SVM Based Agricultural Crop Price Prediction Model

Hongliang Cheng, Aijv Huang

***Abstract*—Considering the impact of crop prices on the agricultural market and economy, in order to manage crop prices more accurately, an intelligent crop price prediction model is proposed in this study. Firstly, agricultural crop data information is collected and analyzed; Factors affecting crop prices is explored. Support vector mechanisms is introduced to build a crop price prediction model. In addition, by analyzing crop time series data through autoregressive moving average models, the problem of insufficient handling of linear problems in prediction models can be improved, thereby achieving effective analysis of crop prices. In the prediction of the autoregressive moving average model, through the residual test, the constructed model conforms to the normal distribution, and the model parameters are determined to be (1,1,8). In the comparison of peanut price prediction, the average prediction accuracy of the autoregressive moving average model and the long-term and short-term network model is 81.65% and 81.65%. In the combined model prediction, the average accuracy of the combined model is 95.65%. Compared with the long-term and short-term network model and the autoregressive moving average model, the prediction accuracy is improved by 23.65% and 12.65%. In addition, the study also conducted predictions on crop investment risks. Compared with similar technologies, the investment risk prediction of the research model is close to the actual value, and the accuracy of investment risk prediction is 0.9725, which is significantly better than similar models. It can be seen that research models can accurately analyze and predict crop prices and risks, providing effective reference opinions for agricultural growers and meeting the needs of agricultural development.**

***Index Terms*—autoregressive moving, data mining, influencing factors, price prediction model, support vector machine**

I. INTRODUCTION

As the call for sustainable development continues to grow, the sustainable development of agriculture has also received widespread attention. In this context, information-based agricultural management technology provides important technical support for the sustainable development of modern agriculture [1]. As one of the core indicators of the agricultural market, crop prices have an important impact on farmers and the entire social economy. Since crop prices are affected by many nonlinear complex factors, traditional linear models are often difficult to accurately predict the trend of crop prices [2]. At the same time, traditional prediction models also have certain difficulties in analyzing crop time series data, and linear models are often ineffective when processing such data [3]. Therefore, in order to effectively predict crop prices, the study proposes an intelligent crop price prediction model, aiming to overcome the shortcomings of traditional crop price management technology. First, through the collection and analysis of agricultural crop data information, the research will explore the influencing factors of crop prices in order to more accurately predict changes in crop prices. Secondly, in order to solve the problem that crop prices are affected by nonlinear complex factors, the study introduced support vector machine (SVM) to build a crop price prediction model. At the same time, in order to consider the linear influencing factors of crop prices, the study also introduced the Autoregressive Integrated Moving Average Model (ARIMA) to analyze the time series data of crops to achieve effective analysis of crop prices. There are two innovations in the research. Firstly, in order to accurately analyze crop data, SVM and ARIMA are introduced to model time series data, ensuring the accuracy of price analysis. The second is to introduce residual testing to modify the price prediction model and improve the application effect of the technology through parameter optimization. Research technology can effectively accelerate the construction of agricultural informatization, and at the same time, effective agricultural data analysis can provide technical support for predicting agricultural prices and managing investment risks.

The research content is divided into four parts. The first part introduces the technologies related to the development of agricultural digital transformation, and conducts research and discussion on the technologies related to agricultural price prediction. The second part studies crop price-related data. In order to better manage prices, a machine learning model is introduced to build a price prediction model to achieve prediction and scientific management of agricultural crops. The third part is to apply the mentioned technology to specific scenarios and verify the application effect of the proposed price prediction model in actual scenarios. The fourth part summarizes and analyzes the full text and elaborates on the improvement direction of the research.

II. RELATED WORK

Agriculture is the foundation of national economic development. Under a sustainable background, promoting the development of agricultural digitalization is of great significance to agricultural and regional economic development. Birner R and others conducted research on the digital transformation of modern agriculture. The integration

Manuscript received January 22, 2024; revised November 11, 2024.

Hongliang Cheng is a teacher of Linzhou College of Architectural Technology, Anyang 456550, China. (corresponding author, email: lzjzchl@163.com).

Aijv Huang is a lecturer of Jiaozuo Normal College, Jiaozuo 454000, China. (e-mail: ada_hd@126.com).

of information digital technology provides important technical support for modern agricultural production and management. However, some people believe that the construction of digital agriculture will pose a threat to market concentration. However, the current alliance between participants and public actors is actually beneficial to agricultural development and has good application effects [4]. Sharifi A and others used remote sensing and digital technology to analyze wheat yield. At the same time, a machine learning model is introduced to build a prediction model to achieve effective prediction of wheat yield by detecting wheat development and growth status. Through testing, the prediction model has good performance in practical applications and is better than other prediction models. This technology can provide more scientific management opinions for wheat planting [5]. Storm H and others analyzed the characteristics of machine learning from the perspective of digital economics. Advanced digital statistical technology can improve the calculation and statistical effects of economic data, but it still faces many limitations in complex data management and predictive analysis. In order to improve the application effect of the model in actual scenarios, the source data and prediction model are parameterized to improve the processing effect of complex data. The results show that in the digital context, machine learning models can improve the monitoring and management effects of agricultural data and provide technical support for agricultural digital transformation [6]. Misra NN and others found that digital agriculture based on big data technology can effectively achieve food safety management through a large number of sensors and digital processing technologies. In this regard, the application effect of Internet of Things technology in agricultural digital transformation is analyzed, and the application effect of technology in food processing, testing, and safety fields is evaluated. Actual results show that IoT technology has significantly improved the environmental management effects of agricultural food in production, processing, testing and safety, and promoted sustainable agricultural development [7].

The SVM model has excellent application effects in agricultural price prediction and promotes the modernization of agricultural management. Purohit S KL and others found that agricultural price fluctuations will have a severe impact on the development of the agricultural market, and effective crop value monitoring will help farmers better evaluate the value of crops and choose the appropriate time to sell. Therefore, in order to effectively monitor crop values, the SVM model is used to build a crop price prediction model. First of all, considering that crops are affected by many adverse factors, including weather, demand, quantity, etc., multiple period models are introduced to monitor crops, and digital statistical models are introduced for data analysis to estimate the price of agricultural products. By implementing the analysis, the proposed technique has good prediction results. However, the study did not consider macroeconomic and policy factors and needs further improvement [8]. Wang L et al. reviewed agricultural product price prediction methods and related data in recent years to achieve effective prediction of agricultural crops. A SVM crop price prediction model based on machine learning is proposed. At the same time, considering that crop prices are affected by many factors, the main influencing factors are screened through data mining technology, and the prediction of agricultural crop prices is achieved through model training. Through experiments, it can be seen that the yield and quality of crops have a direct impact on the value of crops. At the same time, compared with related price prediction models, it is found that the overall error of the proposed technology is lower and the prediction accuracy is higher. Compared with the proposed technology, this technology faces certain difficulties in linear data processing [9]. Das P et al. found that effective crop price prediction has an important impact on agricultural crop development and is of great significance in preventing agricultural risks. Therefore, a crop price prediction model is constructed based on SVM and empirical mode decomposition. The data set is decomposed into a limited and small number of sub-signals through empirical mode decomposition, and the SVM model is used to process and predict the signal values to obtain the final set prediction result. Through experiments, it can be seen that the proposed prediction model has good application effects, has lower error accuracy than other models, and is better than other technologies [10].

According to the above research, under sustainable development, digital transformation has accelerated the development of modern agriculture and promoted the upgrading and transformation of the agricultural industry. At the same time, agricultural crop prices have always been an important indicator of agricultural development, and the application of digital technology provides important technical support for agricultural construction. In this regard, an information-based crop price prediction model is constructed based on machine learning to more effectively monitor crop prices and provide important technical support for agricultural risk control and crop management.

III. Construction of Agricultural Crop Price Prediction Model Based on SVM

This part mainly excavates and analyzes the relevant influencing factors of crop prices, and builds a price prediction model based on SVM based on the characteristics of crop prices. At the same time, considering the linear characteristics of crop price data, ARIMA is introduced to analyze linear data to build a combined price prediction model.

A. Analysis of Factors Affecting Crop Prices

In the context of sustainable development, promoting the digital development of the agricultural trade industry is of great significance to the development of modern agriculture. Agricultural crop prices are affected by various environmental factors in the market, and crop prices fluctuate, which will have an impact on the entire agricultural economic industry chain [11]. In order to effectively respond to the impact of crop price fluctuations on agricultural development, factors affecting crop prices will be analyzed based on the agricultural crop information management system. The schematic diagram of the agricultural crop information management system is shown in Fig. 1.

In order to evaluate the influencing factors of crop prices, we first need to collect data information on crop prices.

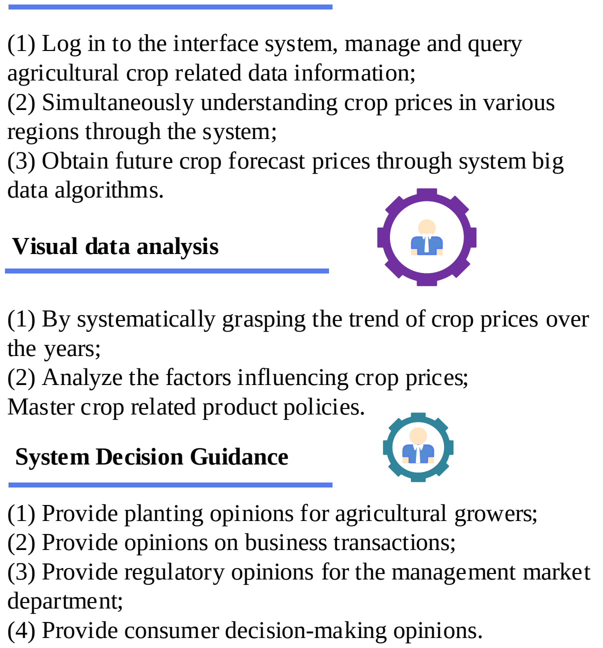


Fig. 1. Schematic diagram of agricultural crop information management system.

In the study, we mainly obtained data through the "National Compilation of Agricultural Products Composition Income Data" and online agricultural information in various regions. In order to screen out the main influencing factors, the study used Principal Component Analysis (PCA) to screen the data index characteristics and find out the main correlation factors that affect crop prices [12]. First, the data is standardized to obtain a standardized matrix Z. Standardization can eliminate dimensional differences between different indicators so that they have the same scale. The expression of data normalization is shown in Equation (1) [13].

$$Z_{ij} = \frac{X_{ij} - \mu_i}{\sigma_i} \tag{1}$$

In Equation (1), Z_{ij} is the standardized value of the i-th indicator in the j-th sample, X_{ij} is the raw data, μ_i and σ_i are the mean and standard deviation of the ith indicator respectively. Next, calculate the sample data matrix R, whose expression is shown in Equation (2).

$$R = \frac{1}{n}\sum_{j=1}^{n} Z_j Z_j^T \tag{2}$$

In Equation (2), n is the number of samples and Z_j is j the standardized matrix of the sample. Then, solve the characteristic equation to obtain the characteristic roots P and unit eigenvectors b. The eigenvalue P represents the importance of the principal component, and the unit eigenvector b represents the direction of the principal component. The expression of the characteristic equation is shown in Equation (3).

$$Rb = Pb \tag{3}$$

In Equation (3), R is the sample data matrix, P is the characteristic root matrix, and b is the unit eigenvector matrix. According to the characteristic roots P, the contribution rate of the principal components can be calculated. The contribution rate of the principal component indicates the extent to which the principal component explains the total variance and is an indicator for evaluating the importance of the principal component [14]. The calculation formula of the contribution rate of the principal component is shown in Equation (4).

$$\frac{\lambda_i}{\sum_{i=1}^{m}\lambda_i} \times 100\% \tag{4}$$

In Equation (4), λ is the i characteristic root. In order to select the main components, it is necessary to retain the principal components whose cumulative contribution rate exceeds 85%. That is, the first principal component is selected k from the characteristic roots P, as shown in Equation (5).

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{m}\lambda_i} \geq 0.85 \tag{5}$$

Finally, the selected m principal components are weighted and summed to obtain the score contribution ratio of each sample. The score contribution ratio indicates the degree of contribution of each principal component to the sample variables [15]. The calculation formula is as shown in Equation (6).

$$G_j = \frac{\sum_{i=1}^{m} b_{ij}\sqrt{\lambda_i}}{\sum_{i=1}^{m}\sqrt{\lambda_i}} \times 100\% \tag{6}$$

In Equation (6), G_j is j the score contribution ratio of the th sample, b_{ij} is the unit eigenvector matrix of the j-th sample, and λ_i is the i-th eigenroot. Through the above steps, the main correlation factors that affect crop prices can be obtained, as shown in Table I.

These factors can help agricultural producers and dealers better understand the changing trends of crop prices and thus formulate reasonable agricultural production and sales strategies. At the same time, based on the above factors, agricultural crop prices are predicted through the informationized agricultural crop system.

B. Construction of Agricultural Crop Price Prediction Model Based on SVM

The agricultural crop price prediction model based on SVM is a regression problem, and the nonlinear data processing capability of SVM can handle this problem well. Therefore, an agricultural crop price prediction model is constructed based on SVM. First, assume that the study has N sample data of agricultural crops, and each sample has d a characteristic. The study expresses the i-th sample as shown in Equation (7).

$$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{id})^T \tag{7}$$

In Equation (7), the corresponding target value is y_i. where $\mathbf{x}_i$ is a one- d dimensional vector, y_i a real number. The goal of the research is to construct an agricultural crop price prediction model and make the prediction results as accurate as possible. Assume that the output of the model is $f(\mathbf{x})$.

TABLE I
FACTORS INFLUENCING AGRICULTURAL CROP PRICES

Serial number	Factor	Factor symbol	Area
1	Consumer price index of residents	A1	Macro-factors
2	Material and service expenses	A2	Supply factor
3	Crop prices	A3	Demand factors
4	Exchange rate	A4	Macro-factors
5	Agricultural financial funds	A5	Macro-factors
6	Disposable income of rural residents	A6	Demand factors
7	Total population of the country	A7	Demand factors
8	Retail price index of goods	A8	Macro-factors
9	Other crop prices	A9	Demand factors
10	Investment in fixed assets	A10	Macro-factors
11	Regional guide price	A11	Demand factors
12	Land cost	A12	Supply factor
13	Average income of urban population	A13	Demand factors
14	Labor costs	A14	Supply factor
15	Crop yield	A15	Supply factor
16	Sown area	A16	Supply factor
17	Inflation rate	A17	Macro-factors

Research can define the objective function as minimizing the squared difference between the prediction result and the true value, that is, minimizing the loss function $J(\mathbf{w},b)$, and the calculation is as shown in Equation (8).

$$J(\mathbf{w},b_1)=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}L(y_i,f(\mathbf{x}_i)) \tag{8}$$

In Equation (8), $\mathbf{w}$ is the weight vector of the model, b_1 is the bias of the model, C is the regularization parameter, which is used to control the complexity of the model, L and is the loss function. The establishment process of SVM model is a convex optimization problem. Research needs to find the optimal solution when minimizing the objective function $J(\mathbf{w},b_1)$ [16]. To this end, the study introduces Lagrange multipliers to transform the problem into a Lagrangian dual problem. The Lagrange multiplier is $\alpha_i \geq 0$, corresponding to each sample. The Lagrangian function is shown in Equation (9).

$$L(\mathbf{w},b_1,\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}L(y_i,f(\mathbf{x}_i))-\sum_{i=1}^{N}\alpha_i(y_i-f(\mathbf{x}_i)) \tag{9}$$

In Equation (9), α_i represents the Lagrange multiplier. For any given $\mathbf{w}$ sum b_1, research can $L(\mathbf{w},b_1,\boldsymbol{\alpha})$ solve the optimal Lagrange multiplier by maximizing the Lagrange function α_i. The optimization problem can be expressed as shown in Equation (9).

$$\begin{aligned}&\max_{\boldsymbol{\alpha}} L(\mathbf{w},b_1,\boldsymbol{\alpha})=\\&\max_{\boldsymbol{\alpha}}\left[\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)\right]\end{aligned} \tag{10}$$

In Equation (10), $K(\mathbf{x}_i,\mathbf{x}_j)$ represents the kernel function. By solving the dual problem, the optimal one can be $\boldsymbol{\alpha}$ obtained . The dual problem is a convex quadratic programming problem that can be solved by some fast algorithms (such as the sequential minimum optimization algorithm).

Finally, by solving the obtained Lagrange multiplier α_i, the output of the model can be obtained $f(\mathbf{x})$, as shown in Equation (11).

$$f(\mathbf{x})=\sum_{i=1}^{N}\alpha_i y_i K(\mathbf{x},\mathbf{x}_i)+b_1 \tag{11}$$

The selection of the kernel function plays a very important role in the SVM model. It can map the original sample data into a high-dimensional feature space, thereby making the classification or regression problem become linearly separable or linearly separable in this high-dimensional space. Linear regression. Radial basis function (RBF) is selected as the kernel function of SVM, which has good nonlinear fitting ability [17]. The mathematical form of the radial basis function is shown in Equation (12).

$$K(\mathbf{x}_i,\mathbf{x}_j)=\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{12}$$

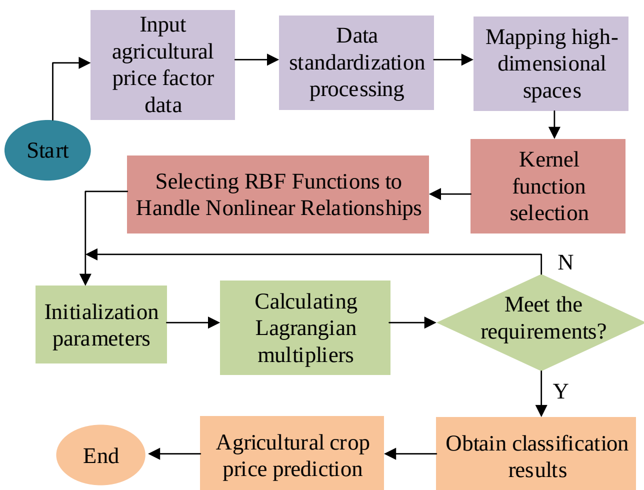


Fig. 2. SVM based agricultural crop price prediction model.

In Equation (12), where σ is the bandwidth parameter of the radial basis function. By selecting the radial basis function as the kernel function, the original agricultural crop characteristics can be mapped into a high-dimensional space, allowing the model to better fit nonlinear changes. The agricultural crop price prediction model based on SVM is shown in Fig. 2.

C. Construction of Agricultural Crop Price Prediction Model Based on ARIMA-SVM

The SVM model has good adaptability when dealing with nonlinear feature components, but faces difficulties when dealing with linear feature components alone. The ARIMA (p, d, q) model is a time series data analysis model, in which the three indices p, d, q, are used to parameterize the ARIMA model, which has good analysis results for linear feature data. Therefore, in order to optimize the shortcomings of the SVM model in linear feature data processing, the ARIMA model and the SVM model are combined to process crop feature data and improve the crop price prediction effect. The combined ARIMA-SVM prediction model includes a linear model, which can effectively extract linear feature components in time series. The nonlinear feature data is extracted and analyzed using the SVM model. The two complement each other, thereby improving the overall analysis effect of the feature data. The ARIMA model data analysis process is shown in Fig. 3.

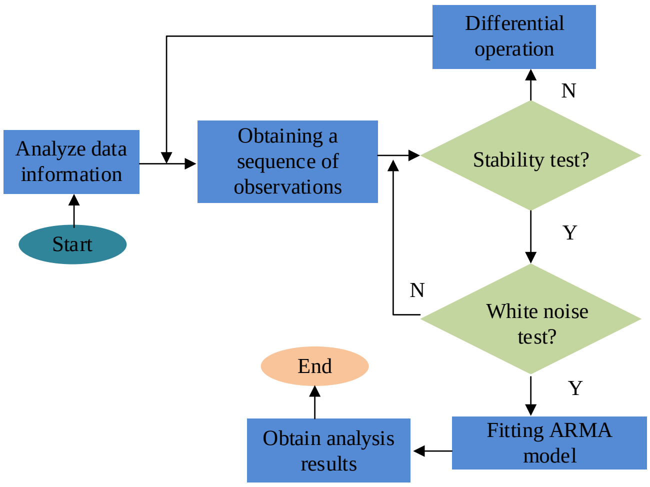


Fig. 3. ARIMA model data analysis process.

The ARIMA model contains a total of three types of models. One is the Autoregressive Model (AR), which mainly divides data categories based on the stability of the original data sequence [18]. The second type is the moving average model (MA), which mainly classifies data based on the different autoregressive terms in the model. The third category is the Autoregressive Moving Average Model (ARMA), which mainly classifies data based on the differences in the moving average terms in the model [19]. In the natural regression model, its expression is shown in Equation (13).

$$\begin{aligned} X_t &= \varphi_0 + \varphi_1 X_{T-1} + \varphi_2 X_{t-2} + \ldots \\ &+ \varphi_p X_{t-p} + \varepsilon_t, \varphi_p \neq 0 \end{aligned} \tag{13}$$

In Equation (13), φ_0, φ_2 and φ_p are all autoregressive coefficients of the model, X_{t-p} and X_{t-2}, and X_{t-p} represent the data sequence. The natural regression model meets the relevant requirements, as shown in Equation (14).

$$\begin{cases} E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_s^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s \\ E(X_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{14}$$

In Equation (14), $E(\varepsilon_t)$ represents the mean function, $\{\varepsilon_t\}$ represents the white noise description, σ^2 represents the variance function, t represents the period, $t, s \in T$. The expression of the moving average model is shown in Equation (15).

$$X_t = \mu + \varepsilon_t - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \theta_q \neq 0 \tag{15}$$

In Equation (15), μ represents a constant, and q the model order is adjusted by changing the model parameter value, θ_2 and θ_q both represent the moving average. At the same time, the moving average model needs to meet relevant conditions, as shown in Equation (16) [20].

$$E(\varepsilon_t) = 0, Var'(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t, \varepsilon_s) = 0, t \neq s \tag{16}$$

In the autoregressive moving average model, $\{X_t\}$ the observed value of the time series X_t at a moment is t, which is not only t related to it, but also related to the previous historical value [21]. The expression of the autoregressive moving average model is shown in Equation (17).

$$\begin{aligned} X_t &= \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots \\ &+ \varphi_p X_{t=p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \end{aligned} \tag{17}$$

In Equation (17), φ_0, φ_2 and φ_p are both autoregressive coefficients of the model, θ_2 and θ_q both represent the moving average of the model. At the same time, the autoregressive moving average model needs to meet relevant requirements, as shown in Equation (18) [22].

$$\begin{cases} \varphi_p = 0, \theta_q = 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_s^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s \\ E(X_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{18}$$

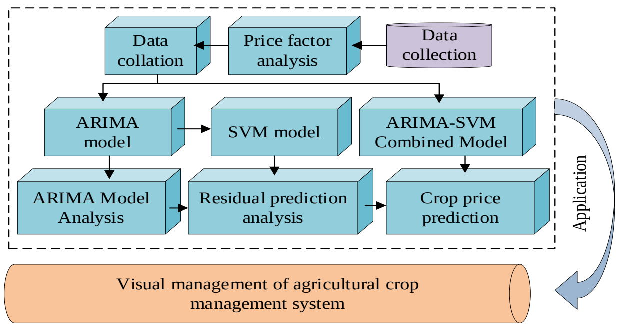


Fig. 4. Flow chart of ARMA-SVM combination prediction model.

Through the above research process, when dealing with linear problems, the ARIMA model is used for analysis and processing, and when dealing with nonlinear problems, the SVM model is used for analysis and processing [23]. The ARIMA-SVM model prediction process is shown in Fig. 4.

By using this combination of processing methods, agricultural crop characteristic data can be effectively processed, thereby improving the overall prediction performance of agricultural crop price prediction models. The entire study is based on the crop information management system, using PCA method to mine crop influencing factors and construct indicators of crop price correlation factors. In addition, an improved SVM model is introduced to train crop price data, achieving prediction and management of prices.

IV. Crop Price Prediction Model Simulation Test

This part mainly analyzes the selection of ARIMA model prediction parameters, time series data processing and comparison of prediction effects. At the same time, the performance of the proposed combined model was tested based on the ARIMA model parameters.

A. Crop Price Forecast Based on ARIMA Model

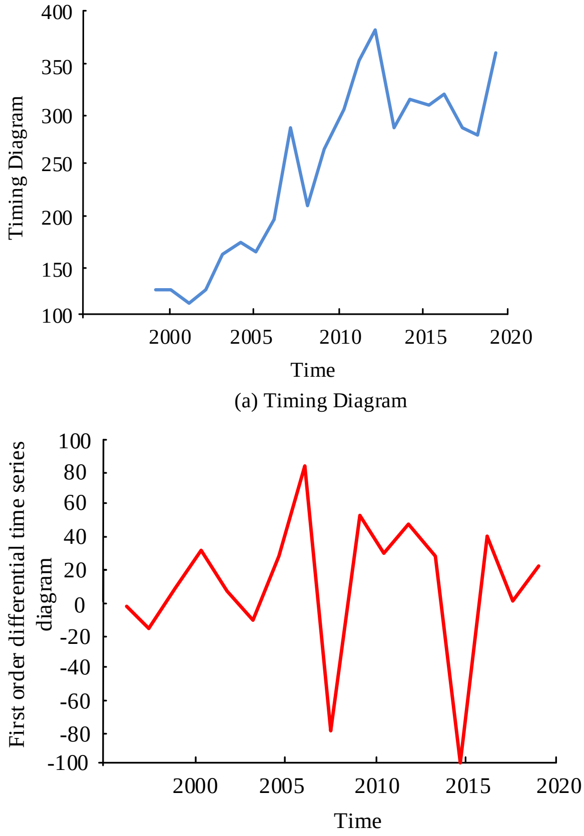
Test the performance of the proposed crop price prediction model, experimental testing will be carried out on the WINDOWS 10 64-bit platform, and simulation experimental analysis will be completed on the Matlab platform. Among them, the agricultural crop information management system is developed using JavaScript, and the system development framework is JQuery, which is compatible with various browser devices. At the same time, the KO framework is used as the system data model, and its advantages can achieve data processing and presentation. Therefore, based on the agricultural crop information management system, the peanut price data from 1995 to 2020 was selected as the ARIMA model training time series analysis data set, and the collected 17 crop price influencing factors were used as the SVM model training data set. The initialization parameters of the AR I MA-SVM model are shown in Table II.

TABLE II
MODEL INITIAL PARAMETERS

Parameter indicator type	Numerical value
SVM model gamma value	0.1768
SVM model epsilon value	0.11
SVM model penalty value	0.1
Bandwidth parameter value	92.8405
SVM model threshold	0.5

Drawing the time series of the peanut crop based on the data set is the key to model analysis. At the same time, the stationarity test is performed based on the crop time series diagram. The stationary processing method for non-stationary sequence data is difference. Fig. 5 shows the peanut price time series and first-order difference time series diagram.

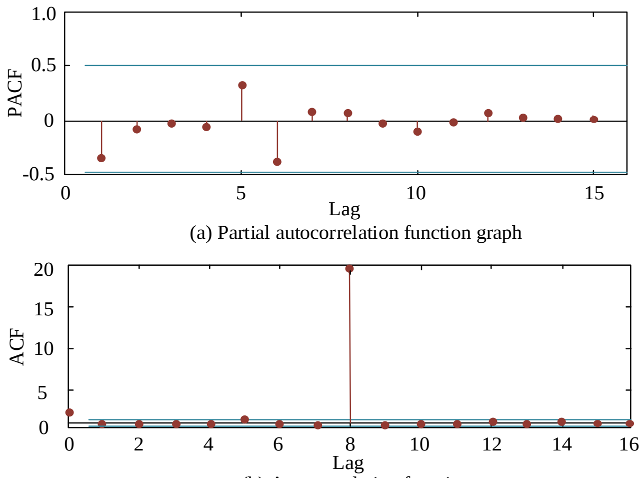
Partial autocorrelation function and autocorrelation function are used to determine d and q of the ARIMA model. As shown in Fig. 6(a) and (b), they are the partial autocorrelation function and the autocorrelation function respectively. Within the 95% confidence interval, the autocorrelation function shows the characteristic of eighth-order truncation, that is, it quickly approaches 0 after being greater than 8. From this, the value of ARIMA model q can be set to 8. In the analysis of the partial autocorrelation function, it has the characteristic of first-order truncation, and gradually tends to zero after being greater than 1. The value of calling the MATLAB built-in function for DW test is 1.819. The closer the value is to 2, it means that there is no first-order correlation in the residual value. Based on the above analysis, the ARIMA parameter value obtained is (1,1,8). Next, the residual test is performed, as shown in Fig. 7.



(a) Timing Diagram

(b) First order differential time series diagram

Fig. 5. Peanut price time series and first order difference time series.



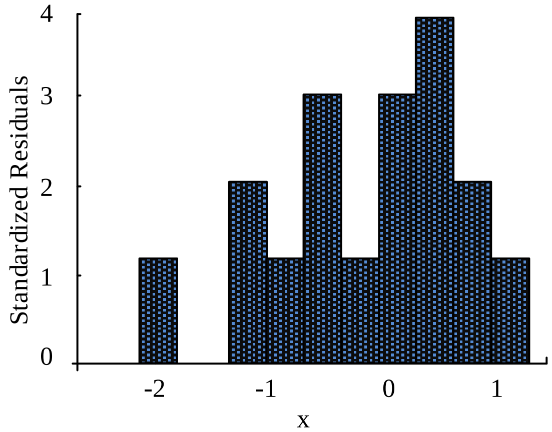
(a) Partial autocorrelation function graph

(b) Autocorrelation function

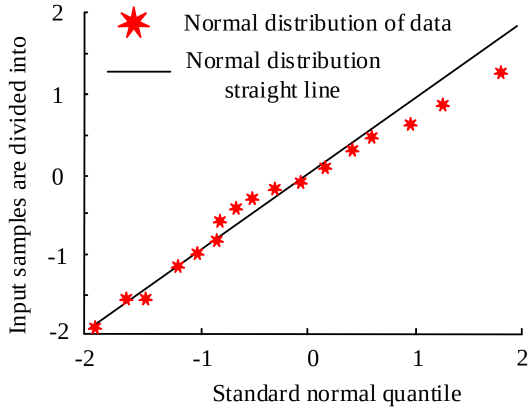
Fig. 6. Time series partial autocorrelation and autocorrelation function diagram.

Residual testing is an important step in the ARIMA model and is represented by a quantile-quantile plot (Quantile-Quantile, QQ). Fig. 7(a) and (b) are the residual test histogram and QQ plot. Judging from the residual test histogram, the model basically meets the requirements of normal distribution. Continue to test through the QQ plot. The value approaches a straight line, indicating that the ARIMA model meets the requirements of normal distribution. The ARIMA model is determined by the residuals as the number

ARIMA (1,1,8) model. Select the long-short-term neural network for comparison (Long Short Term Memory, LSTM), as shown in Fig. 8.



(a) Standardized Residuals



(b) Standard Normal Graph of Sample Data

Fig. 7. Residual test results of ARIMA model.

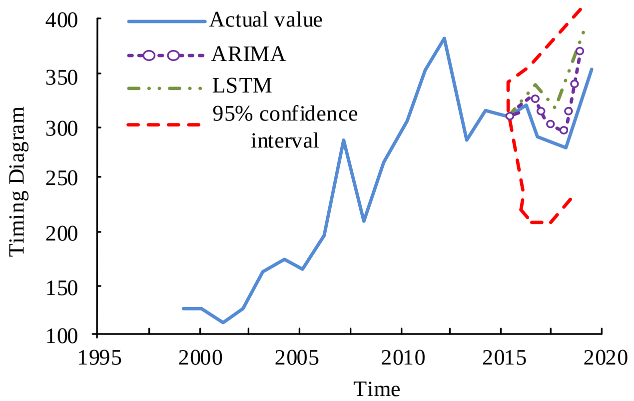


Fig. 8. Comparison of peanut price predictions using multiple price prediction models.

Judging from the data in Fig. 8, the peanut price from 2016 to 2020 is selected for prediction. The average prediction accuracy of LSTM is 81.65%, and the ARIMA model is 86.65%. Overall, the ARIMA model has higher prediction accuracy, but the model prediction accuracy is reduced for nonlinear data prediction. Therefore, the combined ARIMA -SVM model was used for further testing.

B. Crop Price Forecast Based on ARIMA-SVM Model

In the combined prediction model test, the predicted value of peanuts from 2016 to 2020 was obtained through ARIMA model linear analysis, and the difference (residual error) between the predicted value and the actual value fitting value was obtained through analysis. This part will use the RBF radial basis function in the SVM model to predict and analyze the data. The training set data is the nonlinear residuals in the ARIMA prediction values and 17 crop impact indicator factors. Set the training year range from 1995 to 2020. The model training residual values are shown in Table III.

TABLE III
ARIMA-SVM MODEL RESIDUAL PREDICTION VALUES

Time	ARIMA residual	SVM residual prediction
2016	-9.00622	-9.00865
2017	-12.00441	-13.99475
2018	-14.02852	-14.01893
2018	-8.08562	-8.61684
2019	9.44782	9.43783

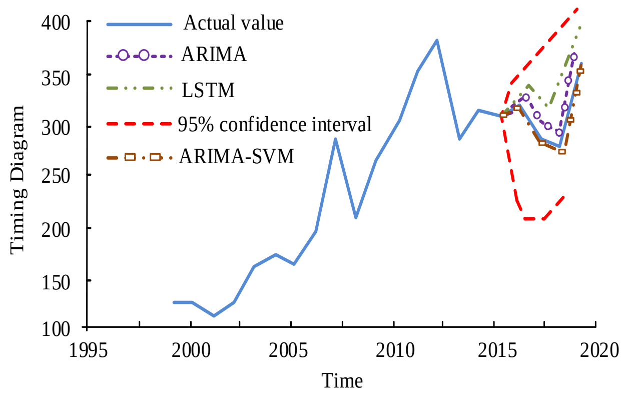


Fig. 9. Comparison results of peanut price predictions using multiple prediction models.

In Table III, the ARIMA-SVM prediction results consist of ARIMA prediction values and SVM prediction values. The specific prediction results are shown in Fig. 9.

As shown in Fig. 9, compared to the LSTM model and ARIMA model, the ARIMA -SVM model is closer to the actual blue value. Among them, the average prediction accuracy of the ARIMA -SVM model is 95.65%. Compared with the LSTM model and the ARIMA model, the prediction accuracy is improved by 23.65% and 12.65%. Table IV shows the comprehensive comparison results of the three models.

Mean Absolute Percentage Error (MAPE), JUNFANG, Mean Square Error (MSE) and Root Mean Squared Error (RMSE) were selected to evaluate the prediction effects of the three models. Judging from the data in Table V, the ARIMA-SVM prediction accuracy is above 0.945, while LSTM and ARIMA have the lowest prediction values in 2019, which are 0.803 and 0.864 respectively.

There are two innovations in the research. Firstly, in order to accurately analyze crop data, SVM and ARIMA are introduced to model time series data, ensuring the accuracy of price analysis. The second is to introduce residual testing to modify the price prediction model and improve the application effect of the technology through parameter optimization.

TABLE IV
COMPREHENSIVE COMPARISON RESULTS OF THREE PRICE PREDICTION MODELS

Method type	Time	Prediction accuracy	MAPE(%)	MSE	RMSE
LSTM	2016	0.856	564.54	0.1353	0.1551
	2017	0.835	656.45	0.1652	0.1567
	2018	0.803	1565.5	0.2157	0.1951
	2019	0.805	456.76	0.1753	0.1865
	2020	0.846	354.57	0.1553	0.1672
ARIMA	2016	0.875	356.45	0.1153	0.1565
	2017	0.895	286.64	0.1153	0.1452
	2018	0.864	358.56	0.1953	0.1769
	in 2019	0.896	253.54	0.1153	0.1452
	in 2020	0.915	125.57	0.1154	0.1069
ARIMA-SVM	2016	0.956	0.054	0.0153	0.0252
	in 2017	0.945	0.085	0.0153	0.0265
	2018	0.957	1.854	0.0993	0.0865
	in 2019	0.949	0.964	0.0153	0.0146
	2020	0.986	0.056	0.0153	0.0165

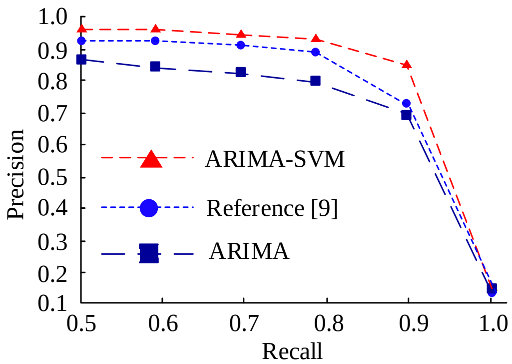
TABLE V
Comparison Results of Multi technical Performance

Different crops	Method type	Stability (%)	Average prediction accuracy
Corn	ARIMA	84.45	0.915
	ARIMA-SVM	96.56	0.965
	Reference [9]	92.54	0.926
Soybean	ARIMA	81.45	0.906
	ARIMA-SVM	95.86	0.959
	Reference [9]	88.64	0.926
Wheat	ARIMA	85.64	0.905
	ARIMA-SVM	89.45	0.967
	Reference [9]	96.45	0.915

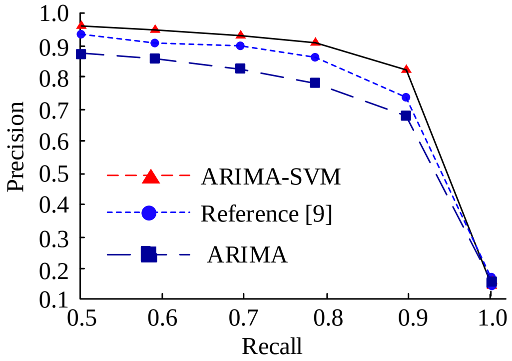
In the comparison of MAPE errors, the MAPE values of the ARIMA-SVM model are all in the range of 1.854%, with the smallest error. At the same time, the MSE values and RMSE values of the three prediction models were compared, and the ARIMA-SVM model errors were the lowest, indicating that the combined model has better application effects in agricultural crop price prediction. Finally, the study added literature [9] to compare the stability and accuracy of 10 rounds of prediction using techniques and research methods. Considering the average prediction performance of LSTM, it was not selected for testing. The results are shown in Table V.

Table V shows the comparison results of multi technology performance. According to the test results, the research model ARIMA-SVM has excellent stability in 10 rounds of testing, with stability above 95%, which is better than the literature [9] and ARIMA model. In addition, the ARIMA-SVM model also shows significantly better average prediction accuracy, indicating that the technology proposed in the study has better application effects. In addition, a precision and recall evaluation model is introduced, and Fig. 10 shows the results of the precision recall curve.

The accuracy recall training results are shown in Fig. 10, where Fig. 10 (a) shows the training results in the corn crop scenario. The ARIMA-SVM model has the best representation, with an accuracy of 95.6, while the technology proposed in reference [9] and ARIMA model are 0.923 and 0.882, respectively, and the recall rate of ARIMA-SVM model is also better than the other model.



(a) Corn



(b) Soybean

Fig. 10 Precision recall curve results

Fig. 10 (b) shows the training results in the soybean scenario, where the accuracy of the ARIMA-SVM model, and the technology proposed in reference [9], ARIMA model are

0.968, 0.942, and 0.893, respectively. The ARIMA-SVM model still has better training performance, indicating that the research model has better performance in predicting actual grain crop prices.

Finally, seven crops including corn, soybeans, wheat, onions, carrots, garlic, and rice were selected for future price prediction. The result is shown in Fig. 11.

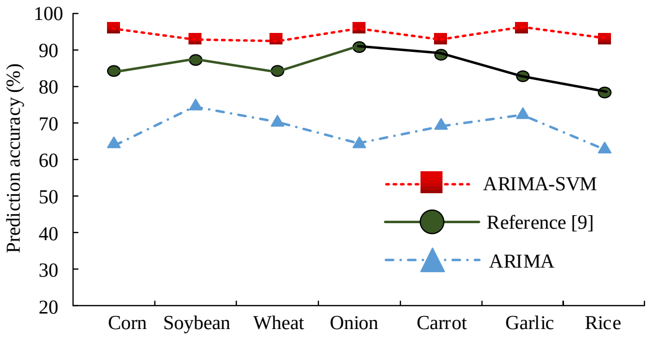


Fig. 11 Predicting the future prices of 7 crops.

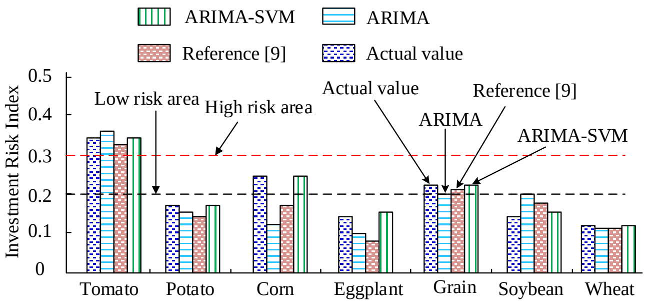


Fig. 12 Prediction of investment risks for different crops.

In Fig. 11, seven common crops were selected for price prediction, including corn, soybeans, wheat, onions, carrots, garlic, and rice. The ARIMA-SVM model had corresponding prediction accuracies of 98.2%, 91.2%, 91.8%, 97.8%, 91.5%, 97.9%, and 92.8%, respectively. According to the prediction results, the ARIMA-SVM model has a prediction accuracy of over 90%, demonstrating excellent performance. However, literature [9] showed relatively average performance, such as poor overall performance in predicting corn and rice, with accuracy rates of 84.5% and 78.2%, respectively. The ARIMA model with the worst performance has poor overall prediction accuracy, such as only 63.2% for rice prediction, which cannot meet the requirements. Finally, based on the data from China Agricultural and Rural Information Network from 2022 to 2023, a predictive analysis was conducted on the comprehensive agricultural prices to evaluate their investment risks. The results are shown in Fig. 12.

Fig. 12 shows the prediction results of investment risks for different crops. According to the test results, tomato planting has the highest investment risk, followed by corn and grains, mainly considering factors such as price and planting cost. Among them, the ARIMA-SVM model's risk prediction results are close to the actual values, with a risk prediction accuracy of 0.9725, followed by literature [9] at 0.8725, while the ARIMA model is only 0.8125. It can be seen that the ARIMA-SVM model has the best comprehensive performance and meets the requirements of agricultural development.

V. Discussion

In the field of crop price prediction, accurate prediction of crop prices is of great significance for agricultural production, market regulation, and policy formulation. Traditional prediction methods often rely on historical data and statistical models, but with the development of information technology, advanced technologies such as time series analysis and machine learning have provided new means for crop price prediction. The ARIMA model, as a classic time series forecasting method, has shown good performance in predicting crop prices. The study conducted a time-series analysis of peanut prices based on the ARIMA-LSTM model and conducted simulation experiments using the Matlab platform. At the same time, the agricultural crop information management system developed by combining JavaScript and KO framework has achieved effective processing and presentation of crop price data.

According to the experimental test results, the composite model used in the study showed high accuracy in predicting peanut prices. Through the analysis of the time series chart of peanut prices, it was found that peanut prices showed significant fluctuations after 2005, especially during the period of significant increase from 2010 to 2014, and then fell back in 2015. Through ADF and KPSS tests, it was confirmed that the first-order differenced time-series data meets the stationarity requirements, and the parameters of the ARIM model were determined to be (1,1,8). The residual test results indicate that the ARIMA model meets the requirements of normal distribution, further verifying the reliability of the model.

To further explore the performance of the ARIMA model and improve it, the ARIMA-LSTM model is proposed. And compared with similar technologies, the ARIMA-LSTM model achieved an average accuracy of 95.65% in peanut price prediction, which is superior to LSTM and ARIMA models. The main reason is that it has excellent processing ability for time series and the flexibility of ARIMA model for data processing, which enhances the processing effect on nonlinear data. For example, in the subsequent performance analysis of the model, the technique proposed in reference [9] was introduced for comparison, and the ARIMA-LSTM model performed well in recall and accuracy. In addition, various crop data were selected for experiments in the study, including corn, soybeans, tomatoes, etc. The research techniques have good data analysis capabilities, which can identify crop prices well and provide risk opinions.

According to the results, the ARIMA-SVM combination model performs well in crop price prediction, outperforming individual ARIMA models, LSTM models, and other similar models not mentioned. This combination model can effectively handle linear and nonlinear data of crop prices, improve prediction accuracy, and accurately evaluate crop investment risks. However, this method is mainly based on market data and does not fully consider the impact of natural phenomena on crop prices. In the future, it is necessary to further collect and analyze the impact of natural phenomena such as climate change and disaster events on crop prices. In addition, more socio-economic factors such as policy changes, international trade conditions, etc. are introduced to improve the comprehensiveness and predictive accuracy of the model.

VI. Conclusion

Agricultural crop prices are affected by many factors such as the environment and policies, and have always been difficult to manage. In order to accurately predict and manage crop prices, first collect and analyze crop price information to obtain crop price influencing factors. Then, considering that the influencing factors have nonlinear characteristics, a crop price prediction model is built based on SVM. At the same time, considering that the crop data over the years is linear data, the ARIMA model is used to analyze the linear data of the crops, and the effective prediction of crop prices is achieved through the combination of the two models. In the prediction of the ARIMA model, the time series stability of the data is first tested, and the sequence meets the stability requirements after differential processing. The prediction performance of the ARIMA model was tested. The average prediction accuracy of peanut prices from 2016 to 2020 was 86.65%, which was better than the LSTM model, but the accuracy was average. In the prediction of the combined model, the SVM model is introduced to process nonlinear data. The prediction results show that the average prediction accuracy of the ARIMA-SVM model is 95.65%. Compared with the LSTM model and the ARIMA model, the prediction accuracy is improved by 23.65% and 12.65%. At the same time, in the comparison of MAPE errors, the MAPE values of the ARIMA-SVM model are all in the range of 1.854%, with the smallest error. In addition, in order to compare the comprehensive performance of the model, the recall rate and accuracy rate were compared. According to the results, the accuracy rate of arima-svm model in soybean training was 0.968, and the comprehensive performance was the best. In addition, the model is used to predict the investment risk of a variety of crops, including corn, Cereals, soybeans and tomatoes. According to the test results, the arima-svm model is close to the actual risk situation, and the prediction accuracy is 0.9725, which is superior to similar models, and performs well. In summary, the combination model ARIMA-SVM has shown excellent overall performance in agricultural price forecasting, with more accurate and stable price predictions, as well as the ability to accurately assess crop investment risks. The overall effect is outstanding and meets the needs of agricultural development. However, this research technique mainly considers crop prices based on market data, without fully considering natural phenomena. In the future, further consideration is needed to improve the practical application effectiveness of this technology.

From this, it can be seen that the combination model has good application performance in crop price prediction, with good performance compared to similar models, and high accuracy in agricultural price prediction, meeting the requirements of agricultural crop price management. However, this research technology mainly considers crop prices based on market data and does not fully consider natural phenomena. In the future, further consideration is needed to enhance the practical application effect of the technology.

References

[1] R. Akhter, S. A. Sofi, “Precision agriculture using IoT data analytics and machine learning”, *Journal of King Saud University: Computer and Information Sciences*, vol. 34, no. 8, pp. 5602-5618, 2022.

[2] G. Sambasivam, G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks", *Egyptian Informatics Journal*, vol. 22, no. 1, pp. 27-34, 2021.

[3] B. J. Chelliah, T. P. Latchoumi, A. Senthilselvi, “Analysis of demand forecasting of agriculture using machine learning algorithm. Environment", *Development and Sustainability*, vol. 26, no. 1, pp. 1731-1747, 2024.

[4] R. Birner, T. Daum, C. Pray, "Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges", *Appl Econ Perspect P*, vol. 43, no. 4, pp. 1260-1285, 2021.

[5] A. Sharifi, "Yield prediction with machine learning algorithms and satellite images", *J Sci Food Agr*, vol. 101, no. 3, pp. 891-896, 2021.

[6] H. Storm, K. Baylis, T. Heckelei, "Machine learning in agricultural and applied economics", *Eur Rev Agric Econ*, vol. 47, no. 3, pp. 849-892, 2020.

[7] N. N. Misra, Y. Dixit, A. Al-Mallahi, "IoT, big data, and artificial intelligence in agriculture and food industry", *IEEE Internet Things*, vol. 9, no. 9, pp. 6305-6324, 2020.

[8] S. K. Purohit, S. Panigrahi, P. K. Sethy, S. K. Behera, "Time series forecasting of price of agricultural products using hybrid methods", *Appl Artif Intell*, vol. 35, no. 15, pp. 1388-1406, 2021.

[9] L. Wang, J. Feng, X. Sui, Q. Chu, W. Mu, "Agricultural product price forecasting methods: research advances and trend", *British Food Journal*, vol. 122, no. 7, pp. 2121-2138, 2020.

[10] P. Das, G. K. Jha, A. Lama, R. Parsad, M. Dwijesh, "Empirical mode decomposition based support vector regression for agricultural price forecasting", *Indian Journal of Extension Education*, vol. 56, no. 2, pp. 7-12, 2020.

[11] C. Wang, B. Liu, L. Liu, Y. J. Zhu, J. L. Hou, X. Li, "A review of deep learning used in the hyperspectral image analysis for agriculture", *Artif Intell Rev*, vol. 54, no. 7, pp. 5205-5253, 2021.

[12] P. Trivedi, C. Mattupalli, K. Eversole, "Enabling sustainable agriculture through understanding and enhancement of microbiomes", *New Phytol*, vol. 230, no. 6, pp. 2129-2147, 2021.

[13] S. S. Bora, A. L. Katchova, T. H. Kuethe, "The accuracy and informativeness of agricultural baselines", *Am J Agr Econ*, vol. 105, no. 4, pp. 1116-1148, 2023.

[14] Y. Liu, X. Ma, L. Shu, "From Industry 4.0 to Agriculture 4.0: Current status, enabling technologies, and research challenges", *IEEE T Ind Inform*, vol. 17, no. 6, pp. 4322-4334, 2020.

[15] G. A. Meehl, J. H. Richter, H. Teng, A. Capotondi, K. Cobb, "Initialized earth system prediction from subseasonal to decadal timescales", *Nat Rev Earth Env*, vol. 2, no. 5, pp. 340-357, 2021.

[16] D. J. Pannell, R. Claassen, "The roles of adoption and behavior change in agricultural policy", *Appl Econ Perspect P*, vol. 42, no. 1, pp. 31-41, 2020.

[17] T. M. Maaz, T. B. Sapkota, A. J. Eagle, "Meta-analysis of yield and nitrous oxide outcomes for nitrogen management in agriculture", Global Change Biol, vol. 27, no. 11, pp. 2343-2360, 2021.

[18] Z. Chen, "Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm", *Journal of Computational and Cognitive Engineering*, vol. 1, no. 3, pp. 103-108, 2022.

[19] R. Hanif, S. Mustafa, S. Iqbal, S. Piracha, "A study of time series forecasting enrollments using fuzzy interval partitioning method", *Journal of Computational and Cognitive Engineering*, vol. 2, no. 2, pp. 143-149, 2023.

[20] Khadija SLIMANI, Yassine RUICHEK, and Rochdi MESSOUSSI, "Compound Facial Emotional Expression Recognition using CNN Deep Features," Engineering Letters, vol. 30, no.4, pp1402-1416, 2022.

[21] S. Cho, S. Lee, S. I. Ahn, "Design and engineering of organ-on-a-chip", *Biomedical Engineering Letters*, vol. 13, no. 2, pp. 97-109, 2023.

[22] Kanita Petcharat, "Designing the Performance of EWMA Control Chart for Seasonal Moving Average Process with Exogenous Variables," IAENG International Journal of Applied Mathematics, vol. 53, no.2, pp757-765, 2023.

[23] C. Chananet, and S. Phanyaem, "Improving CUSUM Control Chart for Monitoring a Change in Processes Based on Seasonal ARX Model," IAENG International Journal of Applied Mathematics, vol. 52, no.3, pp589-596, 2022.