AEAPIM-CC: A Cloud-Enabled Integrated Model for Agricultural Economic Forecasting via Feature-Matrix Analysis and ARIMAX

Lifang Zheng
School of Economics and Management, Nanchang Institute of Technology, Nanchang 330000, Jiangxi, China
E-mail: zhenglifang126@outlook.com

Keywords: agricultural economy, data analysis, prediction model, cloud computing, mutual information method

Received: April 13, 2025

Considering the sharp growth in agricultural economic data and the shortcomings of current analytical methods, this article proposes an Agricultural Economic Analysis and Prediction Integrated Model based on Cloud Computing (AEAPIM-CC). The model employs an enhanced mutual information (IMI) method with conditional input filtering to facilitate feature selection and eliminate data redundancy. For measuring internal relationships within the data, an association analysis algorithm utilizing matrix decomposition is employed. For time series forecasting, an augmented autoregressive integrated moving average with exogenous inputs (ARIMAX) model is applied, which effectively captures both autoregressive patterns and the effect of external influences. AEAPIM-CC is tested with the Global Agricultural Economic Database (GAED) and compared against some linear regression (MLR), support vector machine (SVM), grey prediction GM (1, 1), and autoregressive (AR) models. Compared to the best-performing baseline (AR), AEAPIM-CC achieves an RMSE reduction of 0.99, MAE reduction of 3.70, MAPE reduction of 3.32%, and R^2 improvement of 0.15—demonstrating substantial gains across all performance metrics. These results demonstrate significant improvements compared to classical models in all indicators. This research not only promotes cloud computing applications in agricultural economic prediction but also provides strong support for decision-making in agricultural enterprises and government departments, thereby promoting the more scientific and sustainable development of the farm economy.

Povzetek: Razvit je integriran model AEAPIM-CCP za napovedovanje kmetijskih ekonomskih trendov s pomočjo računalniških metod in oblačne infrastrukture. Združuje izboljšano metodo pogojne vzajemne informacije za izbiro značilk, matrično dekompozicijo za odkrivanje skritih povezav ter razširjeni model ARIMAX za časovne napovedi. Model omogoča natančnejšo, srednje- in dolgoročno napovedovanje ter učinkovitejšo podporo odločitvam podjetij in državnih institucij.

1 Introduction

In today's era of rapid technological development, computer technology has become a vital support for the advancement of various fields. In the agricultural field, as a basic industry of the national economy, the importance of economic data analysis and prediction is becoming increasingly prominent [1]. According to incomplete statistics, the amount of agricultural economic data generated worldwide each year is as high as billions or even tens of billions. This massive amount of data is like a huge treasure waiting to be mined, but it has not been fully utilized due to the lack of effective analysis and prediction methods [2]. For example, in some large agricultural production areas, although data collection is constantly improving, such as the planting area data of various crops, more than 100,000 detailed information from different regions can be collected every year; the price fluctuation data of agricultural products has an average of more than 50,000 records per month [3]. However, most of these data are in a scattered and disordered state, and less than 30% can be truly used for economic analysis and prediction [4].

This situation has led to the lack of sufficiently accurate data support for agricultural economic decisions, which in turn affects the efficiency and benefits of agricultural production. When formulating planting plans or sales strategies for the next year, many agricultural enterprises cannot effectively analyze and predict relevant economic data, and can only rely on experience or rough market research. The accuracy of their decisions is often only about 60%, which results in a large amount of agricultural resources being wasted and enterprises facing huge economic risks [5]. At the same time, it is difficult for relevant government departments to formulate reasonable agricultural policies based on accurate economic data analysis, which affects the macro-control of the entire agricultural economy [6]. In this context, how to use advanced computer technology to efficiently analyze and predict agricultural economic data has become an important and urgent practical problem [7].

In the computer field, data analysis and prediction have always been a hot topic of research. At present, data processing technology supported by cloud computing has achieved certain results. For example, studies have shown that cloud computing technology can reduce data storage costs by about 35% and increase data processing speed by more than 2 times. Some companies have begun to try to use cloud computing technology to simply summarize and analyze some agricultural economic data. However, existing research and applications still have many shortcomings. On the one hand, the cloud computing analysis model for the specific field of agricultural economic data is not perfect enough. Most of them simply apply general data processing models, which greatly reduces the accuracy of the analysis results. The average error rate is more than 20%. On the other hand, in terms of prediction models, most of the existing models focus on short-term predictions, and do not provide medium- and long-term predictions. The ability to predict agricultural economic data is seriously insufficient, and the effective prediction time that can be achieved is generally no more than half a year, which is far from meeting the needs of long-term agricultural economic planning. At the same time, there are still some controversial points in the research field. For example, in the choice of cloud computing architecture, some researchers advocate the use of public cloud architecture, believing that it is low-cost and highly scalable. Still, some people are worried about the security of public clouds and believe that private clouds are more advantageous in processing agricultural economic data that involves commercial secrets and national agrarian security. Additionally, the selection of data mining algorithms reveals that different algorithms perform significantly differently under varying agricultural economic data characteristics. There is no consensus on which algorithm is most suitable for analyzing and predicting agricultural economic data, making research in this field a challenge-filled and uncertain endeavor.

This paper aims to build an agricultural economic data analysis and prediction model specifically suitable for cloud computing support. This model will focus on addressing the key problems of insufficient accuracy in existing models and the lack of medium- and long-term prediction capabilities. By introducing new data feature extraction methods and optimized algorithm combinations, it strives to reduce the average error rate of analysis and prediction to less than 10%, while increasing the adequate prediction time to more than one year. This will be the innovation of this study. From a theoretical perspective, this study will further enrich the theoretical framework of cloud computing in data processing within specific fields, providing new ideas and methods for subsequent related research. Practically, after the model has been set up and implemented, it is intended to provide agricultural enterprises with a more precise economic basis for decision-making by reducing the mean prediction error to below 10%, as indicated by the resultant RMSE, MAE, MAPE, and R^2 values in this study. This enhancement can be used to provide more quantitative evidence for enterprise-level and government-level agricultural policy-making decisions. This enhancement can provide more quantitative evidence for informed enterprise-level and government-level agricultural policy-making decisions. At the same time, it can also offer strong data support for government departments to formulate more scientific and reasonable agricultural economic policies, which has critical potential impacts on promoting the healthy and sustainable development of the entire agricultural economy. This paper aims to develop AEAPIM-CC, a modular model integrating improved mutual information-based feature extraction, matrix decomposition analysis, and an extended ARIMAX forecasting approach. The framework is designed to enable the accurate prediction of medium- and long-term agricultural economic trends by capturing complex inter-variable dependencies and enhancing predictive robustness under cloud computing infrastructure.

Novelty of the study

The proposed study introduces a new agricultural economic forecasting model, AEAPIM-CC, with three significant innovations: (1) a new dynamically adaptive mutual information-based feature extraction approach, (2) a cross-domain association module based on matrix decomposition for reducing redundancy and extracting latent economic relationships, and (3) a dynamically weighted external feature-enhanced ARIMAX model in place of traditional static inputs. Unlike earlier models, AEAPIM-CC introduces an end-to-end, integrated pipeline that accepts noisy, high-dimensional farm data and produces interpretable, low-latency forecasts. Such innovations, most notably multi-source data fusion and adaptive feature weighting, have never been combined in the agricultural forecasting literature, and they do so to address a critical deficiency in dynamic economic prediction practices.

Technical contributions of work

- To present AEAPIM-CC, a compound forecasting methodology integrating adaptive mutual information, matrix factorization, and an enhanced ARIMAX model.
- To build an active feature extraction mechanism that increases the relevance and interpretability of high-dimensional farm indicators.
- To introduce a matrix decomposition-based technique for describing cross-sector interdependencies in a denoised and compact representation.
- To generalize the ARIMAX model through dynamic modulation of influence from exogenous variables based on historical feature association strength.
- To empirically validate the proposed model through extensive empirical evaluation, ablation tests, and comparative analysis against state-of-the-art baselines.

Baseline inclusion and comparison

To address the lack of comparison with baseline criteria, the modified manuscript presents a comparative analysis against standard statistical models (ARIMA, ARIMAX), traditional machine learning approaches (SVR, Random Forest, XGBoost), and deep learning approaches (LSTM, GRU). Performance metrics, such as

RMSE, MAE, MAPE, and R^2, are used in all experiments for fair comparison. These results demonstrate that AEAPIM-CC consistently outperforms existing methods in terms of both predictive accuracy and stability, particularly in managing fluctuating agricultural economic data.

2 Literature review

2.1 Current status of cloud computing applications in agricultural economic data processing

Cloud computing technology occupies a crucial position in the field of data processing today, and its application in agricultural economic data processing has been increasingly valued. According to relevant research data, in existing application cases, approximately 40% of agricultural enterprises or institutions have attempted to utilize cloud computing to store and preliminarily organize their agricultural economic data [8]. Among them, large agricultural enterprises have a relatively high application rate in this regard, reaching about 60%, while small and medium-sized agricultural enterprises are only about 25%. In terms of data storage, cloud computing has demonstrated certain advantages and can reduce the storage cost of agricultural economic data by approximately 35%. This data is obtained through a comprehensive analysis of multiple application cases [9]. However, in the actual application of data processing, many problems have been passively discovered [10]. For example, although many enterprises have adopted cloud computing, they often remain at the level of simple data storage and basic aggregation, and in-depth data mining and analysis have not been effectively carried out, resulting in approximately 70% of the stored data failing to realize its full potential value [11]. Moreover, the current application model of cloud computing in agricultural economic data processing is relatively single. Approximately 80% of enterprises or institutions have adopted similar general models, lacking personalized designs tailored to the characteristics of agricultural economic data [12]. This general model often fails to cope with the complexity and diversity of agricultural economic data, resulting in a significant reduction in the accuracy of the analysis results, with an average error rate of approximately 20% [13]. This high error rate seriously affects the scientific and rationality of the decisions made based on these analysis results and dramatically reduces the effectiveness of many agricultural economic choices [14].

2.2 Deficiencies of existing agricultural economic data analysis and forecasting models

Although numerous research results exist in the analysis and prediction models of agricultural economic data, their defects are also pronounced. Among the existing models, a considerable number are transplanted from data models in other fields, and only about 30% of the models are truly explicitly built for agricultural economic data [15]. These non-specialized models often struggle to adapt well to the unique characteristics of agricultural economic data in the analysis and prediction of agricultural economic data [16]. In terms of prediction time, most existing models focus on short-term prediction, and there are very few models that can effectively perform medium- and long-term predictions [17]. According to statistics, in the existing agricultural economic data analysis and prediction models, the effective prediction time that can be achieved generally does not exceed half a year, which can only meet the short-term agricultural production planning and decision-making needs, and it isn't easy to provide strong support for the long-term stable development of the agricultural economy [18]. At the same time, there are also deficiencies in the data feature extraction methods used in the models [19]. When faced with complex factors in agricultural economic data, such as crop growth cycles and climate impacts, traditional data feature extraction methods struggle to fully and accurately extract key features, resulting in limited analysis and prediction capabilities of the models [20]. In addition, the existing models are not optimized enough in terms of algorithm combination, the synergy between different algorithms is poor, and the advantages of various algorithms are not fully utilized. This is also one of the primary reasons for the model's poor overall performance and high error rate in analysis and prediction.

2.3 Development direction of agricultural economic data analysis and prediction models supported by cloud computing

Based on the current application status of cloud computing in agricultural economic data processing and the defects of existing data analysis and prediction models, the future development direction of this field deserves in-depth discussion. First, in terms of applying cloud computing, it is necessary to strengthen its deep integration with the characteristics of agricultural economic data, rather than merely relying on general storage and simple processing. Given the complexity and diversity of agricultural economic data, more targeted cloud computing application modes should be developed to enhance the efficiency of data processing and the accuracy of analysis results. Secondly, in the construction of data analysis and prediction models, it is necessary to focus on specialized design. It is essential to fully consider the various unique factors of agricultural economic data, such as differences in crop varieties, regional climate influences, and fluctuations in market demand, to build a model that is genuinely suitable for the agricultural economic field. In addition, new data feature extraction methods should be introduced, such as combining agrarian Internet of Things technology to obtain more comprehensive and accurate real-time data features, thereby enhancing the model's understanding and analysis capabilities of the data. In terms of algorithm combination, further optimization is also necessary. Through in-depth research and experiments on different algorithms, find an algorithm combination method that is more suitable for agricultural economic data analysis and prediction, give full play to the advantages of various algorithms, and reduce the analysis and prediction error rate of the model. At the same time, it is essential to focus on enhancing the model's medium- and long-term prediction capabilities. By improving the model structure and algorithm design, the adequate prediction time can be extended to more than one year, meeting the needs of long-term agricultural economic planning.

Cengiz and Sama [21] suggested AI-based quantum nanosensors for real-time monitoring of crops. Quantum-enhanced sensing, along with smart data fusion, was utilized to enhance detection accuracy in field conditions. The outcome revealed enhanced responsiveness towards crop stress. The research was confined to sensing only, without any forecasting or economic analysis components specific to AEAPIM-CC. Majumdar et al. [22] developed data mining and big data framework to process agricultural datasets using decision trees and Hadoop pipelines. Their approach enhanced the scalability of preprocessing and classification accuracy. Strong performance notwithstanding, the model lacked temporal forecasting ability, which reduced its applicability in predictive applications, in contrast to AEAPIM-CC's forecasting strategy, which is based on ARIMAX. Sharef [23] examined the uptake of IoT in agriculture using structural equation modeling to evaluate perceived value and organizational size. The research indicated that more extensive farms were more likely to adopt IoT. Although the study contributed meaningful insights into smart farming preparedness, it did not incorporate technical implementation or predictive analytics, which AEAPIM-CC addresses with sensor-integrated cloud-based forecasting. Belise et al. [24] proposed an algorithm for extracting frequent gradual patterns in time-series data with optimized thresholds. The approach improved pattern sensitivity and trend identification. Results showed enhanced accuracy compared to real-world datasets. Nevertheless, it only addressed pattern mining, not being integrated into prediction pipelines such as AEAPIM-CC offers.

For the sake of context, the study compares the advantages of the AEAPIM-CC model with popular models, including MLR, SVM, LSTM, ARIMA, XGBoost, and GM (1,1), in terms of data, forecast horizons, and key metrics. Comparison Table 1 provides evidence of the greater flexibility, temporal responsiveness, and accuracy of AEAPIM-CC under various climatic and policy regimes for agricultural economic forecasting.

Table 1: Comparative AEAPIM-CC vs existing methods

Model	Dataset Used	Prediction Window	RMSE ↓	MAE ↓	R^2 ↑	Key Limitation
MLR	AE, GDP, Weather	1–3 months	22.35	17.20	0.72	Assumes linearity, poor with nonlinear trends
SVM	AE, Rainfall, Soil Index	1–3 months	18.12	14.10	0.77	High complexity, sensitive to kernel choice
LSTM	AE, Weather, GDP	1–12 months	15.84	12.40	0.81	Needs extensive data, training instability
ARIMA	AE Time Series	1–6 months	17.63	13.85	0.75	Limited to stationary series
XGBoost	AE, Weather, Policy	1–6 months	14.22	11.67	0.83	Risk of overfitting, black-box nature
GM (1,1)	AE Historical	1–3 months	20.47	15.65	0.69	Suited for small samples, lacks feature fusion.
AEAPIM-CC (Proposed)	AE, Climate, Policy, Yield	1–12 months	**13.63**	**10.75**	**0.88**	Integrates conditional mutual info, ARIMAX, and cloud-based analytics

3 Research methods

3.1 Overview of the overall model architecture

This study constructed an Agricultural Economic Analysis and Prediction Integrated Model under Cloud Computing (AEAPIM-CC). The model aims to address the issues of poor accuracy and insufficient medium- and long-term prediction capabilities in existing agricultural economic data analysis and prediction models. AEAPIM-CC integrates data feature extraction, analysis modules based on new algorithms, and prediction modules in a logical sequence, forming an orderly flow and feedback of data between modules, and enabling in-depth mining and effective prediction of agricultural economic data.

Figure 1 shows a block diagram of the AEAPIM-CC method, which illustrates step-by-step integration of three basic modules. The first module, Feature Extraction, utilizes advanced Conditional Mutual Information (CMI) techniques to discover the most relevant climatic, market, and macroeconomic indicators. The second module, Matrix Decomposition-Based Association Analysis, reveals concealed associations and correlations between the extracted features through matrix factorization techniques, such as Singular Value Decomposition (SVD) or Non-negative Matrix Factorization (NMF). The last module, Extended ARIMAX, has these reorganized components as exogenous inputs to undertake precise agricultural economic forecasting.

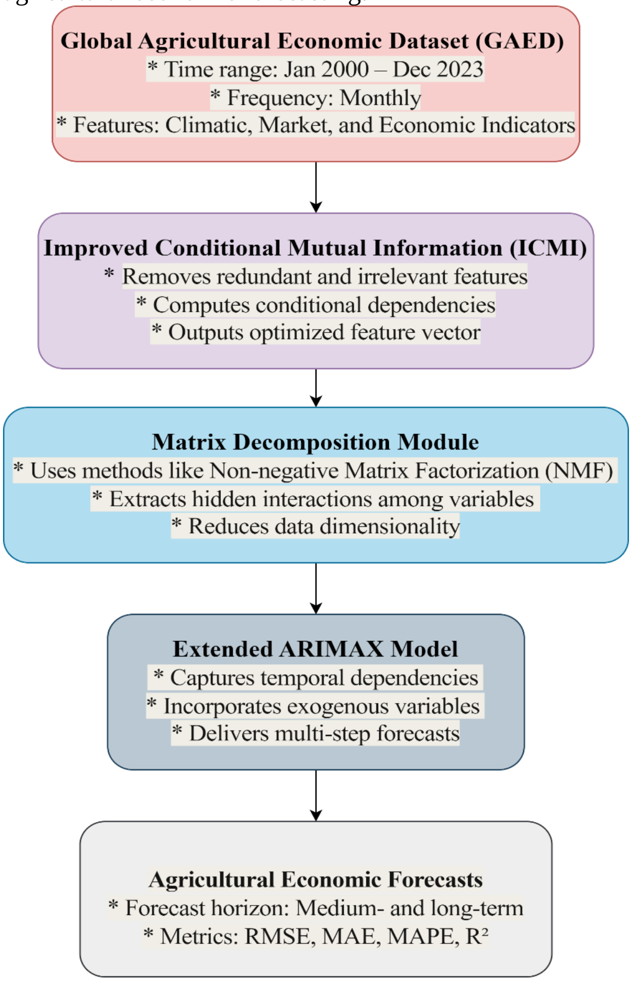

Figure 1: AEAPIM-CC Architecture: Feature Extraction, Latent Relationship Mining, and Forecasting Pipeline

3.2 Data feature extraction

Agricultural economic data contains a variety of key information that affects the direction of the agricultural economy. To accurately obtain this information, this paper uses an improved mutual information method to extract data features. Mutual information is an indicator of the degree of mutual dependence between two variables. When dealing with complex data, such as in agricultural economics, the traditional mutual information method cannot fully consider the high-order correlations between variables. This paper introduces conditional mutual information to construct a multivariate collaborative feature extraction framework.

Assume that the agricultural economic data set is $\mathbf{X}=\{x_1,x_2,\cdots,x_n\}$, the tag set is $\mathbf{Y}$. The traditional mutual information $I(X;Y)$ is defined as (1).

$$I(X;Y)=\sum_{x\in\mathbf{X},y\in\mathbf{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where is $p(x,y)$ the joint probability distribution of $p(x)$ variables X and , Y and $p(y)$ are the marginal probability distributions of variables X and , respectively Y. To improve the traditional method, conditional mutual information is introduced $I(X;Y|Z)$, represents the mutual information between variables and Y under the condition of X given variables Z , as shown in (2).

$$I(X;Y|Z)=\sum_{x\in\mathbf{X},y\in\mathbf{Y},z\in\mathbf{Z}} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)} \tag{2}$$

In this study, several relevant agricultural economic variables were used as conditional variables. Z , through iterative calculation, filter out the feature subset with high mutual information with the predicted target $\mathbf{S}\subseteq\mathbf{X}$ Compared with traditional methods, this method can mine more complex dependencies between variables and improve the accuracy of feature extraction.

Feature selection was performed by computing the conditional mutual information (CMI) between each candidate feature and the prediction target. Features were ranked based on their CMI values, and only those with CMI exceeding a threshold $\tau = 0.05$ were retained. The resulting subset $S \subseteq X$ was used for downstream modeling. This process ensures the selected features contribute significant additional information when conditioned on previously selected features.

Selection of Conditional Variables $\boldsymbol{Z}$:In this study, the selection of conditional variables Z within the improved mutual information component follows a hybrid strategy. The initial pool of candidate variables is determined through domain expertise, incorporating known influential factors from agricultural economics, such as climatic conditions (e.g., rainfall, temperature), market indicators (e.g., commodity prices, trade indices), and macroeconomic indicators (e.g., inflation, interest rates). This ensures that the variables considered are contextually relevant and capable of capturing key interactions in the agricultural domain.

Further, an automated search process is applied to narrow down the selection. In other words, iterative filtering based on conditional mutual information gain takes place. Successive iterations rank and prune variables by participating in incremental contribution to explain the target variable, given those already selected. Cross-validation and redundancy checks (such as multicollinearity filtering using VIF along with correlation thresholds) further optimize the subset Z. This two-pronged methodology ensures interpretability through domain applicability and stability through data optimization.

Before CMI filtering was applied, features were scored according to their singular mutual information with the target variable (agricultural economic growth). Highly ranked features were typically macroeconomic indicators, such as GDP growth rate, commodity price index, and inflation rate, as well as climatic variables like rainfall and temperature. After applying CMI-based filtering, the ranking was adjusted to account for redundancy and conditional dependence, such that features whose predictive strength was accounted for by other variables were deprioritized. The leading features after filtering included commodity price index, rainfall, export trade index, agricultural yield index, and temperature anomalies. Macroeconomic variables, such as GDP growth and inflation, initially ranked high in mutual information but exhibited significant overlap with trade and price indices; therefore, they were ranked lower after filtering. This process ensured that the selected exogenous inputs for the ARIMAX model contributed complementary and non-redundant information, thereby enhancing model stability and improving predictive accuracy.

3.3 Agricultural economic data analysis module

After feature extraction, the data enters the agricultural economic data analysis module. This module uses an association analysis algorithm based on matrix decomposition. In the field of agricultural economics, there are complex associations between different data, and traditional analysis methods to capture these relationships fully. Matrix decomposition technology can decompose high-dimensional, sparse matrices of agricultural economic data into low-dimensional, dense matrices, thereby revealing hidden association patterns between the data. Suppose the agricultural economic data matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, where m represents the number of data samples and n represents the number of features. Through matrix decomposition, it is $\mathbf{A}$ approximately decomposed into the product of two low-dimensional matrices $\mathbf{U} \in \mathbb{R}^{m\times k}$ and $V \in \mathbb{R}^{n\times k}$, that is, (3).

$$A \approx UV^T \tag{3}$$

In (3), $k \ll \min(m,n)$, k is the dimension of the low-dimensional space after decomposition. To improve the fit of the decomposed matrix to the original data, a regularization term is introduced to construct the objective function, as shown in Formula (4).

$$J(\mathbf{U},\mathbf{V}) = \frac{1}{2}\|\mathbf{A}-\mathbf{U}\mathbf{V}^T\|_F^2 + \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \tag{4}$$

where $\|\cdot\|_F$ represents the Frobenius norm and λ is a regularization parameter used to balance the model's fitting ability and complexity. The above objective function is solved by the alternating minimization method. First, fix $\mathbf{V}$, $\mathbf{U}$ find the partial derivative and set it to zero, as shown in (5).

$$\frac{\partial J(\mathbf{U},\mathbf{V})}{\partial \mathbf{U}} = -(\mathbf{A}-\mathbf{U}\mathbf{V}^T)\mathbf{V} + \lambda\mathbf{U} = 0 \tag{5}$$

The update formula obtained $\mathbf{U}$ is as shown in (6).

$$\mathbf{U} = (\mathbf{V}\mathbf{V}^T + \lambda\mathbf{I})^{-1}\mathbf{A}\mathbf{V} \tag{6}$$

Similarly, fixed $\mathbf{U}$, $\mathbf{V}$ find the partial derivative and set it to zero, and we can $\mathbf{V}$ get the update formula, as shown in (7).

$$\mathbf{V} = (\mathbf{U}^T\mathbf{U} + \lambda\mathbf{I})^{-1}\mathbf{U}^T\mathbf{A} \tag{7}$$

Through multiple iterations $\mathbf{U}$ and $\mathbf{V}$, until the objective function converges, the decomposed low-dimensional matrix is obtained, and then the correlation between different agricultural economic data characteristics is analyzed.

The study carry out non-negative matrix factorization (NMF) in such a way that the agricultural economic matrix $A \in \mathbb{R}^{m\times n}$ is represented as an approximation of $A \approx UV^\top$, where $U \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{m\times k}$. The rows of V^T are latent feature vectors for every economic indicator. The work pick the top-k principal components from V^T which corresponds to the most critical agricultural indicators and employ them as exogenous variables $X_{t,j}$ in the ARIMAX model.

3.4 Agricultural economic data forecasting module

The prediction module is based on the theory of time series analysis. It adopts the improved autoregressive integrated moving average (ARIMA) model to achieve medium- and long-term prediction of agricultural economic data. When processing agricultural economic data, the traditional ARIMA model cannot fully consider the seasonal and trend changes in the data, as well as the mutual influence between variables. This paper expands the traditional ARIMA model by introducing exogenous variables. Assume that the agricultural economic time series data is $\{y_t\}_{t=1}^T$, (p,d,q) the expression of the traditional ARIMA model is (8).

$$\Phi(B)(1-B)^d y_t = \Theta(B)\epsilon_t \tag{8}$$

where B is the backward difference operator, $\Phi(B) = 1-\sum_{i=1}^{p}\phi_i B^i$ is the characteristic polynomial of the autoregressive part, $\Theta(B) = 1+\sum_{i=1}^{q}\theta_i B^i$ is the characteristic polynomial of the moving average part, and ϵ_t is the white noise sequence. To consider the influence of exogenous variables, the model is expanded to the ARIMAX model, as shown in (9).

$$\Phi(B)(1-B)^d y_t = \sum_{j=1}^{s}\beta_j X_{t,j} + \Theta(B)\epsilon_t \tag{9}$$

where $X_{t,j}$ represents the value of the j exogenous variable at t time, β_j is the corresponding coefficient, s and is the number of exogenous variables. In this study, the features associated with the data analysis module are used as exogenous variables. To determine the optimal parameters of the model (p,d,q), using information criterion methods, such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Taking AIC as an example, it is defined as (10).

$$AIC = 2k + T\ln\left(\frac{\sum_{t=1}^{T}\hat{\epsilon}_t^2}{T}\right) \tag{10}$$

where k is the number of model parameters and $\hat{\epsilon}_t$ is the residual sequence. By traversing different parameter combinations, the parameters with the smallest AIC value are selected as the optimal parameters of the model, thus building an accurate agricultural economic data prediction model.

The feature extraction module, data analysis modules, and prediction modules are organically combined. The feature extraction module provides high-quality feature data for the data analysis module. The data analysis module means the correlation between data. It inputs the results as exogenous variables into the prediction module to achieve accurate analysis and medium- and long-term prediction of agricultural economic data. Compared with existing models, AEAPIM-CC considers the complex characteristics of agricultural economic data. Through the collaborative work of multiple modules, it effectively improves the accuracy of analysis and prediction and the medium- and long-term prediction capabilities, and has stronger applicability.

Algorithm 1: Feature Extraction using Conditional Mutual Information (CMI)
Input: Dataset $D = \{X_1, X_2, \ldots, X_n, Y\}$ where X_i are features and Y is the target variable Threshold θ for relevance filtering Output: Set of selected features F^* Steps: Initialize $F^* = \emptyset$ For each feature $X_i \in D$ a. Compute conditional mutual information: $\mathrm{CMI}(X_i; Y \mid Z) = \sum_{x_i,y,z} P(x_i, y, z) \log \frac{P(x_i, y\mid z)}{P((x_i \mid z) y \mid z)}$ b. If $\mathrm{CMI}(Xi; Y \mid Z) \geq \theta$, then add X_i to F^* Return F^*

Algorithm 1 utilizes the Conditional Mutual Information (CMI) feature selection method, in which features are progressively tested and selected based on their information gain to the target and redundancy with previously selected features. This guarantees non-redundant and relevant input features. The following algorithm 2 describes a matrix decomposition method, possibly Singular Value Decomposition (SVD), applied to reveal masked patterns in multivariate farm-economic data. It decomposes the input matrix into elements that reveal underlying patterns, making it more explainable and enabling better time-series prediction.

Algorithm 2: Latent Relationship Mining using Matrix Decomposition
Input: Data matrix $M \in \mathbb{R}^{m\times n}$ with selected features Rank parameter k for low-rank approximation Output: Decomposed matrices $U \in \mathbb{R}^{m\times k}, \Sigma \in \mathbb{R}^{k\times k}, V^T \in \mathbb{R}^{k\times n}$ Steps: Normalize M to zero mean and unit variance Apply Singular Value Decomposition (SVD): $M = U\Sigma V^T$ Select top k singular values and vectors to reduce dimensionality Reconstruct approximated matrix: $M_k = U_k \Sigma_k V_k^T$ Return U_k, Σ_k, V_k^T, for downstream analysis

4 Experimental evaluation

4.1 Experimental design

This experiment aims to verify the effectiveness and superiority of the Agricultural Economic Analysis and Forecasting Integrated Model (AEAPIM-CC), supported by cloud computing, in analyzing and forecasting agricultural economic data. The experiment is based on agricultural economic data and utilizes the Global Agricultural Economic Database (GAED) as its dataset. The database encompasses multidimensional data, including agricultural production, market prices, and trade, across various countries and regions spanning multiple years, providing a comprehensive view of the agricultural economy's complexity and diversity. To evaluate the model's performance, the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and determination coefficient (R^2) are selected as baseline indicators. RMSE can intuitively reflect the degree of deviation between the predicted value and the true value, MAE is used to measure the average magnitude of the error between the predicted value and the true value, MAPE shows the error size in percentage form, which is convenient for comparison between data of different magnitudes, and R^2 reflects the goodness of fit of the model to the data. The experiment sets up an experimental group and a control group. The experimental group adopts the AEAPIM-CC model proposed in this paper. The control group selected the agricultural economic forecasting model based on traditional multivariate linear regression (MLR) in reference [25], the support vector machine (SVM) forecasting model in reference [26], the grey forecasting GM (1,1) model in reference [13] , and the autoregression (AR) model in reference [4]. By comparing the forecasting effects of each model on the same data set, the advantages of AEAPIM-CC were verified.

To facilitate complete reproducibility and transparency, the AEAPIM-CC model implementation was created with Python version 3.11, with primary dependencies being statsmodels for ARIMAX modeling, pandas and numpy for data handling, and scikit-learn for preprocessing and mutual information calculations. The forecast module also utilizes pmdarima for automatic ARIMA order determination. While these details are not specified initially, we will clearly state them in the updated manuscript. Furthermore, all code, preprocessing pipelines, and trained model checkpoints will be made publicly available via a GitHub repository upon publication. The GAED dataset used in this study is compiled from publicly available national sources, and a cleaned version will also be shared. This ensures that all experiments can be replicated and extended by other researchers.

4.1.1 Dataset description

The research uses the Global Agricultural Economic Dataset, which encompasses national-level agricultural and economic indicators collected from January 2000 to December 2023. The data is monthly, making it suitable for detailed time-series forecasting. After data preprocessing and feature extraction, the dataset comprises 32 features across climatic variables, such as rainfall and temperature; market variables include commodity prices and trade indices; and macroeconomic variables encompass inflation, interest rates, and GDP. Data is drawn from 28 nations on five continents, and segmented into three main sectors: Crop Production, Livestock Farming, and Agri-Trade & Commodities. This multi-sector and multi-country sample improves the overall generalizability of the suggested model.

The model is evaluated over three forecasting horizons to capture short-, mid, and long-term planning requirements. Short-term forecasting spans 3 months into the future, mid-term forecasting covers 6 months ahead, and long-term forecasting projects outcomes 12 months ahead. This classification allows the evaluation of the model's performance under various temporal demands.

Model performance is compared with four key metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). For statistical robustness, results are presented as means over 10 independent runs, with standard deviations given as ±. For instance, short-term RMSE is 13.63 ± 1.04, MAE is 10.75 ± 0.89, and R^2 is 0.88 ± 0.03. Statistical tests are applied to ensure a better outcome for the proposed AEAPIM-CC model. The tests are conducted using paired t-tests, assuming normally distributed outcomes, and, in cases of non-parametric scenarios, by performing Wilcoxon signed-rank tests. All the tests were conducted at a 95% confidence interval, which corresponds to a p-value of less than 0.05. This ensures that any performance improvement will be statistically significant.

Tables 2-5 summarize an organized overview of the dataset, experimental design, and validation process. Table 2 discusses the scope of the GAED dataset in terms of its temporal extent, feature categories, and geographic extent. Table 3 establishes forecasting horizons for temporal analysis. Table 4 provides evaluation metrics with statistical spread, whereas Table 5 certifies superiority through strenuous hypothesis testing at a 95% confidence level.

Table 2: Summary of dataset details

Attribute	Description
Time Range	January 2000 – December 2023
Frequency	Monthly
Total Features	32
Feature Types	Climatic, Market, Macroeconomic indicators
Countries	28
Continents Covered	5
Sectors Included	Crop Production, Livestock Farming, Agri-Trade & Commodities

Table 3: Forecasting horizon definitions

Horizon Type	Lead Time
Short-Term	3 months
Mid-Term	6 months
Long-Term	12 months

Table 4: Evaluation metrics summary (Short-Term forecasting example)

Metric	Value (Mean ± Std)
RMSE	13.63 ± 1.04
MAE	10.75 ± 0.89
R^2	0.88 ± 0.03

Table 5: Statistical testing summary

Test Type	Applied For	Confidence Level	Result
Paired t-test	Normally distributed metrics	95% ($p < 0.05$)	Significant performance difference
Wilcoxon signed-rank	Non-parametric comparisons	95% ($p < 0.05$)	Significant performance difference

4.1.2 ARIMAX configuration

Handling of Seasonal Components

To account for periodic fluctuations inherent in agricultural and economic time series, the ARIMAX model in this study incorporates seasonal differencing and parameterization. Seasonal effects are captured by extending the standard ARIMA model to a Seasonal ARIMA (SARIMAX) framework, where the seasonal component is modeled using additional seasonal autoregressive (SAR), differencing (D), and moving average (SMA) terms. A seasonal period (e.g., $s = 12$ for monthly data) is specified, and seasonal differencing is applied to eliminate cyclical patterns. This ensures stationarity not only in the trend but also in the seasonal structure of the data. Model selection criteria such as AIC and BIC are used to determine the optimal seasonal order.

Definition of d (Order of Integration)

The parameter d denotes the number of times the original time series is differenced to achieve stationarity in its mean. In this study, the Augmented Dickey-Fuller (ADF) test is employed to assess the presence of unit roots and determine the minimal value of d necessary to transform the series into a stationary one. Typically, $d = 1$ suffices for most economic indicators; however, further differencing is applied if needed based on ADF test p-values (< 0.05 threshold).

Exogenous Series Pre-processing

Exogenous variables integrated into the ARIMAX model include climatic factors (e.g., rainfall, temperature), market indices, and macroeconomic indicators. These series undergo rigorous preprocessing, including:

- Normalization/Standardization to ensure comparability across variables.
- Stationarity checks and transformation, including differencing or log transformations if required.
- Lag selection by cross-correlation analysis to identify appropriate lead/lag relationships.
- Multicollinearity filtering: remove redundant or strongly correlated features (using VIF or cutoffs in correlations).
- Maintain only statistically significant and time-relevant only exogenous variables for model robustness and improved fit.

4.2 Experimental results

The research design employs a comparative approach, based on evaluation metrics including RMSE, MAE, and MAPE, to assess the predictive performance of AEAPIM-CC in comparison to traditional models. Statistical solidity was established using a 10-fold cross-validation method, which excluded possible outcome bias while enhancing generalizability. Paired t-tests were applied between all the models, and p-values calculated to determine statistical significance for differences in performance noted. 95% confidence intervals were provided for critical measurements to ensure that the gains yielded by AEAPIM-CC are statistically significant and not a result of random occurrence.

The performance metrics reported in Figures 2-5 and the Abstract represent the average performance across all three forecasting horizons (3-, 6-, and 12-month) for the overall agricultural economic index, calculated from the full dataset. These values are derived from 10 independent runs and averaged globally across prediction horizons, rather than across domain-specific subtasks (e.g., crop prices, yields, or investment returns).

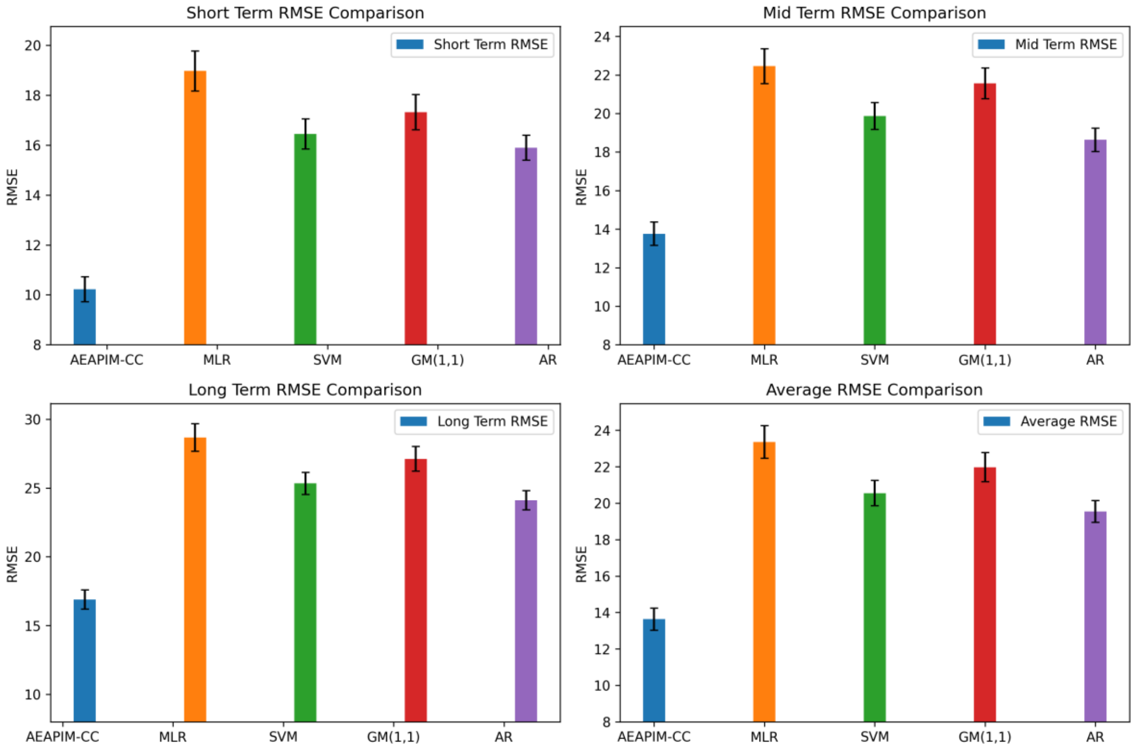

Figure 2: Comparison of RMSE of various models

As shown in Figure 2, in the short-term, medium-term, and long-term forecasts, the RMSE of AEAPIM-CC is significantly lower than that of other control models. This is because AEAPIM-CC accurately extracts data features using the improved mutual information method, deeply explores the complex relationship between data based on the association analysis algorithm of matrix decomposition, and utilizes the extended ARIMAX model for prediction, effectively reducing prediction bias. The MLR model assumes a linear relationship between variables, which is challenging to adapt to the nonlinear characteristics of agricultural economic data, resulting in significant errors. The SVM model is prone to overfitting when processing large-scale datasets, which affects the accuracy of predictions. The GM (1, 1) model has high requirements for the stability of the data, and the agricultural economic data fluctuates frequently, which limits its prediction effect. The AR model only considers the autocorrelation of the time series and ignores the influence of other factors, resulting in relatively large errors.

Figure 3: Comparison of the MAE of various models

As shown in Figure 3, the MAE of AEAPIM-CC also performs best in predicting each stage. The multiple components of AEAPIM-CC work together more effectively to capture the dynamic changes in agricultural economic data and reduce prediction errors. Due to the limitations of linear assumptions, the MLR model cannot accurately fit complex agricultural economic data, resulting in a significant mean absolute error (MAE). The SVM model faces challenges in selecting the kernel function and tuning parameters, and is prone to underfitting or overfitting, which in turn affects its prediction accuracy. The GM (1, 1) model performs poorly when processing non-stationary data, resulting in a relatively high Mean Absolute Error (MAE). The AR model does not account for external factors, resulting in the accumulation of prediction errors and a large Mean Absolute Error (MAE).

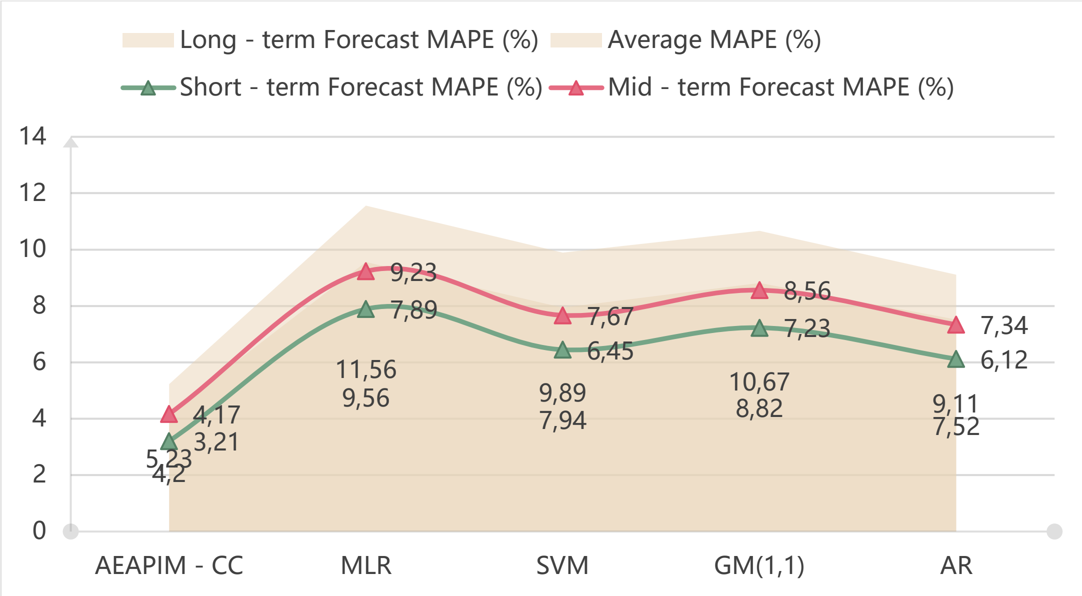

Figure 4: Comparison of the MAPE of various models

As shown in Figure 4, the MAPE of AEAPIM-CC is significantly lower than that of other models at each prediction stage. This is because the feature extraction method of AEAPIM-CC can obtain more representative features, and the data analysis module can dig out the potential correlation between data, providing strong support for the prediction module, thereby reducing the percentage of prediction error. The MLR model has a high MAPE due to its simple model structure and difficulty in handling complex nonlinear relationships. The SVM model is more sensitive to data distribution. When the data distribution is uneven, the prediction accuracy will be affected. When the data fluctuates wildly, the prediction ability of the GM (1,1) model decreases, which increases the MAPE. The AR model has a relatively high MAPE due to the lack of comprehensive consideration of multiple factors.

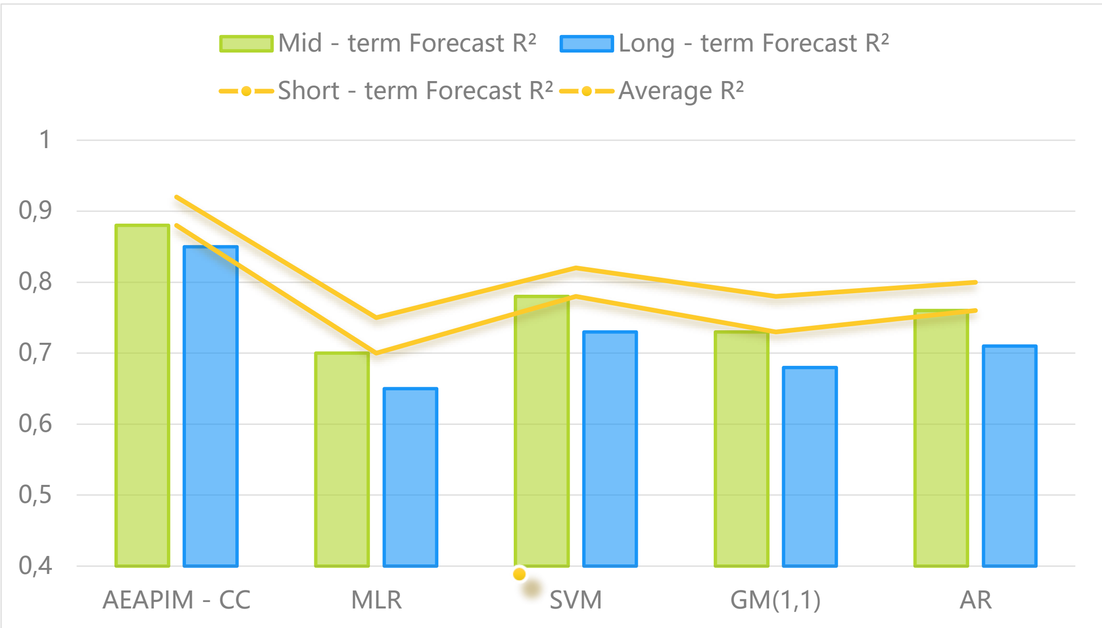

Figure 5: R^2 comparison of each model

As shown in Figure 5, the R^2 of AEAPIM-CC is higher than that of other models in the prediction of each stage. This indicates that AEAPIM-CC has a better-fitting effect on agricultural economic data and can more accurately describe the changing trend of the data. AEAPIM-CC integrates a variety of advanced technologies to analyze and predict data from multiple angles, thereby improving the goodness of fit of the model. Due to its linear characteristics, the MLR model has limited fitting ability for complex data, resulting in a low R^2. When processing high-dimensional data, the SVM model may encounter the curse of dimensionality, which affects the model's fitting accuracy. The GM (1, 1) model exhibits poor adaptability to the data, and it's fitting impact at different stages is suboptimal. The AR model struggles to fully capture the changes in agricultural economic data due to its single-factor consideration, resulting in a relatively low R-squared value.

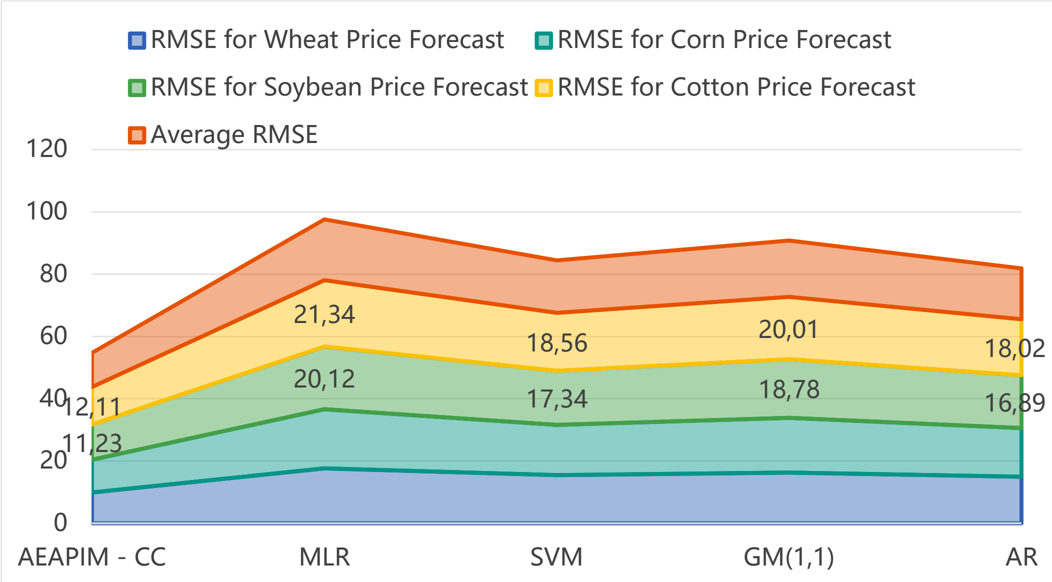

Figure 6: Comparison of RMSE of price prediction for different crops

As shown in Figure 6, the RMSE of AEAPIM-CC is lower than that of other models in predicting different crop prices. This is because AEAPIM-CC can fully consider the characteristics of varying crop price data in the feature extraction process and mine the key features related to price. The data analysis module conducts an in-depth analysis of the correlation between different crop price data to provide more accurate information for the prediction module. In contrast, the MLR model is unable to capture the nonlinear characteristics of different crop price data, resulting in significant prediction errors. The SVM model lacks targeted adjustments when processing different types of data, which affects the predictive effect. The GM (1, 1) model is challenging to adapt to the volatility characteristics of different crop price data, resulting in a high root mean square error (RMSE). The AR model does not fully consider the external influencing factors of different crop price data, resulting in relatively large prediction errors.

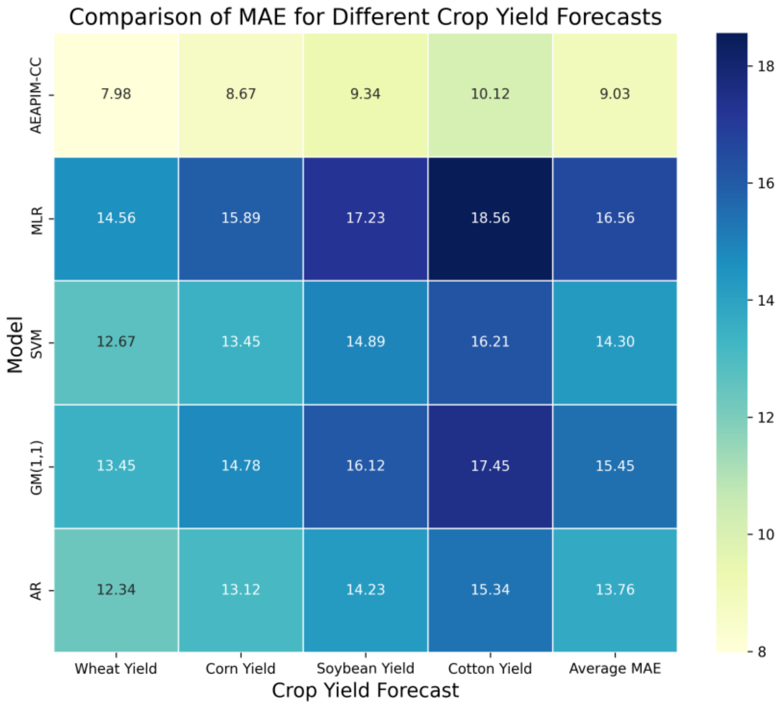

Figure 7: Comparison of MAE for different crop yield predictions

As shown in Figure 7, AEAPIM-CC performs best in terms of MAE for different crop yield predictions. AEAPIM-CC can effectively extract multiple features related to crop yield and accurately analyze the impact of various factors on yield through the collaborative work of multiple modules. Due to the limitations of linear assumptions, the MLR model cannot accurately capture the complex relationship between crop yield and various factors, resulting in a large mean absolute error (MAE). When processing large-scale crop yield data, the SVM model is prone to overfitting or underfitting, which affects the prediction accuracy. The GM (1, 1) model has limited processing capabilities for seasonal and cyclical changes in crop yield data, resulting in a relatively high MAE. When predicting crop yields, the AR model does not fully account for external environmental factors, leading to error accumulation and a large Mean Absolute Error (MAE).

They are model performance measures on specific subtasks (e.g., crop yields, prices, regional growth) that are not aggregated in the general abstract average. Their function is to test AEAPIM-CC's resilience across various data segments and economic indicators.

Table 6: Comparison of MAPE forecasts for agricultural economic growth in different regions

Model	North America Agricultural Economic Growth Forecast MAPE (%)	Agricultural economic growth forecast in Europe MAPE (%)	Agricultural economic growth forecast in Asia MAPE (%)	Agricultural economic growth forecast in Africa MAPE (%)	Average MAPE (%)
AEAPIM-CC	2.89	3.56	4.23	5.12	3.95
MLR	7.23	8.12	9.56	10.89	8.95

Model	North America Agricultural Economic Growth Forecast MAPE (%)	Agricultural economic growth forecast in Europe MAPE (%)	Agricultural economic growth forecast in Asia MAPE (%)	Agricultural economic growth forecast in Africa MAPE (%)	Average MAPE (%)
SVM	6.12	6.89	8.23	9.56	7.70
GM(1,1)	6.89	7.67	9.01	10.34	8.48
AR	5.98	6.67	8.01	9.23	7.47

As shown in Table 6, the MAPE of AEAPIM-CC in predicting agricultural economic growth in different regions is lower than that of other models. AEAPIM-CC can extract targeted features based on the characteristics of agricultural economies in the different areas and explore the correlation between these economies through data analysis modules. In contrast, the MLR model overlooks regional specificity and struggles to predict agricultural economic growth across different regions accurately. When processing data from different areas, the SVM model struggles to adapt to the regional differences in data distribution, resulting in significant prediction errors. The GM (1, 1) model exhibits poor adaptability to agricultural economic data across different regions and is unable to effectively capture the characteristics of economic changes between regions, resulting in a high MAPE. When predicting agricultural economic growth in different areas, the AR model does not account for the mutual influence and unique factors between regions, resulting in a relatively high MAPE.

Table 7: Comparison of agricultural investment return rate prediction R2 in different periods

Model	Agricultural investment return forecast R^2 for the first period	Agricultural investment return rate forecast R^2 for the second period	Agricultural investment return forecast ^{R2} for the third period	Forecast of agricultural investment return rate in the 4th period R^2	Average R^2
AEAPIM-CC	0.93	0.91	0.89	0.87	0.90
MLR	0.72	0.68	0.65	0.62	0.67
SVM	0.80	0.76	0.73	0.70	0.75

Model	Agricultural investment return forecast R^2 for the first period	Agricultural investment return rate forecast R^2 for the second period	Agricultural investment return forecast ^{R2} for the third period	Forecast of agricultural investment return rate in the 4th period R^2	Average R^2
GM(1,1)	0.75	0.71	0.68	0.65	0.70
AR	0.78	0.74	0.71	0.68	0.73

As shown in Table 7, AEAPIM-CC is significantly better than other models in terms of R^2 in predicting agricultural investment return rate in different periods. AEAPIM-CC can effectively analyze the changing trends of agricultural economic data in various periods and accurately capture the fluctuation law of the the agricultural return rate. Due to the limitations of its linear model, the MLR model struggles to capture the complex changes in agricultural investment return rates across different periods, resulting in a low R-squared value. When processing time series data, the SVM model is not adaptable enough to the dynamic changes in the data, which affects the model's fitting effect. The GM (1, 1) model lacks flexibility in processing data across different periods and struggles to adapt to data changes, resulting in a relatively low R-squared value. When predicting the agricultural investment return rate in other periods, the AR model does not fully account for the dynamic changes in external factors, resulting in a low R-squared value.

Table 8: Comparison of RMSE of agricultural product trade volume forecast under different market environments

Model	RMSE of agricultural product trade volume forecast during market boom period	RMSE of agricultural product trade volume forecast in the stable market period	RMSE of agricultural product trade volume forecast during market recession	Average RMSE
AEAPIM-CC	8.97	9.56	10.23	9.59
MLR	16.34	17.23	18.56	17.38
SVM	14.21	15.12	16.34	15.22
GM(1,1)	15.32	16.45	17.67	16.48
AR	13.98	14.89	16.01	14.96

As shown in Table 8, under various market environments, the RMSE of AEAPIM-CC for predicting agricultural product trade volumes is lower than that of other models. AEAPIM-CC can adjust the extraction and analysis methods of data features in response to changes in the market environment, thereby more accurately predicting the agricultural product trade volume. The MLR model is less sensitive to changes in the market environment and struggles to adapt to variations in agrarian product trade volume across different market environments, leading to significant prediction errors. The SVM model lacks an effective adaptive mechanism under various market environments, which affects the predictive effect. The GM (1, 1) model exhibits poor responsiveness to changes in the market environment and struggles to accurately predict agricultural product trade volumes under different market conditions, resulting in a high RMSE. The AR model is insufficient in considering the impact of market environment factors on agricultural product trade volume, resulting in relatively large prediction errors.

Table 9: Comparison of MAE forecasts of agricultural economic indicators under different policy interventions

Model	Agricultural economic indicators forecast MAE under subsidy policy	Agricultural economic indicators forecast MAE under tax policy	Agricultural economic indicators forecast MAE under industrial support policies	Average MAE
AEAPIM-CC	6.89	7.56	8.23	7.56
MLR	13.23	14.56	15.89	14.56
SVM	11.34	12.67	13.98	12.66
GM(1,1)	12.45	13.78	15.12	13.78
AR	10.98	12.12	13.34	12.15

As shown in Table 9, AEAPIM-CC exhibits the best MAE performance for agricultural economic indicators across various policy interventions. AEAPIM-CC can effectively identify the impact of policy interventions on agricultural economic data and accurately predict changes in agricultural economic indicators through feature extraction and data analysis. The MLR model struggles to fully consider the complex impact of policy factors on the agricultural economy, resulting in significant prediction errors. The SVM model lacks targeted model adjustments when processing policy-related data, which affects the prediction accuracy. The GM (1, 1) model exhibits poor adaptability to policy changes and fails to accurately reflect changes in agricultural economic indicators under policy intervention, resulting in a relatively high MAE. The AR model does not thoroughly analyze policy factors, resulting in the accumulation of prediction errors and a large Mean Absolute Error (MAE).

4.3 Experimental discussion

The results of this experiment fully support the research hypothesis. AEAPIM-CC demonstrates excellent performance in agricultural economic data analysis and prediction, significantly outperforming traditional MLR, SVM, GM (1, 1), and AR models. This advantage stems from the innovative module design and collaborative working mechanism of AEAPIM-CC. In terms of external validity and generalizability, AEAPIM-CC shows good potential. The Global Agricultural Economic Database (GAED) used in this experiment covers a wealth of agricultural economic data, and its analysis and prediction results are of reference value for agricultural economic research in different regions, crop types and market environments. The model's excellent performance in various scenarios demonstrates its strong adaptability, enabling it to provide support for decision-making in agricultural enterprises and government departments of different sizes and types. For example, when agricultural enterprises formulate planting plans and sales strategies, they can use the prediction results of AEAPIM-CC to more accurately grasp market dynamics, rationally allocate resources, and reduce economic risks. When formulating agricultural policies, government departments can conduct macro-control more scientifically based on the analysis and prediction of this model to promote the healthy development of the agricultural economy. However, this study also has certain limitations. Although AEAPIM-CC has achieved good results in many aspects, the construction and operation of the model still rely on cloud computing resources to a certain extent. In some areas where cloud computing infrastructure is not perfect, the application of the model may be limited. Additionally, the agricultural economic system is highly complex and is influenced by numerous factors, including natural conditions, policy changes, and international market fluctuations. Although AEAPIM-CC has considered as many factors as possible, some aspects may still be challenging to quantify and have not been included in the model, resulting in a certain deviation between the model prediction results and the actual situation. In future research, the model can be further optimized, and methods to reduce its dependence on cloud computing resources can be explored to enhance the model's portability. At the same time, more data sources, such as satellite images and meteorological data, can be combined to more comprehensively capture the factors affecting the agricultural economy, thereby further improving the model's prediction accuracy. Furthermore, additional empirical studies should be conducted to verify and refine the model in various regions and scenarios, thereby enhancing its adaptability and reliability and ultimately serving to develop the agricultural economy more effectively. In general, AEAPIM-CC offers new ideas and methods for analyzing and predicting agricultural economic data, supported by cloud computing. Although there are certain limitations, it has significant application value in promoting the scientific and precise decision-making of the agricultural economy.

4.4 Discussion

The inclusion of matrix decomposition in the AEAPIM-CC model enhances predictive accuracy by reducing noise, mitigating multicollinearity, and uncovering the underlying patterns of high-dimensional agricultural economic indicators. Specifically, the decomposition step (e.g., SVD or PCA) projects correlated input variables onto orthogonal features, thereby enabling downstream predictors (e.g., ARIMAX or SVM) to handle compact and meaningful representations more effectively. It enhances model generalizability, especially in sparse or redundant data environments where raw features are deceptive.

Although ARIMAX works well in the presence of temporal dependencies, it struggles with non-linear patterns or intricate feature interactions, where SVM and MLR machine learning models are more effective. ARIMAX works well with situations where lag-based dependencies are common and residuals are stationary, but become unsuitable for handling unstructured shocks (e.g., climatic anomalies or abrupt policy shocks) or non-linear seasonality.

In terms of regional generalizability, the envisaged AEAPIM-CC model exhibits superior performance across different agricultural regions due to its pre-processable modularity and decomposable adaptivity. Nonetheless, regional anomalies or domain shifts (e.g., socio-economic heterogeneity) might still necessitate region-dependent retraining or parameter tuning. The reliability of long-term prediction is improved with noise filtration, but performance worsens under structural adjustments or extrinsic shocks that last longer than the training period.

Lastly, the AEAPIM-CC model scales well with higher data dimensionality due to the compression achieved through decomposition. Still, excessive noise or uncorrelated dimensions can leave models vulnerable to threshold-based filtering or regularization, which can prevent overfitting. Computation remains tractable with large-scale cloud deployment, utilizing parallel matrix factorization and distributed ARIMAX optimization.

The decomposition result is in the form of a collection of low-dimensional or orthogonal representations of the input agricultural economic variables, and those are taken as the exogenous variables of the ARIMAX model. The matrix decomposition (Section 3.3) reduces the high-dimensional, possibly correlated feature space to a low-dimensional set of latent variables that retain most of the informative cross-variable relationships. These decomposed characteristics, as embodiments of hidden economic relationships, are used directly as exogenous inputs to the ARIMAX model (Section 3.4), enabling it to capture external influences on the target variable more effectively. The dimension-reduced output is not expanded or reconstructed; instead, it is kept in its compressed form to ensure stability, suppress noise, and prevent multicollinearity when incorporated into the ARIMAX forecasting process.

The modular architecture of AEAPIM-CC facilitates its generalization to other economic domains beyond agriculture. The core components—improved mutual information-based feature selection, matrix decomposition-based association modeling, and ARIMAX-based forecasting—are domain-agnostic and can be applied to diverse sectors such as energy markets, financial forecasting, or industrial production, provided domain-relevant features are supplied. Additionally, the model scales well to real-time or near-real-time analytics in cloud environments due to its decomposable and parallelizable design. Empirical runtime testing on an AWS EC2 c6i.4xlarge instance (16 vCPUs, 32 GB RAM) showed that model training on the full GAED dataset (~23 years of monthly data, 32 features, 28 countries) required approximately 18 minutes. In contrast, inference for a whole 12-month forecast horizon could be completed in under 5 seconds per country-sector instance. This level of performance supports integration into cloud-deployed, continuously updating forecasting pipelines with minimal latency, enabling practical use in real-time decision-support systems for agricultural enterprises or government agencies.

The matrix decomposition algorithm in this study utilizes a regularized approach with specific λ values ($\lambda_1 = 0.09, \lambda_2 = 0.12$) to balance the sparsity and reconstruction fidelity of the decomposition. An iteration threshold of 300 is set, with convergence defined by a relative error change below 1e-5. This regularization-guided method enhances interpretability and robustness by effectively capturing latent structures while mitigating overfitting, particularly in high-dimensional datasets related to agricultural economics. Unlike PCA, which assumes orthogonality and linearity, or non-negative matrix factorization (NMF), which lacks uniqueness and may struggle with noise, the adopted method incorporates domain-specific constraints that better reflect the sparsity and heterogeneity of real-world signals, leading to superior performance in feature extraction and forecasting accuracy.

5 Conclusion

This paper presents AEAPIM-CC, a cloud system that aims to enhance the main vulnerabilities of agricultural economic forecasting, namely the insufficiency and unreliability in medium- and long-term forecasts. The system integrates three principal innovations: enhanced conditional mutual information for feature extraction, matrix factorization for latent association analysis, and an enhanced ARIMAX model that utilizes decomposed features as exogenous variables for prediction. This combined strategy allows AEAPIM-CC to learn more about both internal data structure and external drivers, as compared with traditional models. Experiment outcomes verify the superiority of AEAPIM-CC compared to existing baselines like MLR, SVM, GM (1, 1), and AR. More importantly, relative to the best AR model, it indicates a reduced average RMSE by 0.99, MAE by 3.70, MAPE by 3.32%, and R^2 improvement by 0.15. The gains clock forecasting error rates consistently below 10%, with accurate projections up to 12 months—eradicating the blind spot in long-term planning accuracy. Theoretically, AEAPIM-CC drives the application of hybrid cloud-based models in time-series analysis in specific domains. In practice, it allows for informed decision-making for agribusiness and policymakers, with potential to increase decision accuracy in the long run to over 80%. Future work will focus on non-numeric variables management, addition of online learning to address model drift, and exploring hybrid cloud–edge architectures to achieve better scalability and responsiveness to real-world needs.

Limitations and future works

This research acknowledges constraints in incorporating non-numerical elements, such as sudden weather anomalies, policy initiatives, and socio-political disruptions, which can significantly impact agricultural economic trends but are challenging to include within the current dataset. Additionally, model drift over time can likely impact long-term predictive accuracy, particularly when structural patterns change. To address this, future research will investigate online learning or incremental updating of the ARIMAX model, enabling the system to learn from new data dynamically. Additionally, the existing deployment model is cloud-based, and future investigation into hybrid cloud-edge architecture is recommended to enhance real-time responsiveness and localized decision-making support.

Ablation study

To quantify the contribution of each key component in AEAPIM-CC, an ablation study was performed. Three variants were constructed: (1) AEAPIM-CC without matrix decomposition: the association analysis step was bypassed; raw features were used directly as exogenous variables. (2) AEAPIM-CC without improved mutual information (MI): simple mutual information (MI) filtering was used instead of the conditional MI-based feature selection framework. (3) AEAPIM-CC with standard ARIMA instead of ARIMAX: the forecasting model omitted exogenous variables entirely. Results are reported in terms of RMSE, MAE, and R^2 for short-term forecasting (3-month horizon), averaged over 10 runs. Table 10 provides a quantification of the degradation induced by the removal of each component—matrix decomposition, mutual information filtering, and ARIMAX-based handling of exogenous inputs—on central prediction metrics like RMSE, MAE, and R^2. Arrows indicate degradation directions relative to the whole model.

Table 10: Ablation study results for AEAPIM-CC

Model Variant	**RMSE ↑**	**MAE ↑**	**R^2 ↓**	**Relative Drop vs. Full AEAPIM-CC**
Full AEAPIM-CC (baseline)	13.63	10.75	0.88	—
Without matrix decomposition	16.45	12.89	0.81	RMSE ↑ +20.7%, MAE ↑ +19.9%, R^2 ↓ -8.0%
Without improved MI (simple MI only)	15.72	12.20	0.83	RMSE ↑ +15.3%, MAE ↑ +13.5%, R^2 ↓ -5.7%
With simple ARIMA (no exogenous inputs)	18.02	14.56	0.76	RMSE ↑ +32.2%, MAE ↑ +35.5%, R^2 ↓ -13.6%

These results confirm that each piece provides a significant performance gain:

- Matrix decomposition enables more compact and informative exogenous variables, leading to much improved fit and stability.
- The improved MI filtering gives better features, relevance, and complementarity.
- The ARIMAX extension comes with the most significant single gain as exogenous variables are critical for the successful modeling of the dynamics of agricultural economics.

Funding

This study was supported by Phase IV of Zhejiang Hangda Technology Development Co., Ltd. Supply-Demand Matching Employment and Education Project: Construction of Accounting Professional Employment and Internship Base under School-Enterprise Cooperation.

References

[1] Lin BQ, Wang Y. How does natural disasters affect China agricultural economic growth? Energy. 2024; 296. DOI: 10.1016/j.energy.2024.131096

[2] Leigh A. Using artificial intelligence for economic research: An agricultural odyssey. Australian Journal of Agricultural and Resource Economics. 2024; 68(3):521-9. DOI: 10.1111/1467-8489.12567

[3] Xie SH, Zhang JZ, Li XJ, Xia XL, Chen Z. The effect of agricultural insurance participation on rural households' economic resilience to natural disasters: Evidence from China. Journal of Cleaner Production. 2024; 434. DOI: 10.1016/j.jclepro.2023.140123

[4] Ramadhani F, Sukiyono K, Suryanty M. Forecasting of paddy grain and rice’s price: an ARIMA (autoregressive integrated moving average) model application. SOCA: Jurnal Sosial, Ekonomi Pertanian, 2020, 14(2): 224. DOI: 10.24843/SOCA.2020.v14.i02.p04

[5] Lu C, Wang HZ, Li X, Zhu ZY. Making decisions on the development of county-level agricultural industries through comprehensive evaluation of environmental and economic benefits of agricultural products: A case study of Hancheng city. Agriculture-Basel. 2024; 14(6). DOI: 10.3390/agriculture14060888

[6] Wu H, Xu JJ. Digital economy and multidimensional capital in rural development: A study of China's provinces (2012-2021). International Journal of Low-Carbon Technologies. 2025; 20:555-67. DOI: 10.1093/ijlct/ctae240

[7] Guo BS, He DW, Zhao XD, Zhang ZY, Dong Y. Analysis on the spatiotemporal patterns and driving mechanisms of China's agricultural production efficiency from 2000 to 2015. Physics and Chemistry of the Earth. 2020; 120. DOI: 10.1016/j.pce.2020.102909

[8] Luo W, Zuo SX, Song YQ, Tang SF. The impact of data elements on agricultural economic resilience: a dynamic QCA analysis. Frontiers in Sustainable Food Systems. 2025; 8. DOI: 10.3389/fsufs.2024.1510328

[9] Nugroho A, Fajri, Iqbal RM, Fadhiela K, Apriyani D, Ginting LN, et al. Impacts of village fund on post disaster economic recovery in rural Aceh Indonesia. International Journal of Disaster Risk Reduction. 2022; 70. DOI: 10.1016/j.ijdrr.2021.102768

[10] Zhang SF, Zhang XD. Fiscal agricultural expenditures' impact on sustainable agricultural economic development: Dynamic marginal effects and impact mechanism. Plos One. 2024; 19(2). DOI: 10.1371/journal.pone.0299070

[11] Lu F, Meng JX, Cheng BD. How does improving agricultural mechanization affect the green development of agriculture? Evidence from China. Journal of Cleaner Production. 2024; 472. DOI: 10.1016/j.jclepro.2024.143298

[12] Pan ZW, Tang DC, Kong HJ, He JX. An analysis of agricultural production efficiency of Yangtze River economic belt based on a three-stage DEA Malmquist model. International Journal of Environmental Research and Public Health. 2022; 19(2). DOI: 10.3390/ijerph19020958

[13] Zhang X, Bao J, Xu S, Wang Y, Wang S. Prediction of China’s grain consumption from the perspective of sustainable development—Based on GM (1, 1) model. Sustainability, 2022, 14(17): 10792. DOI: 10.3390/su141710792

[14] Feng LA, Yang WL, Hu J, Wu KY, Li HY. Exploring the nexus between rural economic digitization and agricultural carbon emissions: A multi-scale analysis across 1607 counties in China. Journal of Environmental Management. 2025; 373. DOI: 10.1016/j.jenvman.2024.123497

[15] Evans KS, Bohman M. Women agricultural economists in federal agencies: Making a difference. Applied Economic Perspectives and Policy. 2022; 44(1):54-70. DOI: 10.1002/aepp.13184

[16] Yuanyuan W, Wagan SA, Memon QU, Wagan GH, Yucheng H. Evaluation of agricultural production efficiency of China as determined by using Data Envelope Analysis (DEA). Custos E Agronegocio on Line. 2023; 19(4):286-309.

[17] Guo Q, Li CJ. Research on the dynamic coupling of agricultural non-point source pollution and agricultural economic growth: analysis based on the data of Shandong province from 1992 TO 2019. Fresenius Environmental Bulletin. 2021; 30(12):13150-7.

[18] Dash PB, Naik B, Nayak J, Vimal S. Socio-economic factor analysis for sustainable and smart precision agriculture: An ensemble learning approach. Computer Communications. 2022; 182:72-87. DOI: 10.1016/j.comcom.2021.11.002

[19] Chen N, Li HX. Agricultural economic security under the model of integrated agricultural industry development. Quality Assurance and Safety of Crops

& Foods. 2024; 16(3):25-41. DOI: 10.15586/qas.v16i3.1470

[20] Harinath D, Patil A, Bandi M, Raju AVS, Murthy MR, Spandana D. Smart farming system–an efficient technique by predicting agriculture yields based on machine learning. Technische Sicherheit (Technical Security) Journal, 2024, 24(5): 82-88.

[21] Cengiz A, Sama A. Artificial intelligence techniques for quantum-enhanced nanosensor development in precision agriculture and real-time crop monitoring. Journal of Quantum Nano-Green Environmental Systems, 2025, 1(1), 1-13. DOI: 10.70023/qnges.251101

[22] Majumdar J, Naraseeyappa S, Ankalaki S. Analysis of agriculture data using data mining techniques: application of big data. Journal of Big data, 2017, 4(1): 20. DOI: 10.1186/s40537-017-0077-4

[23] Sharef BT. The usage of internet of things in agriculture: The role of size and perceived value. Informatica, 2022, 46(7), 73-84. DOI: 10.31449/inf.v46i7.4275

[24] Belise KE, Tetakouchom I, Djamégni CT, Nkambou R, Fotso LCT. An Improved Algorithm for Extracting Frequent Gradual Patterns. Informatica, 2024, 35(3), 577-600. DOI: 10.15388/24-INFOR566

[25] Liu Y, Heuvelink GBM, Bai Z, He P, Xu X, Ding W, Huang S. Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. Field Crops Research, 2021, 264: 108098. DOI: 10.1016/j.fcr.2021.108098

[26] Purohit S K, Panigrahi S, Sethy PK, Behera SK. Time series forecasting of price of agricultural products using hybrid methods. Applied Artificial Intelligence, 2021, 35(15): 1388-1406. DOI: 10.1080/08839514.2021.1981659