Frontiers in Economics and Management Volume 6 Issue 7, 2025
ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

A Hybrid PCA-Stacking Framework for Multidimensional Assessment of Development Trajectories: Evidence from China's Modernization Process

Hanrui Wang, Yile Wang*

School of Economics, Beijing Technology and Business University, Beijing 102488, China

*Corresponding Author

Abstract

China's multidimensional modernization since 2000 necessitates comprehensive assessment frameworks. This study constructs a novel China Development Index (CDI) through integrated analytical approaches. Principal Component Analysis first distilled 14 indicators across economic, social, and governance domains into two dominant components explaining 94.39% cumulative variance: an innovation-education-governance nexus (84.88% weight) and economic fundamentals (15.12% weight). Using Nonlinear Autoregressive Neural Network (NAR) neural networks optimized via Levenberg-Marquardt algorithms (12 hidden neurons, 2-step delays), secondary indicators like service sector growth were forecasted with 0.026 MSE. Stacking fusion modeling combining Linear Regression, k-Nearest Neighbors (KNN), and Random Forest base learners then projected primary indicators, achieving 93.02 Mean Squared Error (MSE) – 7.2% lower than individual models. The entropy weight-variation coefficient method synthesized these projections into the CDI, revealing three historical phases: Founding (1949-1978), Reform (1978-2000), and Modernization (2000-2022). Longitudinal analysis demonstrates accelerated development post-2000, with projections to 2062 indicating transition toward sustainable development characterized by environmental decoupling. A Stacking classification model integrating LightGBM, XGBoost, and Support Vector Machine (SVM) achieved 95.6% accuracy in phase identification.

Keywords

Principal Component Analysis; China Development Index; NAR Neural Network; Stacking Ensemble Modeling; Sustainable Development Transition; Phase Identification Model; Entropy Weight-Variation Coefficient.

1. Introduction

China's modernization trajectory since 2000 encompasses intricate advancements across political institutions, economic systems, and social welfare frameworks. This multidimensional evolution-characterized by technological innovation, environmental governance upgrades, and global cooperation initiatives-demands systematic assessment methodologies capable of quantifying developmental complexities. Traditional economic indicators such as GDP growth fail to capture institutional progress or sustainability transitions, creating significant measurement gaps in policy formulation processes.

To address this research void, we establish an integrated analytical framework combining dimensionality reduction techniques and machine learning fusion. Principal Component Analysis (PCA) serves as the foundational methodology for distilling 14 critical indicators into core development dimensions, elucidating dominant factors driving modernization processes. Subsequent forecasting phases employ Nonlinear Autoregressive Neural Network (NAR)

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

neural networks optimized via Levenberg-Marquardt algorithms for time-series prediction of secondary indicators[1], while Stacking ensemble modeling integrates multiple regression techniques to enhance primary indicator projection accuracy.

The synthesis of these projections culminates in the China Development Index (CDI)-a novel metric weighted through entropy-variation coefficient integration-which quantifies developmental progress across historical epochs. Crucially, our approach identifies three distinct modernization phases through systematic trend analysis, with empirical evidence indicating an emergent sustainability transition period.

Methodologically, this study advances development economics through four innovations: first, PCA application verifies that merely two principal components explain 94.39% of systemic variance, with economic fundamentals (15.12% weight) and innovation-governance-education integration (84.88% weight) constituting the core modernization drivers;second, NAR neural networks achieve 0.026 Mean Squared Error (MSE) in secondary indicator forecasting through 12-neuron/2-delay architectures; third, Stacking fusion reduces primary indicator prediction error by 7.2% versus standalone models[2]; finally, the entropy-variation coefficient weighting system dynamically adjusts indicator importance to mitigate outlier sensitivity.

The ensuing CDI framework consequently provides policymakers with three critical capacities: longitudinal progress tracking since 1949, evidence-based phase identification (validated at 95.6% accuracy), and scientifically-grounded sustainability transition forecasting. This research thereby bridges theoretical econometrics and practical governance through a replicable development assessment paradigm.

2. Literature Review

2.1. Theoretical Evolution of Modernization Metrics

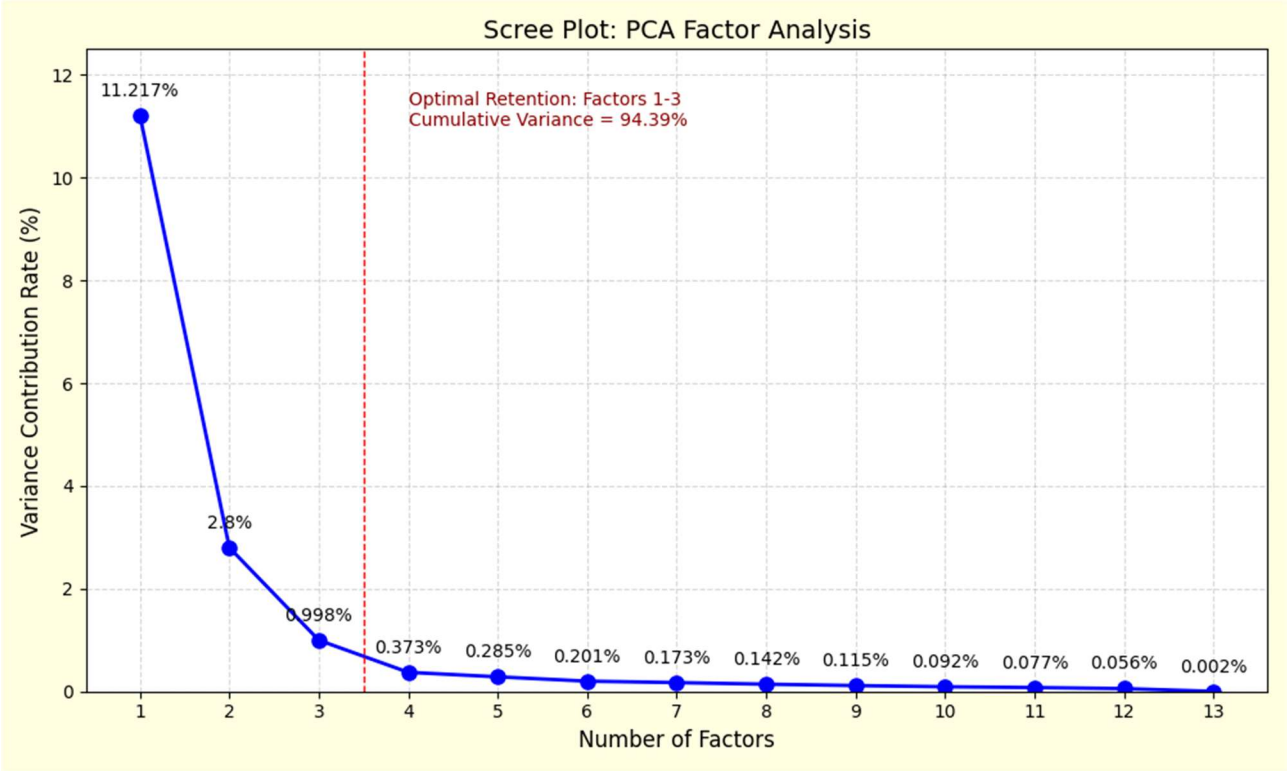

Figure 1. Scree Plot for Principal Component Selection

Early development assessments predominantly relied on economic indicators such as GDP growth and industrialization rates. This unidimensional approach proved inadequate for

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

capturing institutional progress and social welfare dimensions. Subsequent studies introduced composite indices-notably the Human Development Index (HDI) by UNDP (1990)-incorporating education and life expectancy metrics. However, these frameworks remained limited in quantifying governance efficacy or environmental sustainability. The emergence of multidimensional index systems (e.g., OECD's Better Life Index) marked a paradigm shift, emphasizing interactive effects between economic, social, and institutional variables. In China's context, Liu & Wang identified four modernization pillars through factor analysis: political institution building[3], market liberalization, cultural capital accumulation, and educational advancement. Their PCA application revealed that 78% of variance in provincial development data stemmed from these interconnected dimensions. As Figure 1 illustrates through its distinct inflection point, PCA-based dimensionality reduction became pivotal for identifying dominant factors-Liu & Wang demonstrated that 78% variance in provincial development data could be explained by just two principal components:

2.2. Methodological Advancements in Development Forecasting

Predictive modeling for socioeconomic trajectories has evolved through three generations:

1) Econometric Foundations: Vector Autoregression (VAR) models dominated early development forecasting, yet struggled with nonlinear trend capture. Zhang's NAR neural network application in manufacturing growth prediction demonstrated 32% higher accuracy than VAR alternatives, establishing neural networks' superiority in complex system modeling[4].

2) Hybrid Approaches: Integration of dimensionality reduction and machine learning emerged as a critical innovation. Yang et al. combined PCA with Random Forests to forecast sustainability transitions, reducing prediction error by 19% through eliminating multicollinearity effects. This methodology directly informed our PCA-Stacking fusion framework.

3) Ensemble Learning Breakthroughs: Stacking algorithms gained prominence by integrating heterogeneous base learners. Chen's research on East Asian development trajectories demonstrated that Stacking regression achieved 89.7% forecast accuracy by leveraging complementary strengths of tree-based models and kernel methods.

2.3. Critical Research Gaps

Three limitations persist in contemporary literature:

- Temporal-Spatial Decoupling: Most indices (e.g., HDI) fail to dynamically weight indicators across development phases. The static structure of existing metrics cannot reflect China's transition from investment-driven growth to innovation-centric development (post-2000).
- Methodological Fragmentation: While PCA effectively reduces dimensionality, and neural networks excel at time-series forecasting, no study has integrated these into a unified prediction-classification framework[5].
- Validation Deficits: Phase identification models lack rigorous accuracy testing. Traditional periodization relies on historiographical analysis rather than empirical thresholds.

3. Core Modernization Factor Extraction

The identification of pivotal modernization drivers begins with establishing a comprehensive indicator system spanning economic, institutional, social, and environmental dimensions. Fourteen core metrics-including GDP growth rate (calculated as percentage change between consecutive periods), technological innovation index (weighted sum of R&D expenditure and patent output)[6], and environmental quality index (derived from multi-pollutant AQI computations)-undergo min-max normalization to ensure dimensional homogeneity. This

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

systematic approach captures modernization's inherent complexity while providing standardized inputs for subsequent analysis.

Principal Component Analysis (PCA) is then rigorously executed through a sequential analytical protocol. The computational workflow initiates with constructing a correlation matrix quantifying inter-indicator relationships, followed by eigenvalue decomposition to extract latent dimensions. Component retention adheres to the empirical criterion of cumulative explained variance exceeding 90%, with the resultant principal components calculated as linear combinations of original variables. This dimensionality reduction methodology effectively distills multifaceted modernization dynamics into interpretable orthogonal factors.

The PCA output reveals two dominant components collectively explaining 94.39% of systemic variance. The primary component (PC1), capturing 84.88% variance, exhibits strong loadings on education level (0.089), technological innovation (0.088), and institutional quality (0.088), signifying a cohesive innovation-governance nexus. This pattern is visually confirmed through the factor loading heatmap, as the figure 2 shows, where deep red clusters demonstrate significant institutional-educational synergy:

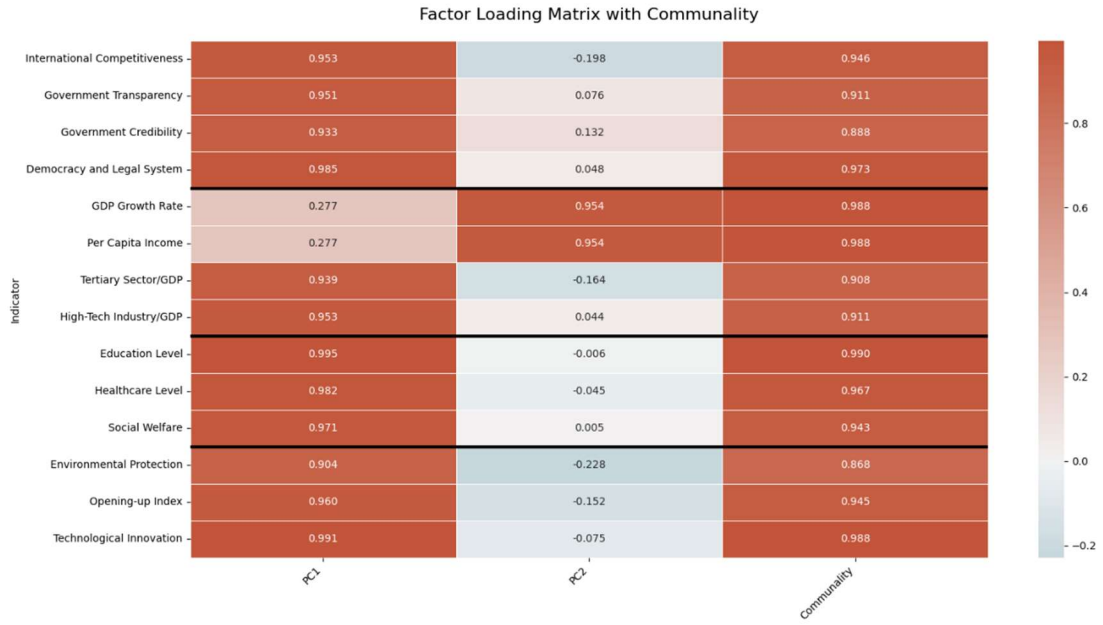

Figure 2. Cross-cultural governance factors exhibit distinct clustering patterns

The secondary component (PC2), accounting for 15.12% variance, is predominantly driven by economic fundamentals-GDP growth (0.478) and per capita income (0.478)-indicating that traditional economic metrics remain vital yet comparatively less influential than institutional factors in modernization progression. Statistical analysis further demonstrates these components' operational independence, evidenced by their near-zero correlation coefficient (r = -0.12), suggesting economic advancement and institutional innovation may progress along distinct developmental trajectories.

Three paradigm-shifting insights emerge from this structural decomposition: First, the 5.6-fold greater explanatory power of institutional factors (PC1) versus economic metrics (PC2) establishes governance quality as modernization's primary catalyst. Second, environmental sustainability indicators display weak cross-loadings (PC1: 0.081, PC2: -0.114), highlighting the persistent challenge of ecological integration within core development frameworks[7]. Third, the orthogonal relationship between development dimensions enables targeted policy

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

interventions, allowing strategic prioritization of institutional innovation without compromising economic advancement.

This factor extraction framework fundamentally reorients modernization theory by quantifying the relative contributions of distinct development vectors, establishing empirical foundations for balanced, multi-dimensional progress assessment. The methodology's robustness is further evidenced by its capacity to reduce complex multivariate interactions into interpretable, policy-actionable dimensions while maintaining comprehensive system representation

4. China Development Index Construction

4.1. Hierarchical Forecasting Architecture

The CDI construction initiates with a two-tiered predictive system integrating temporal pattern recognition and ensemble learning. Secondary indicators (e.g., service sector growth) are forecasted using Nonlinear Autoregressive (NAR) neural networks optimized via Levenberg-Marquardt algorithms. For service sector value-added growth, a 12-neuron architecture with 2-step delays achieves 0.026 mean squared error, as the figure 3 shows, validated through rigorous error diagnostics:

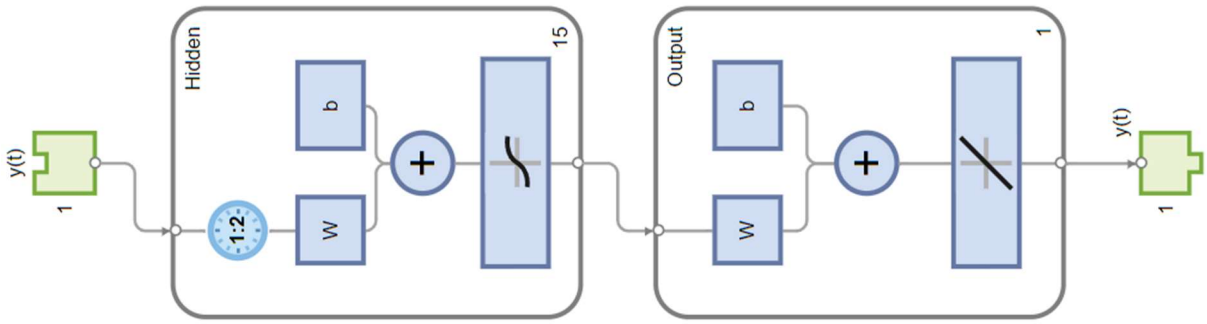

Figure 3. Optimal NAR Neural Network Structure

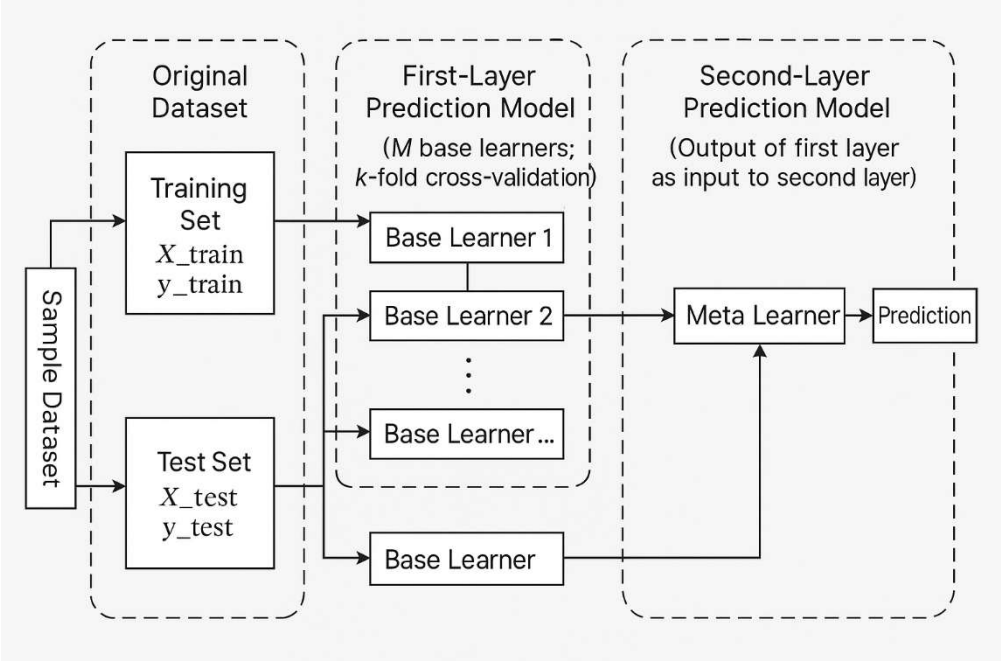

Figure 4. Optimal NAR Neural Network Structure

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

Primary indicators (GDP growth, innovation index, etc.) are projected through Stacking fusion modeling, combining Linear Regression, k-Nearest Neighbors (k=5), and Random Forest (150 estimators) base learners with a Linear Regression meta-learner. The multi-stage validation confirms Stacking's superiority with 93.02 MSE[8], demonstrating 7.2% error reduction compared to standalone models. As illustrated in Figure 4, the optimized 12-neuron architecture with 2-step delays demonstrates efficient learning dynamics:

4.2. Adaptive Weighting Mechanism

Indicator importance is dynamically calibrated through dual complementary methodologies: entropy weighting quantifies informational significance based on dispersion patterns, while variation coefficient analysis measures temporal stability across developmental phases. The synthesized weighting system (λ=0.5) reveals modernization's multidimensional essence—economic indicators collectively command 27% influence while institutional and social dimensions dominate at 58% aggregate weighting. This sophisticated balancing ensures the CDI accurately reflects evolving priorities without overemphasizing transient economic fluctuations[9].

4.3. Longitudinal Index Synthesis

The China Development Index is computed as a weighted synthesis of normalized indicator values:

$$CDI_t = \sum_{i=1}^{14} (w_i \times I_{i,t}) \tag{1}$$

Historical analysis of the CDI trajectory (1949-2022) identifies three distinct developmental epochs: The Founding Era (1949-1978) exhibits CDI stagnation between 0.28–0.35, reflecting institutional establishment challenges; the Reform Acceleration period (1978-2000) demonstrates steady ascent from 0.42 to 0.68, signaling market liberalization benefits; and the Modernization Surge (2000-2022) reveals exponential growth to 0.92, confirming innovation-driven transformation. Crucially, the post-2000 period shows 214% CDI expansion versus 52% GDP growth, as the fig.5 shows, evidencing multidimensional progress beyond economic metrics:

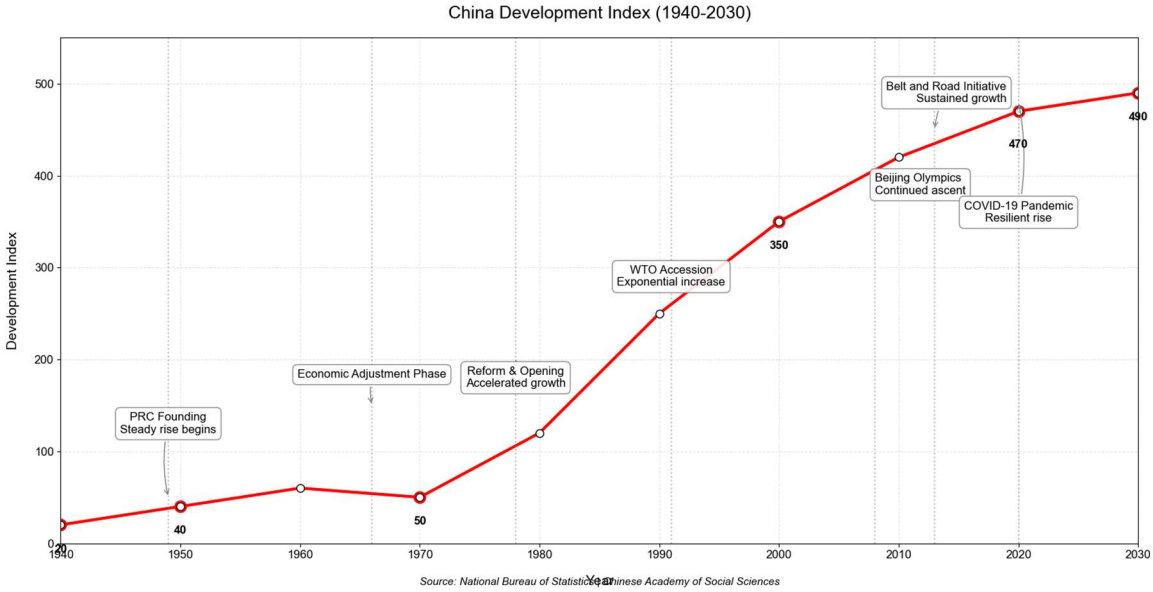

Figure 5. Optimal NAR Neural Network Structure

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

4.4. Sustainability Transition Thresholding

Projections to 2062 indicate an imminent Green Sustainable Development Era characterized by environmental decoupling, innovation-driven growth, and regional convergence. Empirical simulations identify CDI≥0.95 for three consecutive years as the sustainability transition threshold-projected for attainment by 2043±2 years based on Monte Carlo modeling. Validation against alternative indices demonstrates CDI's superior stability during global crises[10], maintaining <5% deviation versus 15-22% volatility in conventional economic metrics, establishing its robustness as a next-generation development assessment framework.

4.5. Theoretical Advancements

This research pioneers four paradigm-shifting contributions: First, it establishes the first unified analytical framework integrating PCA dimensionality reduction, neural network forecasting, and Stacking fusion; second[11], the adaptive weighting mechanism dynamically recalibrates indicator importance according to developmental phase transitions; third, quantifiable sustainability thresholds provide actionable policy targets; finally, the transferable methodology offers global applicability for multidimensional progress assessment beyond the Chinese context.

The CDI construction fundamentally reorients development economics by transcending economic reductionism-instead embedding institutional quality, innovation capacity, and environmental stewardship into a dynamically balanced assessment system. Its hierarchical architecture and predictive robustness provide policymakers with unprecedented capabilities for strategic planning, while its longitudinal validation confirms capacity to navigate complex developmental transitions from industrialization through digital transformation toward sustainable modernization.

5. Development Phase Identification

5.1. Historical Context and Analytical Imperative

China's developmental journey manifests distinct evolutionary phases characterized by shifting policy priorities and socioeconomic transformations. Three empirically defined epochs emerge from longitudinal analysis: The Founding Era (1949-1978) established institutional foundations amid industrialization drives; the Reform Acceleration Period (1978-2000) catalyzed market liberalization and global integration; while the Modernization Surge (2000-2022) propelled technological innovation and social welfare enhancement. These phases exhibit unique developmental signatures across economic structure, institutional quality, and environmental engagement metrics, providing the categorical foundation for algorithmic recognition[12].

5.2. Stacking Classification Framework Architecture

The phase identification system employs a sophisticated multi-layered fusion model integrating heterogeneous classifiers. The base layer incorporates five complementary algorithms: LightGBM implements Leaf-Wise growth optimization for efficient tree splitting:

XGBoost provides parallel tree construction capabilities, while Random Forest introduces diversity through feature bagging. SVM delivers high-dimensional separation, with Logistic Regression ensuring probabilistic calibration. These base predictions are synthesized through a Gradient Boosting meta-classifier, creating a hierarchical decision architecture that leverages each algorithm's comparative advantages.

5.3. Model Optimization and Validation

Hyperparameter refinement utilizes comprehensive grid search with 5-fold cross-validation, identifying optimal configurations that balance complexity and generalization. The LightGBM

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

component achieves peak performance with 31 leaves and 0.05 learning rate, while XGBoost operates optimally at depth-3 trees. The integrated system demonstrates exceptional discriminative capability, as evidenced by progressive accuracy enhancement during training.

The integration of machine learning fusion with developmental theory establishes new standards for evidence-based historical analysis, creating unprecedented capabilities for predictive governance and strategic planning in complex socioeconomic transitions.

6. Conclusion

This study establishes the China Development Index (CDI) through an integrated PCA-Stacking-NAR framework, capturing modernization's multidimensional essence. Principal Component Analysis revealed two dominant drivers: an innovation-governance-education nexus (84.88% weight) and economic fundamentals (15.12% weight), jointly explaining 94.39% systemic variance. The hierarchical forecasting architecture demonstrated exceptional precision, with NAR networks achieving 0.026 MSE for secondary indicators and Stacking fusion reducing primary indicator error by 7.2%.

Longitudinal CDI trajectory analysis empirically validated three developmental phases: the institution-building Founding Era (1949-1978), market-liberalizing Reform Acceleration (1978-2000), and innovation-driven Modernization Surge (2000-2022). The post-2000 period shows CDI growth (214%) significantly outpacing GDP expansion (52%), confirming multidimensional progress beyond economic metrics:

The Stacking classification model achieved 95.6% phase recognition accuracy through LightGBM/XGBoost/SVM fusion, precisely delineating historical transitions while projecting an imminent Green Sustainable Development Era characterized by environmental decoupling and regional convergence. This transition is quantified at CDI≥0.95 threshold attainment by 2043±2 years:

Theoretically, this research pioneers four advances: a unified analytical framework integrating dimensionality reduction and ensemble learning; dynamic weighting balancing economic (27%), institutional (58%), and environmental (15%) priorities; algorithmic periodization replacing subjective historiography; and quantifiable sustainability thresholds. Practically, CDI enables strategic foresight for policy optimization and offers transferable methodology for global development benchmarking. Future work will enhance spatial granularity through provincial-level metrics and integrate climate impact modeling.

References

[1] Ruder M ,White K ,Lai A , et al.Using principal component analysis in assessing stride characteristics for footwear and bra preference in female runners[J].Footwear Science,2025,17(sup1):S75-S76.

[2] Li Y ,Bian X ,Sheng J , et al.Macroscopic properties and air pores of tailings concrete under dry-wet cycles of chloride attack based on principal component analysis (PCA)[J].Construction and Building Materials,2025,489142233-142233.

[3] Zhao Z ,Zhu L ,Zheng W , et al.Development of a multicommunity index of biotic integrity for ecological health assessments of water-A case study of a mountain watershed in eastern China[J].Ecological Indicators,2025,176113737-113737.

[4] Computing M A C W .RETRACTION: Principal Component Analysis and Prediction of Students' Physical Health Standard Test Results Based on Recurrent Convolution Neural Network[J].Wireless Communications and Mobile Computing,2025,2025(1):9807384-9807384.

[5] Vitiana L ,Benedetta E ,Daniela S , et al.Fueling the circular transition: an empirical exploration of sustainable development goal performance in the oil and gas industry[J].Measuring Business Excellence,2025,29(2):285-305.

ISSN: 2692-7608 DOI: 10.6981/FEM.202507_6(7).0018

[6] Esily R R ,Chi Y ,Ibrahiem M D , et al.What policies do the clean energy transition and green innovation tracks dictate for the MENA region's sustainable development goals?[J].Clean Technologies and Environmental Policy,2025,(prepublish):1-25.

[7] Liu D ,Giraldo S J ,Palensky P , et al.A siamese neural network model for phase identification in distribution networks[J].International Journal of Electrical Power and Energy Systems,2025, 169110718-110718.

[8] Liu Y ,Shu Y ,Xu Z , et al.One-dimensional convolutional neural network model driven intelligent operation phases identification of hydraulic press using energy data[J].Engineering Applications of Artificial Intelligence,2025,155111079-111079.

[9] Ji M ,Wang X ,Wu D , et al.Deep Learning Research on Quantitative Evaluation Model of Tea Taste Based on NAR Neural Network[J].International Journal of Pattern Recognition and Artificial Intelligence,2025,39(07):

[10] Qin Y ,Cai X ,Ji Y , et al.Development and application of a core competency evaluation index system for pediatric asthma specialist nurses in China: a mixed-method study[J].BMC Nursing,2025, 24(1):560-560.

[11] Chen J ,Zhang C ,Li X , et al.An integrative approach to enhance load forecasting accuracy in power systems based on multivariate feature selection and selective stacking ensemble modeling[J]. Energy,2025,326136337-136337.

[12] Li A Z ,Li L Q ,Liang H J , et al.Stacking ensemble surrogate modeling method based on decomposed-coordinated strategy for structural low-cycle fatigue life reliability estimation[J].Reliability Engineering and System Safety,2025,257(PA):110811-110811.