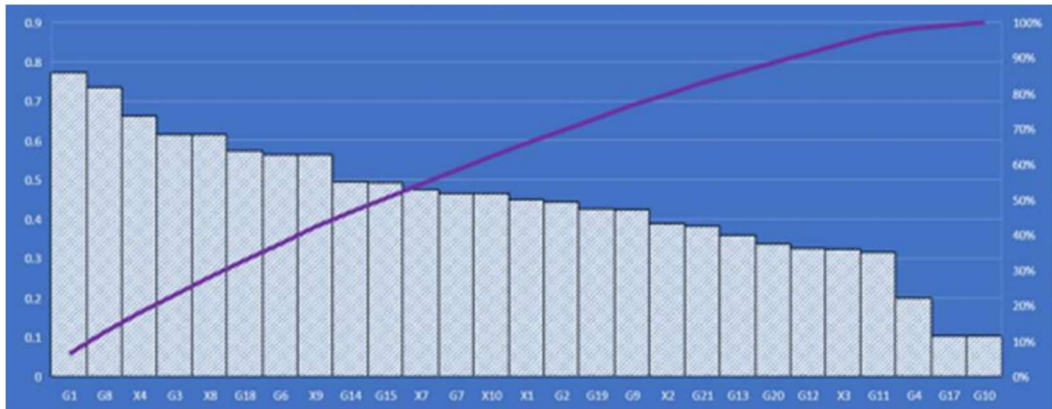


Figure 13: Pareto distribution map of indicators for the first category of provinces

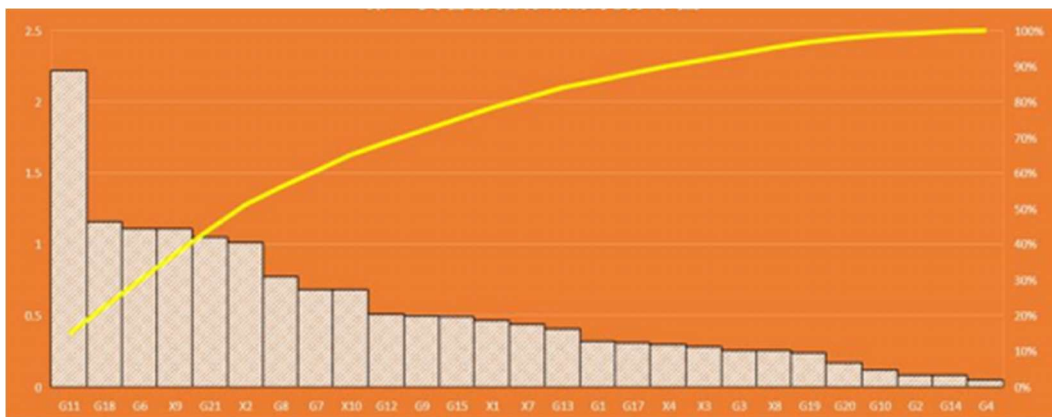


Figure 14: Pareto distribution map of indicators for the second category of provinces

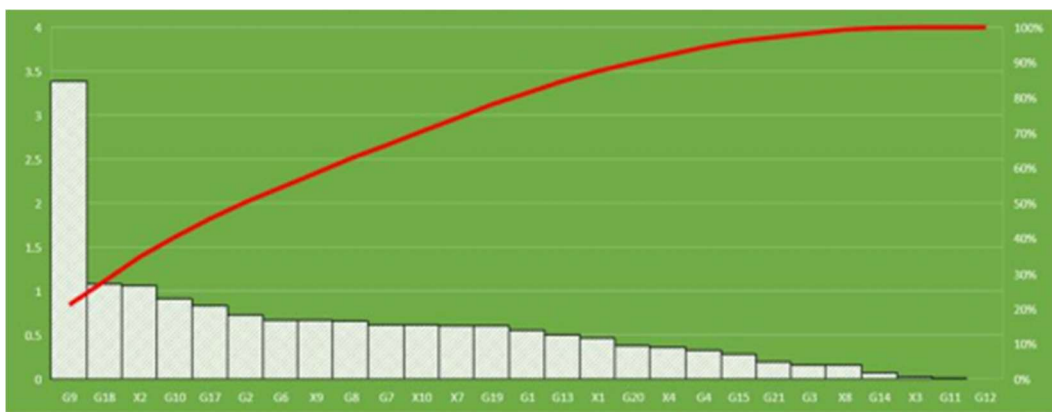


Figure 15: Pareto distribution map of indicators for the third category of provinces

5.2 Analysis and Suggestions on the Degree of Obstruction for the First Type of Provinces

This section focuses on Cluster 1 provinces (Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, and Anhui). We first map each hindering indicator to its corresponding objective layer, Tier-1 indicator, Tier-2 indicator, and specific indicator, thereby forming a structured basis for targeted policy design. The mapping is reported in Table 4.

Table 4 Indicator Correspondence Table (Cluster 1)

| Objective Layer | Tier-1 Indicator | Tier-2 Indicator | Specific Indicators | Hindrance indicator |
|-------------------------------|---------------------------|---|--|---------------------|
| High-Quality Development | Green Development | Green Production | Pesticide Usage Agricultural Producer Price Index | G4 G10 |
| | Openness | Rural Labor Force Cross-border E-commerce Infrastructure | Rural Employment Share in Primary Industry Number of Agricultural Workers Number of Cross-border E-commerce Pilot Zones Length of Paved Roads | G11 G12 G13 G17 |
| | Shared Benefits | Social Security | Number of Rural Doctors Disposable Income-to-Consumption Ratio of Rural Residents | G20 G21 |
| New Quality Productive Forces | Digital Rural Development | Digital Rural Development Environment | Peking University Digital Inclusive Finance Index Rural Electricity Consumption | X1 X3 |

From Table 4, the dominant constraints for Cluster 1 provinces concentrate on four dimensions: green development, openness, shared benefits, and digital rural development. Based on the indicator meanings and the broader policy context, the following targeted recommendations are proposed.

(1) Green development

- Improve fertilizer and pesticide use efficiency by upgrading agricultural machinery and promoting precision application aligned with crop growth stages; precision irrigation can further reduce nutrient loss.
- Increase the share of organic fertilizers and promote integrated organic–chemical fertilization, supported by extension services that communicate the risks of excessive chemical use and the agronomic benefits of organic inputs.
- Develop crop-specific fertilization formulas based on soil testing, and scale soil-test–based fertilization to guide field practice (Ma, 2015).

(2) Openness

Cluster 1 provinces face openness-related constraints associated with rural labor supply and

structural employment dependence. Recommended actions include:

- Expand farmers' income channels by improving the rural operational system and scaling diversified new agricultural business entities, while strengthening the dissemination of pro-farmer policies.
- Accelerate land transfer and moderate-scale operation, using mechanization and improved farming techniques to offset labor shortages.
- Strengthen local talent pipelines, including incentives for college graduates to work in (or return to) their hometowns, accompanied by improvements in supporting infrastructure and rural service capacity (He, 2014).
- Deepen the coverage and usage of digital inclusive finance to improve household financial access and help narrow income disparities among rural laborers (Liu & Han, 2024).

(3) Sharing level

Constraints at the sharing level are linked to infrastructure, rural public services, and social security, which jointly shape residents' welfare and the inclusiveness of modernization gains. Thus:

- In line with local specialty industries, scale livestreaming and e-commerce-based agricultural assistance, using digital marketing to strengthen rural industrial resilience.
- Increase investment in rural public services (healthcare, education, and basic health), and use policy guidance and fiscal support to strengthen living for rural residents and laborers (Liu, Meng, & Niu, 2024).
- Improve bottom-line social protection capacity and strengthen the governance of poverty-prevention mechanisms to reduce the risk of relapse into poverty.

(4) Digital rural development

Digital rural development is hindered by the digital infrastructure and usage environment. Accordingly:

- Expand rural broadband and optical-fiber coverage and improve network access capacity, accelerating progress toward near-universal rural connectivity.
- Promote agricultural digitalization and adoption of basic digital tools, enabling rural e-commerce and improving the efficiency of rural production-circulation-marketing chains.

In summary, the obstacle diagnosis suggests that lagging provinces are constrained by a small set of "critical" factors rather than by uniformly weak performance across all dimensions. Therefore, policy design should shift from generic capacity building to cluster-specific prioritization, focusing resources on the most binding constraints identified by the Pareto-based obstacle analysis. This "measurement-classification-obstacle identification-targeted recommendation" pipeline not only strengthens the interpretability of the provincial benchmarking results but also improves the actionability of policy interventions. The next chapter concludes by summarizing the study's methodological, empirical, and policy contributions, and by outlining limitations and future research directions.

6. Conclusion and Policy Implications

Contributions. This study makes three contributions. First, methodologically, it constructs an integrated provincial evaluation-and-diagnosis pipeline that combines indicator-system design, correlation-based screening, and machine-learning-based scoring and prediction. Second,

empirically, it benchmarks interprovincial heterogeneity in agricultural modernization under the dual lenses of high-quality development and new quality productivity. Third, from a policy perspective, it translates composite scores into cluster-specific obstacle profiles that support targeted intervention prioritization.

This paper develops a provincial-level assessment and diagnostic framework for China's agricultural modernization by integrating high-quality development and new quality productivity into a unified indicator system. Using multi-source panel data for 2020–2022, the study proceeds from indicator construction and screening to composite scoring, spatial comparison, cluster-based typology, and obstacle diagnosis, and further introduces a stacking ensemble to improve the prediction of benchmark indicators. Overall, the proposed framework forms a replicable closed loop—measurement, classification, obstacle identification, and targeted recommendation—that can be used for interprovincial benchmarking and policy sequencing.

6.1 Main findings

First, the composite evaluation results reveal substantial spatial heterogeneity in provincial agricultural modernization. Provinces in the more developed coastal belt generally achieve higher comprehensive scores, whereas several inland provinces remain persistently lagging. The inter-year comparison across 2020–2022 also suggests that the epidemic period coincided with observable performance fluctuations, indicating that modernization quality and digital productivity can be sensitive to shocks in rural production and circulation systems.

Second, the correlation diagnostics show that redundancy differs across dimensions: high-quality development indicators are mostly weakly correlated, while new quality productivity indicators exhibit stronger coupling among digital infrastructure and e-commerce variables. This justifies screening highly correlated indicators before model-based scoring, improving parsimony and mitigating multicollinearity risks in subsequent modeling.

Third, the clustering and obstacle-degree diagnosis provide a policy-actionable view of lagging provinces. For provinces with lower comprehensive scores, constraints do not appear as a uniform “overall weakness” but as development types with distinct bottleneck structures. The “critical few” obstacles concentrate in areas such as green production efficiency, openness-related labor and infrastructure conditions, shared public services, and the digital rural environment. This implies that one-size-fits-all interventions are unlikely to be efficient; targeted policies should be matched to cluster-specific binding constraints.

Fourth, the stacking ensemble improves predictive performance relative to single learners in forecasting benchmark indicators, achieving lower error metrics (e.g., RMSLE, MSE, and MAE). This demonstrates that ensemble learning can serve as a robust auxiliary tool for monitoring key modernization dimensions when indicator sets are high-dimensional and relationships are potentially non-linear.

6.2 Policy implications

The implications below follow directly from the empirical findings: spatial heterogeneity supports differentiated provincial pathways; cluster-based obstacle diagnostics support

bottleneck-oriented governance; and the ensemble results support data-driven monitoring for mid-course correction.

The findings suggest that policy design should prioritize precision governance rather than average upgrading. For green development constraints, reducing pesticide and fertilizer intensity requires integrated measures that combine soil testing, precision application, and extension services (Ma, 2015). For openness constraints, policies should jointly address labor structure, land transfer, mechanization substitution, and inclusive finance to expand income channels and improve factor mobility (He, 2014; Liu & Han, 2024). For sharing constraints, strengthening rural public services—especially healthcare capacity, including rural doctors—remains essential to ensure inclusive modernization gains (Liu, Meng, & Niu, 2024). For digital rural development constraints, infrastructure expansion should be coupled with adoption-oriented interventions (training, platform access, and digital business models) to translate connectivity into productivity. Implementation can adopt a “sequence-and-monitor” approach: (i) prioritize interventions targeting each cluster’s top obstacle indicators (the “critical few” below 40% cumulative contribution); (ii) align resources with these obstacles (precision input reduction for green production; labor mobility/land transfer and mechanization for openness; rural healthcare staffing and service provision for sharing; broadband/optical fiber plus adoption training for digital rural development); (iii) use the stacking ensemble as an early-warning and tracking tool to periodically update G1 and X1 and evaluate whether bottlenecks are relaxing.

6.3 Limitations and future research

This study has several limitations. First, the analysis focuses on 2020–2022 and may not fully capture longer-run structural changes. Future research could extend the time span and evaluate dynamic transitions before and after major policy nodes. Second, although multi-source data were combined, indicators may still be affected by reporting differences across provinces. More granular micro-data (e.g., county-level or household/firm-level evidence) would help validate mechanisms and improve causal interpretability. Third, the current obstacle diagnosis adopts a Pareto-style threshold to identify key constraints. Future work could compare alternative diagnostic rules and explicitly incorporate spatial dependence to account for regional spillovers and policy diffusion.

In sum, this paper provides an integrated measurement and diagnostic toolkit for provincial agricultural modernization under the dual strategic orientation of high-quality development and new quality productivity. By combining indicator-system construction, machine-learning-based scoring, typology identification, and obstacle diagnosis, the proposed framework supports evidence-based prioritization and can be adapted to other regions or sectors facing similar modernization transitions.

References

- Cao, Q., Chen, R., & Ma, N. (2024). China’s new urban industrial productivity is calculated and the regional difference research. *Journal of Lingnan*, 1–16. (Advance online publication).
<http://kns.cnki.net/kcms/detail/44.1005.C.20240425.1635.002.html>

- CSDN. (2022, February 21). LightGBM principle. *CSDN*.
- CSDN. (2023, September 19). CatBoost: Easily understood easily increase performance of machine learning. *CSDN*.
- Development Research Center of the State Council, Rural Economy Research Department, Research Group. (2021). The connotation characteristics and evaluation system of agricultural and rural modernization in the new development stage. *Reform*, (09), 1–15.
- Du, Z., Luo, Q., & Yang, X. (2021). The connotation characteristics, development predicaments and realization paths of high-quality agricultural development: A literature review. *Journal of the Management Cadre College of the Ministry of Agriculture and Rural Affairs*, (04), 14–25.
- Fang, K., Wu, J., Zhu, J., et al. (2011). Review of random forest method research. *Statistics and Information Forum*, 26(03), 32–38.
- Fu, L., & Cai, Y. (2024). New quality productivity in urban areas of China: Temporal evolution, group characteristics and development strategies. *Industrial Economic Review*, 1–18. (Advance online publication). <https://doi.org/10.19313/j.cnki.cn10-1223/f.20240428.001>
- Ghousam, Y. (2017). *Titanic top 4% with ensemble modeling*. Kaggle. <https://www.kaggle.com/code/yassineghousam/titanic-top-4-with-ensemble-modeling>
- Gu, P. (2021). *Based on LightGBM used-car prices forecast* [Dissertation, Shandong Normal University]. <https://doi.org/10.27280/> (as provided: “dc nki. Gdsu. 2021.001186”)
- Han, W., Zhang, R., & Zhao, F. (2024). New mass productivity level is calculated with China’s economic growth momentum. *Quantitative Technical Economics*, 1–22. (Advance online publication). <https://doi.org/10.13653/j.cnki.jqte.20240418.001>
- Hao, X., & Shen, Z. (2022). Assignment of rural electricity to rural revitalization of effectiveness evaluation and obstacle factor analysis. *Journal of Xi’an University of Finance and Economics*, 35(5), 40–52.
- He, H., Song, L., & Li, G. (2017). Time and space differences of China’s agricultural economic growth quality research. *Journal of Economists*, (7), 87–97.
- He, Z. (2014). *Research on the impact of rural labor outflow on rural governance* [Dissertation, Chongqing University].
- Hou, G., Zhang, Z., & Dong, S. (2024). New quality productivity empowering high-quality development of Northeast agriculture: Theoretical logic, key issues and practical paths. *Hunan Social Sciences*, (01), 69–76.
- Huang, P. (2007). The course and development innovation of agricultural modernization in China. *Research on Agricultural Modernization*, (02), 129–134.
- Kaggle. (2019). *All about the metric: RMSLE* (NANASHI). Kaggle. <https://www.kaggle.com/competitions/ashrae-energy-prediction/discussion/113064>
- Kaggle. (2020). *Ensemble learning techniques tutorial* (Pavan Sanagapati). Kaggle. <https://www.kaggle.com/code/pavansanagapati/ensemble-learning-techniques-tutorial>
- Kaggle. (2022). *A statistical analysis & ML workflow of Titanic* (Masum Rumi). Kaggle. <https://www.kaggle.com/code/masumrumi/a-statistical-analysis-ml-workflow-of-titanic#Part-4:-Statistical-Overview>
- Lin, W., & Dong, X. (2024). To the new productivity lead several thinking of the construction of the agricultural power. *Journal of Nanjing Agricultural University (Social Science Edition)*, 1–10. (Advance online publication). <https://doi.org/10.19714/j.cnki.1671-7465.20240430.001>

- Liu, G., Meng, H., & Niu, H. (2024). Practical predicaments and path choices: The role of the state in alleviating rural social security difficulties in China in the new era. *Rural Economy*, (02), 103–113.
- Liu, X., & Han, Q. (2024). Digital inclusive finance and rural labor wage income: An empirical analysis from the perspectives of total income and income gap. *Shandong Social Sciences*, (01), 183–192.
- LightGBM. (n.d.). *Welcome to LightGBM's documentation!* LightGBM. <https://lightgbm.com>
- Ma, B. (2015). *Research on the low-carbon development of modern large-scale agriculture in Heilongjiang Province* [Dissertation, Northeast Agricultural University].
- National Development and Reform Commission. (2024, March 15). The new development of productivity is the inner requirement to promote the development of high quality. *NDRC*. https://www.ndrc.gov.cn/wsdwhfz/202403/t20240315_1364971.html
- Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58, 240–242.
- Peng, S., & Wu, X. (2023). China water resources carrying capacity and the agricultural development coupling coordination study. *High Quality People of the Yangtze River*, (10), 52–59.
- Qi, R., & Li, F. (2024). New mass productivity can assign high quality tour industry research. *Journal of Yantai University (Philosophy and Social Sciences Edition)*, 1–12. (Advance online publication). <https://doi.org/10.13951/j.cnki.issn1002-3194.2024.0411.02>
- Qiu, Z., & Yang, Y. (2019). The development and countermeasures of rural e-commerce in Heilongjiang Province under the background of “Internet +”. *Business Economic Research*, (02), 145–148.
- Shi, J., & Zhang, J. (2019). Load forecasting method based on multi-model fusion stacking ensemble learning approach. *Proceedings of the CSEE*, 39(14), 4032–4042. <https://doi.org/10.13334/j.0258-8013.pcsee.181510>
- Song, D., & Lyu, M. (2024). Dialectical thoughts on new quality productive forces and high-quality development of state-owned enterprises. *Social Sciences Front*, (05), 256–261.
- State Council. (2021, February 21). On comprehensively advancing rural revitalization of the Central Committee of the Communist Party of China and the State Council to speed up the agricultural modernization. *Chinese Government Network*.
- State Council. (2023, March 15). We will accelerate the development of agricultural power: Promoting the modernization of agriculture and rural areas. *Chinese Government Network*. https://www.gov.cn/xinwen/2023-03/15/content_5746861.htm
- State Council. (2024, March 12). The government work report: On March 5, 2024 in the 14th session of the National People's Congress second conference. *Chinese Government Network*. https://www.gov.cn/gongbao/2024/issue_11246/202403/content_6941846.html
- Wang, J., Shi, M., & Li, Z. (2023). Digital drive agricultural high quality economic development: The internal mechanism and empirical evidence. *Journal of Xi'an University of Finance and Economics*, 4(4), 78–89.
- Wang, Q., & Yang, J. (2023). Digital new qualitative research productivity and the high quality of agriculture development in China. *Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition)*, 52(6), 61–72.

- Xiao, L., & Zhang, Y. (2023). The spatio-temporal characteristics and influencing factors of high-quality agricultural development in the Northeast Urban Agglomeration. *Journal of South China Agricultural University (Social Sciences Edition)*, 22(06), 74–87.
- Xin, L., & An, X. (2019). Evaluation system construction and development of agricultural high quality measure analysis. *Journal of Economic Aspect*, (5), 109–118.
- Xu, J., & Yang, Y. (2018). Ensemble learning methods: Research review. *Journal of Yunnan University (Natural Science Edition)*, 40(06), 1082–1092.
- Xu, Z., Liu, K., & Lin, Z. (2024). The new era of agricultural development level of high quality evaluation, based on the supply and demand side perspective. *Journal of Dalian University of Technology (Social Science Edition)*, (01), 52–61.
- Xu, Z., Zheng, L., & Cheng, M. (2023). The intrinsic logic and practical conception of new quality productivity empowering high-quality development. *Contemporary Economic Research*, (11), 51–58.
- Ye, X., Cheng, Y., & Development Research Center of the State Council, Rural Economy Research Department, Research Group. (2021). The connotation characteristics and evaluation system of agricultural and rural modernization in the new development stage. *Reform*, (09), 1–15.
- Yassine Ghouzam. (2017). *Titanic top 4% with ensemble modeling*. Kaggle. <https://www.kaggle.com/code/yassineghouzam/titanic-top-4-with-ensemble-modeling>
- Zhang, H., & Wang, L. (2023). Western region digital measure and promote rural development level path. *Journal of East China Economic Management*, (11), 70–78.
- Zhang, H., & Wang, Y. (2021). Research on the convergence of digital inclusive finance and the high-quality development level of agriculture. *Financial Theory and Practice*, (01), 9–18.
- Zheng, J. (2023). Promoting agricultural modernization with new quality productivity: Theoretical logic and development path. *Price Theory and Practice*, (11), 31–35.