CIT. Journal of Computing and Information Technology, Vol. 29, No. 4, December 2021, 251–263
doi: 10.20532/cit.2021.1005390

Application of Data Visualization and Big Data Analysis in Intelligent Agriculture

Wei Liu

School of Information Engineering, Shaanxi Xueqian Normal University, Xi'an, China

Intelligent agriculture can renovate agricultural production and management, making agricultural production truly scientific and efficient. The existing data mining technology for agricultural information is powerful and professional. But the technology is not well adapted for intelligent agriculture. Therefore, this paper introduces data visualization and big data analysis into the application scenarios of intelligent agriculture. Firstly, an intelligent agriculture data visualization system was established, and the RadViz data visualization method was detailed for intelligent agriculture. Moreover, the intelligent agriculture data were processed using dimensionality reduction through principal component analysis (PCA) and further optimized through k-means clustering (KMC). Finally, the crop yield was predicted using the multiple regression algorithm and the residual principal component regression algorithm. The crop yield prediction model was proved effective through experiments.

ACM CCS (2012) Classification: Software and its engineering → Software organization and properties → Extra-functional properties → Software fault tolerance

Keywords: intelligent agriculture, data visualization, big data analysis

1. Introduction

As a large agricultural country, China far exceeds the other countries in the total population involved in agricultural production and the coverage of agricultural production equipment. However, the country faces obvious defects in productivity, crop yield, and quality of agricultural products [1–4].

Intelligent agriculture is a novel agricultural development model, which effectively improves the backward production techniques and management methods, and truly realizes scientific and efficient agricultural production. With these advantages, intelligent agriculture has received extensive attention from scholars at home and abroad [5–9].

The development of information technology provides a strong support to the wide application of intelligent agriculture [10–12]. Data visualization and big data analysis need to be applied scientifically to process, display, and mine the massive complex production data gathered by intelligent agriculture.

To solve the problems of current systems in information perception and communication security [13–16], Suciu *et al.* [17] combined wired and wireless networks into a Modbus-based communication protocol for network data, and introduced the heartbeat detection mechanism into the network communication between the client and the server, which effectively reduces the misjudgment rate of the system. Putri *et al.* [18] improved the functions of the current intelligent agriculture production system in soil fertility analysis and crop yield prediction, and effectively improved the efficiency, classification speed, and accuracy of algorithm in handling continuous attributes, with the aid of K-means clustering and Pearson correlation coefficient.

Intelligent agriculture, which integrates information technology, sensing technology, and wireless communication technology, can

achieve smart sensing, transmission, and analysis of the information in each link of agricultural planting. The numerous links, various plants, and diverse sensors involved in agricultural planting undermine the efficiency and reliability of data collection and storage [19–23].

For reliable storage and efficient use of massive agricultural data, Saranya *et al.* [24] developed an intelligent agriculture monitoring system with a distributed framework, and realized the identification and web interface alarm of abnormal points in agricultural dataset by appying k-means clustering, which is superior in data processing. Nagaraja *et al.* [25] constructed a web-based intelligent agriculture platform with functions such as expert guidance, growth factor state monitoring, and growth condition control. In addition, they provided the design flows of the browser/server (B/S)-based client and remote-control module (including water pump control and sunshade control) and carried out big data analysis to analyze and integrate user data with agricultural data and to extract the values of deep data. Chen *et al.* [26] introduced cloud storage and cloud computing into the collection and analysis of intelligent agriculture data, compiled the cloud, host, and client-end programs for intelligent agriculture based on development tools like Android Studio and XAMPP.They also achieved real-time acquisition and storage of temperature, humidity, and illuminance of greenhouses. Fully considering the features of agricultural soil data, Khatri-Chhetri *et al.* [27] explored the WebGIS visualized analysis under the big data of agricultural soil and differentiated the visual features of agricultural soil under natural and geographic spatial attributes. They proposed an integrated preprocessing method for agricultural soil data, which covers data cleaning, compaction, regularization, and reduction, and provided an ideal solution to problems like unsatisfactory data display, slow rendering speed, and small visualization space. Mazzetto *et al.* [28] collected real-time planting data with web crawlers, built up a Django Web Framework according to the features of different datasets and the needs of data visualization, displayed the resources of real-time planting data, and supported the visual display of planting data and dynamic interaction of data charts.

With information technology at the core, intelligent agriculture has reshaped the traditional agricultural values. It combines modern intelligent models, highlights detailed management, and controls agricultural production and agricultural products, creating a truly smart production model.

By summing up the existing literature, several problems of China's intelligent agriculture were exposed, such as small scale, and limited coverage of information technology. With the rapid development of information technology, the traditional agricultural information websites can no longer meet the changing user demand, with their basic functions like browsing and information retrieval. Besides, the current agricultural data mining technology does not adapt to the massive amount of data generated by intelligent agriculture, because its powerful functions only apply to specific professional fields.

To solve the defects of data mining technology for agricultural information, this paper explores the data visualization and big data analysis technologies for intelligent agriculture. The main contents of this paper cover the following aspects.

1. Building a visualization system for intelligent agriculture data, and providing the workflow of the system;
2. Detailing the RadViz data visualization method for intelligent agriculture, and explaining the procedure of dimensionality reduction through principal component analysis (PCA) and optimization through k-means clustering (KMC) for intelligent agriculture data;
3. Predicting crop yield with multiple regression algorithm and residual principal component regression algorithm;
4. Providing a visual display of clustered intelligent agriculture data, and verifying the effectiveness of the crop yield prediction model.

2. Construction of Intelligent Agriculture Data Visualization System

As shown in Figure 1, the data visualization system for intelligent agriculture is described by the following process: sensing of crop

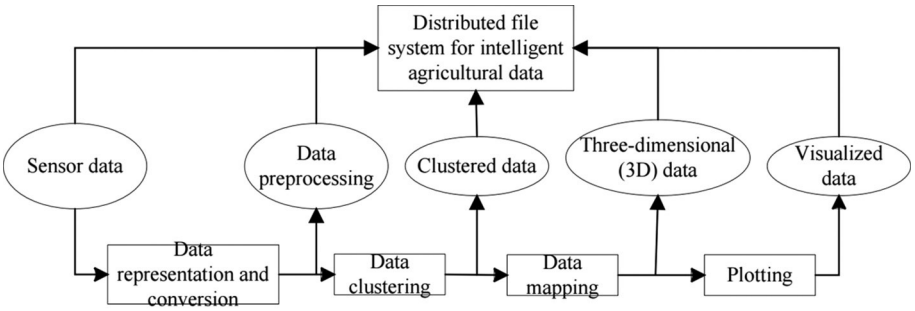


Figure 1. Flow chart of data visualization system for intelligent agriculture.

growth data, preprocessing data, clustering and optimizing data, visual mapping of data, system construction, and system realization.

The Hadoop distributed file system was adopted to store the big data collected and processed by intelligent agriculture. Through the effective formation and orderly distribution of data blocks, the processing functions of intelligent agriculture big data were optimized, including data connection, integration, and storage. The basic framework of the system (Figure 2) consists of a single Master node to management the file system, data nodes to store files in the form of data blocks, and secondary naming nodes for data backup.

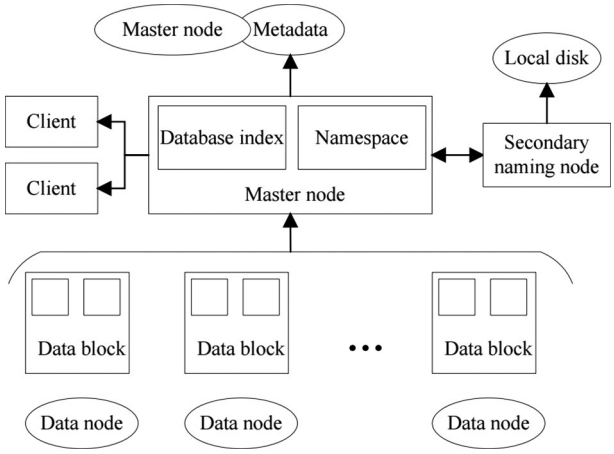


Figure 2. Hadoop distributed file system for intelligent agriculture.

Figure 3 shows the B/S-based visualization structure of intelligent agriculture data. There are four layers of the structure, namely, the storage layer that stores original files and test metadata, the calculation layer for mapping and reduction operations, the application layer for data calling and interaction with remote clients, and the user layer for data uploading and management. The functions on the four layers are closely linked with each other, and operate collaboratively.

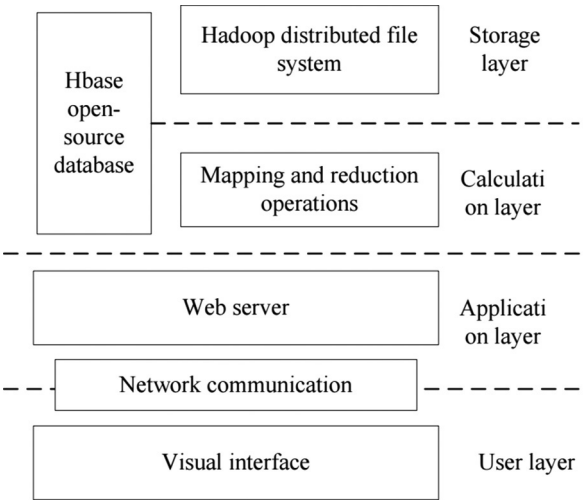


Figure 3. B/S-based visualization structure of intelligent agriculture data.

3. RadViz Data Visualization for Intelligent Agriculture

Figure 4 illustrates the principle of anchor point mapping on RadViz plane. The connection of each anchor point AP_i* is represented in the RadViz plane, such as to map the point set [b_1, b_2, ..., b_m] in M-dimensional space to a point T*. The equilibrium formula of point T* can be given by:

$$\sum_{i=1}^{m}\left(AP_i^* \cdot T^*\right)b_i = 0. \tag{1}$$

Point T^* can be calculated by:

$$T^* = \frac{\sum_{i=1}^{m} AP_i^* \cdot b_i}{\sum_{i=1}^{m} b_i}. \quad (2)$$

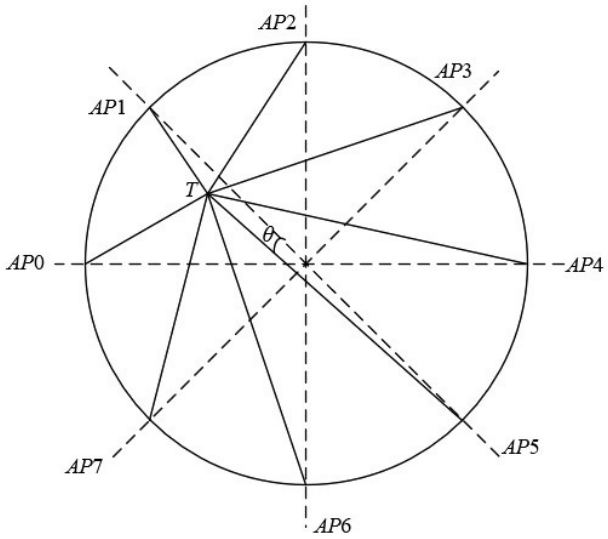


Figure 4. Principle of anchor point mapping on RadViz plane.

Let θ_i be the angle between anchor points on Radviz plane circle with T^* as the vertex, b_i be the dimension of AP_i^*, and m be the total number of dimensions. Then, the final coordinates $[a_T, a_T]$ of T^* after conversion can be obtained as follows:

$$a_T = \frac{\sum_{i=1}^{m} b_i \cos(\theta_i)}{\sum_{i=1}^{m} b_i}, \; b_T = \frac{\sum_{i=1}^{m} b_i \sin(\theta_i)}{\sum_{i=1}^{m} b_i}. \quad (3)$$

From formulas (2) and (3), it can be inferred that the calculations of any point mapped in RadViz plane and its coordinates are linear. When the high-dimensional data have different degrees of data similarity, the visualization can be realized by analyzing the law of data classification. This paper improves the traditional RadViz radial coordinate visualization technology for unclassified data: the PCA was adopted to reduce the dimensionality of the high-dimensional data on the growth and environment of crops in each growth stage, which were collected by intelligent agriculture, before visually displaying the relevant data. Figure 5 explains the principle of dimensionality reduction through PCA.

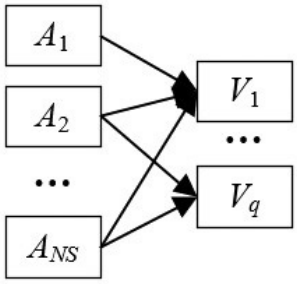


Figure 5. Principle of dimensionality reduction through PCA.

Let a_i' be the mean of NS intelligent agriculture data samples in the i-th row, and A_i be the growth or environment dataset collected by ND samples. Then, a sample matrix A_{ji} of NS rows and ND columns can be established, where the variance of the NS samples in the i-th row is $P^2{}_i$. Then, the matrix can be transformed into a standard matrix E:

$$E_{ji} = \frac{a_{ji} - a_i'}{P_i}, \quad j = 1, 2, ..., N_D; \; i = 1, 2, ..., N_S. \quad (4)$$

The values of a_i' and $P^2{}_i$ can be respectively calculated by:

$$a_i' = \frac{1}{N_S} \sum_{j=1}^{N_S} A_{ji}, \quad (5)$$

$$P_i^2 = \frac{\sum_{j=1}^{N_S} \left(a_{ji} - a_i'\right)^2}{N_S - 1}. \quad (6)$$

The correlation coefficient matrix H of E can be expressed as:

$$H = \left[h_{ji}\right]_{N_D} = \frac{E^T E}{N_S - 1} \quad (7)$$

where, h_{ji} is an element in matrix H:

$$h_{ji} = \frac{\sum e_{lj} - e_{li}}{N_S - 1}, \; j, i = 1, 2, ..., N_S. \quad (8)$$

The variance explained by each principal component can be calculated by:

$$\frac{\mu_i}{\sum_{l=1}^{N_D}\mu_l},\ i=1,\ 2,\ ...,N_D. \tag{9}$$

The cumulative variance explained can be calculated by:

$$\frac{\sum_{l=1}^{i}\mu_l}{\sum_{l=1}^{N_D}\mu_l},\ i=1,\ 2,\ ...,N_D. \tag{10}$$

The N_D eigenvalues and eigenvectors can be computed by the characteristic equation of H and transformed into the corresponding principal components $V_{ji} = e_i^T Z_i$, $i = 1, 2, ..., N_S$. Figure 6 summarizes the workflow of PCA-based dimensionality reduction.

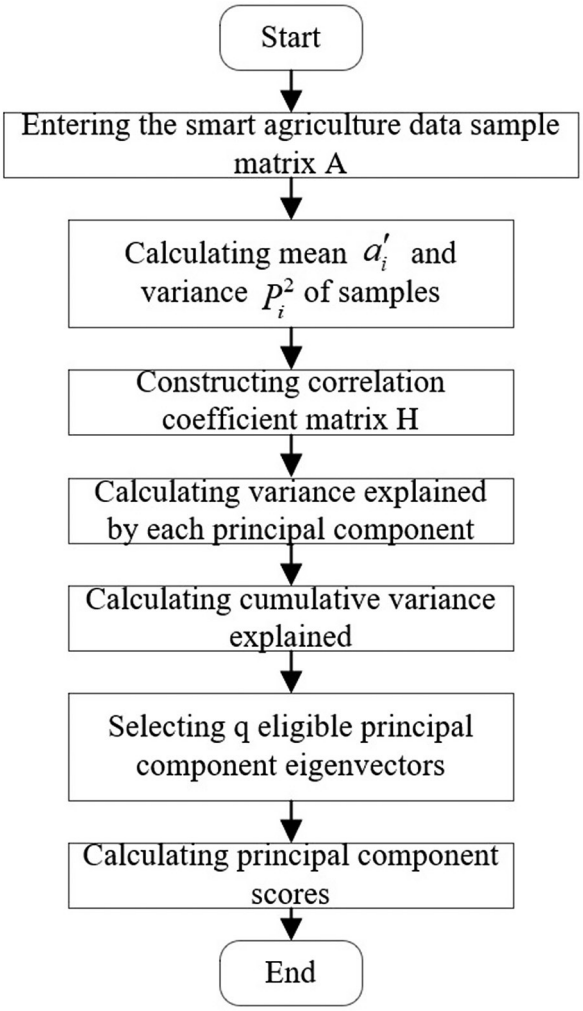


Figure 6. Flow chart of PCA-based dimensionality reduction.

To preserve the original data features in dimensionality reduction, the $q(q < N_D)$ eligible principal component eigenvectors R_i, $i = 1, 2, ..., q$ can be selected based on the results of formula (10), and the principal component scores can be calculated as follows:

$$\begin{cases} V_1 = R_{11}a_1 + R_{12}a_1 + ... + R_{1N_D}a_{N_D} \\ V_2 = R_{21}a_1 + R_{22}a_2 + ... + R_{2N_D}a_{N_D} \\ ... \\ V_q = R_{q1}a_1 + R_{q2}a_1 + ... + R_{qN_D}a_{N_D} \end{cases}. \tag{11}$$

After the PCA-based dimensionality reduction step, KMC is used to optimize the intelligent agriculture data. The purpose is to improve the perception accuracy of visualization technology for the characteristic patterns and association rules between the collected intelligent agriculture data after dimensionality reduction.

During two-dimensional (2D) clustering, the difference between two points (a_{11}, a_{12}) and (a_{21}, a_{22}) can be characterized by Euclidean distance:

$$DIS = \sqrt{\left(a_{21}-a_{11}\right)^2 + \left(a_{22}-a_{12}\right)^2}. \tag{12}$$

In a multi-dimensional space, the Euclidean distance can be calculated by:

$$DIS_{ji} = \sqrt{\left(a_{j1}-a_{j1}\right)^2 + \left(a_{j2}-a_{i2}\right)^2 + ... + \left(a_{jq}-a_{iq}\right)^2} = \sqrt{\sum_{k=1}^{q}\left(a_{jk}-a_{ik}\right)^2}. \tag{13}$$

Let (CC_1, CC_2, ..., CC_k) be the classes obtained by KMC, and γ_i be the mean vector of class CC_j. Then, the clustering objective function of the minimum squared error (MSE) σ can be expressed as:

$$\min \sigma = \sum_{j=1}^{k}\sum_{a\in CC_j}\left\|a-\gamma_j\right\|_2^2. \tag{14}$$

The mean vector γ_i, which can be viewed as cluster head, can be obtained by:

$$\gamma_j = \frac{1}{\left|CC_j\right|}\sum_{a\in CC_j} a. \tag{15}$$

Figure 7 provides the flow of KMC for visualization of intelligent agriculture data.

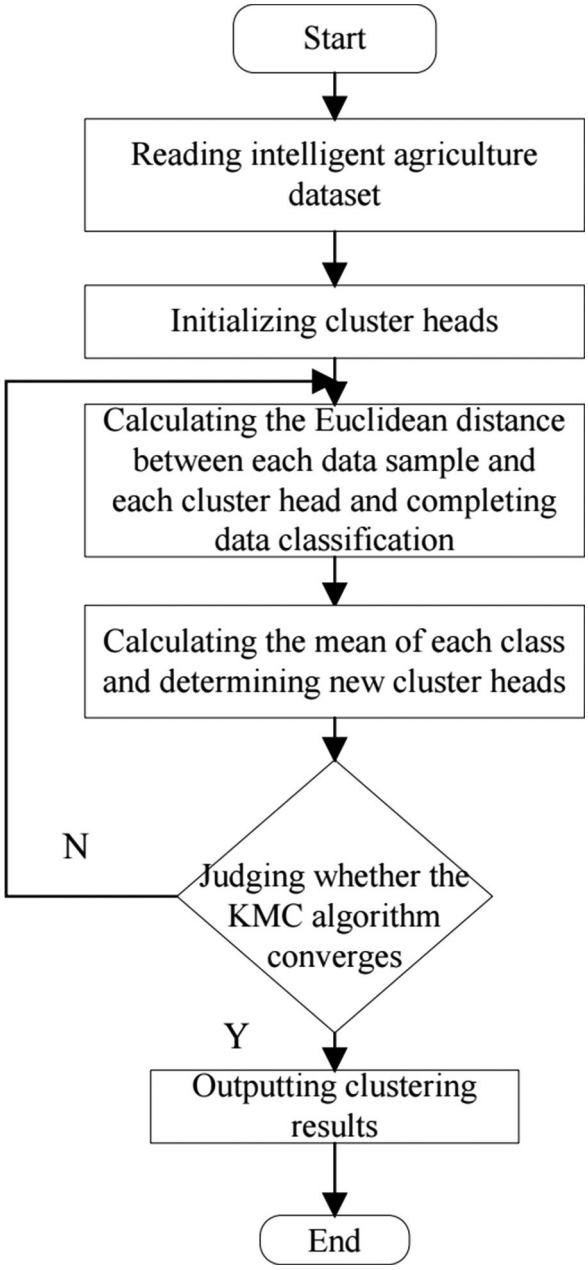


Figure 7. Flow of KMC-based data optimization.

The KMC involves the following steps:

Step 1. Read the intelligent agriculture data set, and randomly choose k samples as the initial cluster heads.

Step 2. Calculate the Euclidean distance from each data sample to each cluster head, and complete data classification.

Step 3. Calculate the mean of each class and determine new cluster heads.

Step 4. Judge whether the KMC algorithm converges. If yes, output the clustering results; otherwise, add 1 to the number of iterations and return to Step 2.

4. Crop Yield Prediction Based on Big Data Analysis

The visualized data of intelligent agriculture usually contain historical data and predicted data. An important aspect of the predicted data is prediction of yield. The description of our crop yield prediction method is given in the continuation of this section. The intelligent agriculture data are complex in form, reflecting the growth and environment states of crops in different growth stages. Crop yield is generally affected by controllable human factors like fertilization and irrigation, environmental factors like temperature and rainfall, and soil factors like various organic matters and elements. Let B be the crop yield, and NIV be the number of crop growth parameters. Then, a multiple regression equation can be constructed:

$$b = \omega_0 + \omega_1 \cdot a_1 + \omega_2 \cdot a_2 + ... + \omega_{N_{IV}} \cdot a_{N_{IV}} + \tau. \tag{16}$$

Suppose there are N_{POV} groups of data samples. Let ω_0, ω_1, ..., ω_{NPOV} be the $N_{IV}+1$ parameters to be measured; τ_j be the degree of influence of factors on crop yield B_j in the j-th acquisition process. Then, the multiple regression equation can be transformed into an equation set:

$$\begin{cases} b_1 = \omega_0 + \omega_1 \cdot a_{11} + \omega_2 \cdot a_{12} + \\ ... + \omega_{N_{POV}} \cdot a_{1N_{POV}} + \tau_1 \\ b_2 = \omega_0 + \omega_1 \cdot a_{21} + \omega_2 \cdot a_{22} + \\ ... + \omega_{N_{POV}} \cdot a_{2N_{POV}} + \tau_2 \\ ... \\ b_{N_{IV}} = \omega_0 + \omega_1 \cdot a_{N_{IV}1} + \omega_2 \cdot a_{N_{IV}2} + \\ ... + \omega_{N_{POV}} \cdot a_{N_{IV}N_{POV}} + \tau_{N_{IV}} \end{cases}. \tag{17}$$

Based on the existing intelligent agriculture data, the parameters to be measured $\omega_0 \sim \omega_{NPOV}$ were estimated by least squares (LS) method to obtain the desired regression model. The LS method aims to minimize the residual square between the actual and predicted values of samples, that is, solve the $\omega_0 \sim \omega_{NPOV}$ when formula (18) is in the minimum state:

$$W\left(\omega_0, \omega_1, ..., \omega_{N_{IV}}\right) = \sum_{j=1}^{N_{POV}} \left(B_j - \dot{B}_j\right)^2 = \sum_{j=1}^{N_{POV}} \left(B_j - \omega_0 - \omega_1 a_{j1} - \omega_2 a_{j2} - ... - \omega_{N_{IV}} a_{jN_{IV}}\right)^2. \quad (18)$$

More to the point, it is necessary to find the partial derivative of $W(\omega_0, \omega_1, ..., \omega_{NPOV})$ relative to every parameter to be measured $\omega_0 \sim \omega_{NPOV}$, and set it to zero:

$$\frac{\partial\left(\omega_0, \omega_1, ..., \omega_{N_{IV}}\right)}{\partial \omega_l} = -2 \sum_{j=1}^{N_{POV}} \left(b_j - \sum_{i=0}^{\omega_{N_{IV}}} \omega_1 a_{ji}\right) = 0, \left(a_{jl} = 0, l = 0,1,..., \omega_{N_{POV}}\right). \quad (19)$$

The crop growth parameters can be effectively fitted by solving the formula (19). Multiple regression aims to clarify the functional relationship between dependent and independent variables. Then, the dependent variable can be predicted based on the known independent variables. Since the variables are often uncertain, the functional model may not pass through every data point. To judge the model quality, it is important to statistically test the model.

4.1. Crop Yield Modeling Based on Multiple Regression Algorithm

It is impossible to obtain the future values of parameters like temperature, rainfall, and organic matters/elements in the soil, for these factors vary with time and environment. Before predicting crop yield, the future values of the said parameters must be forecasted. Here, the forecast is realized by exponential smoothing. The value of initial smoothing point VP_2 was taken as the actual data value b_1. Then, the value of smoothing point VP_t at time t can be calculated by:

$$VP_t = \xi b_t + \left(1 - \xi\right) VP_{t-1}, 0 < \xi \le 1 \quad (20)$$

where, ξ is the smoothing coefficient. It can be seen from the above formula that the ξ value is negatively correlated with the stability of data changes. Then, formula (20) can be expanded into:

$$VP_t = \xi \sum_{i=0}^{t-1} \left(1 - \xi\right)^i b_{t-i} + \left(1 - \xi\right)^t VP_0. \quad (21)$$

The ξ value ranges in [0, 1]. When t approximates + ∞, $(1 - \xi)^2$ approaches zero. Hence, formula (21) can be rewritten as:

$$VP_t = \xi \sum_{i=0}^{\infty} \left(1 - \xi\right)^t b_{t-i}. \quad (22)$$

Formula (22) shows that the value of VP_t can be obtained by weighted average operation on b_t, b_{t-1}, ..., b_{t-i}, using the coefficients ξ, $\xi(1 - \xi)$, $\xi(1 - \xi)^2$, ... respectively. Since the weight coefficient attenuates in geometric progressions, the data samples with large coefficients approach each other, while those with small coefficient move away from each other. The sum of weight coefficients must be equal to 1:

$$\xi \sum_{i=0}^{\infty} \left(1 - \xi\right)^i = 1. \quad (23)$$

The crop yield model can be described by:

$$\dot{b}_{t+1} = VP_t = \xi b_t + \left(1 - \xi\right) \dot{b}_t. \quad (24)$$

4.2. Residual Principal Component Regression Algorithm and Model

Due to the sheer number of influencing factors, the regression model is not very accurate in predicting crop yield, and the prediction results are susceptible to abnormal values. To solve these two problems, this paper presents a residual principal component regression algorithm. Firstly, a multiple regression algorithm was applied to build a model for prediction, followed by solving and predicting the residual. Then, the PCA was performed to reduce the dimensions of the original attributes, and replace the original influencing factors with a few principal components. In addition, the residual was taken

as a principal component to build another regression model for prediction. The first step is to compute the prediction residual error $error_j$, which characterizes the gap between actual value B_j and fitted value $\dot{B}_j$:

$$error_j = B_j - \dot{B}_j. \quad (25)$$

Suppose the crop yield is described by N_C variables $A_1, A_2, ..., A_{NC}$. Then, the N_C-dimensional random vector formed by these variables can be described as $A = (A_1, A_2, ..., A_{NC})^t$. Let γ and CM be the mean and covariance matrix of A, respectively; γ_l be the expectation of the l-th element in CM. Based on the covariance matrix, the linear combination can be transformed linearly on A, in view of the original variables:

$$\begin{cases} E_1 = \gamma_{11}A_1 + \gamma_{12}A_2 + ... + \gamma_{1N_C}A_{N_C} \\ E_1 = \gamma_{21}A_1 + \gamma_{22}A_2 + ... + \gamma_{2N_C}A_{N_C} \\ \quad ... \\ E_1 = \gamma_{N_C1}A_1 + \gamma_{N_C2}A_2 + ... + \gamma_{N_CN_C}A_{N_C} \end{cases}. \quad (26)$$

Next is to standardize the original data matrix of intelligent agriculture, and build up the corresponding covariance matrix G. The elements $G_{ji}(j, i = 1, 2, ..., N_C)$ in G are composed of the correlation coefficients between original variables A_j and A_i:

$$G_{ji} = \frac{\sum_{l=1}^{N_C}\left(A_{li} - A_j\right)\left(A_{li} - A_i\right)}{\sqrt{\sum_{l=1}^{N_C}\left(A_{li} - A_j\right)^2\left(A_{li} - A_i\right)^2}}. \quad (27)$$

By solving the covariance matrix G, the eigenvalues can be sorted in descending order as $\mu_1 > \mu_2 > ... > \mu_j > 0$ and the eigenvectors $error_j(j = 1, 2, ..., N_C)$ can be obtained. Then, the i-th component of prediction residual $error_j$ must satisfy:

$$\left\|error_j\right\| = \sum_{i=1}^{N_C} error_{ji}^2 = 1. \quad (28)$$

The variance explained by principal component E_j can be calculated by:

$$\eta_j = \mu_i / \sum_{i=1}^{N_C} \mu_i. \quad (29)$$

The cumulative variance can be calculated by:

$$\eta = \sum_{i=1}^{N_{IV}} \mu_i / \sum_{i=1}^{N_C} \mu_i. \quad (30)$$

Then, the factor loading representing the meaning of principal component can be defined as the correlation coefficient $G(E_j, A_j)$ between principal component E_j and original variable A_j:

$$k_{ji} = \sqrt{\mu_j}\, error_{ji}\left(j, i = 1, 2, ..., N_C\right). \quad (31)$$

Each principal component vector can be obtained by dividing each column in the factor loading matrix with the square root of the corresponding eigenvalue. Taking the residual as the last principal component, the desired residual principal component model can be obtained by modeling all principal components again.

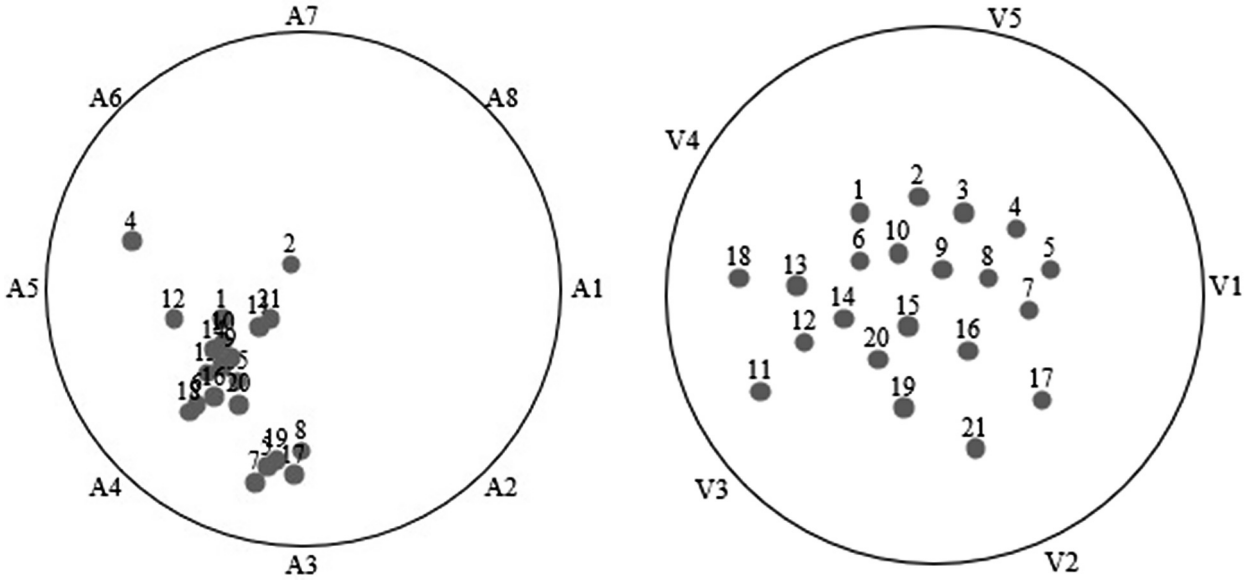
5. Experiments and Results Analysis

Figure 8 visually displays the intelligent agriculture inputs and outputs before and after dimensionality reduction. Figure 9 visually displays the crop growth data of intelligent agriculture on RadViz plane before and after dimensionality reduction.

As shown in Figures 8 and 9, the intelligent agriculture inputs and outputs / crop growth data of intelligent agriculture were overlapped, when they were directly displayed on RadViz plane. The overlap problem could be effectively solved by reducing the dimensionality of intelligent agriculture inputs and outputs from 8 to 4, and that of crop growth data from 15 to 7. After the dimensionality reduction, the data distribution became more dispersed.

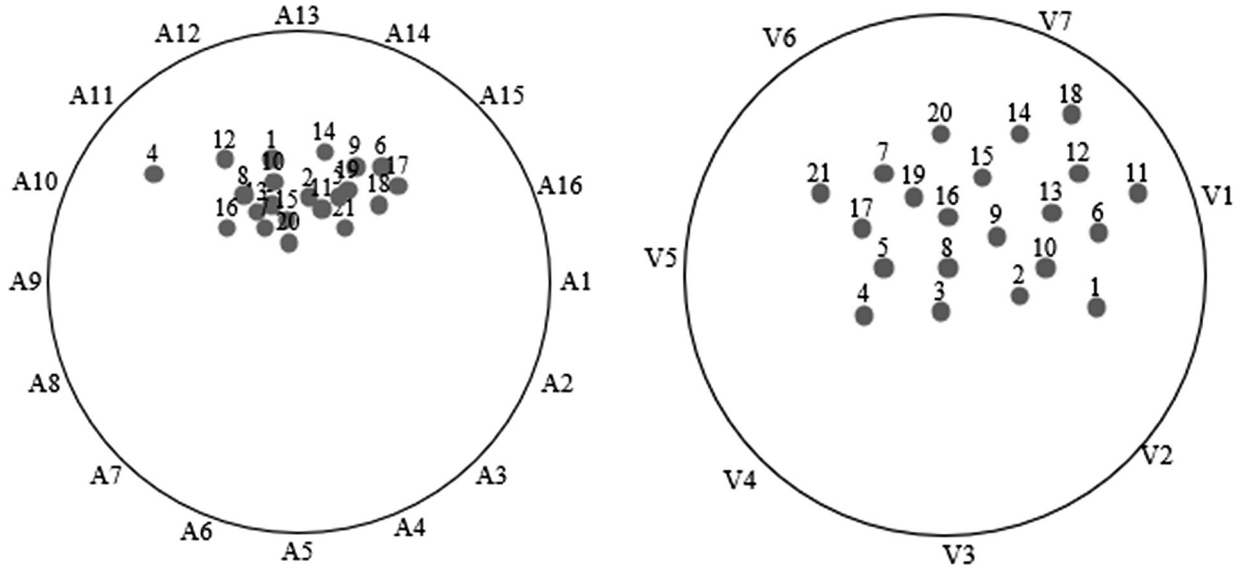
Figure 10 visually displays the clustered intelligent agriculture data after dimensionality reduction. It can be seen that the different clusters of intelligent agriculture data were distinguished clearly by signs in different colors and shapes. By adding digital labels to distinguish the planting areas or crop types on the visual interface, the users can utilize the client more conveniently, and obtain the information of the data samples more easily.

Here, a multiple regression model is constructed based on the sweet potato yields in 2000-2013 of a city in China, as well as the factors



(a) Before dimensionality reduction. *(b)* After dimensionality reduction.

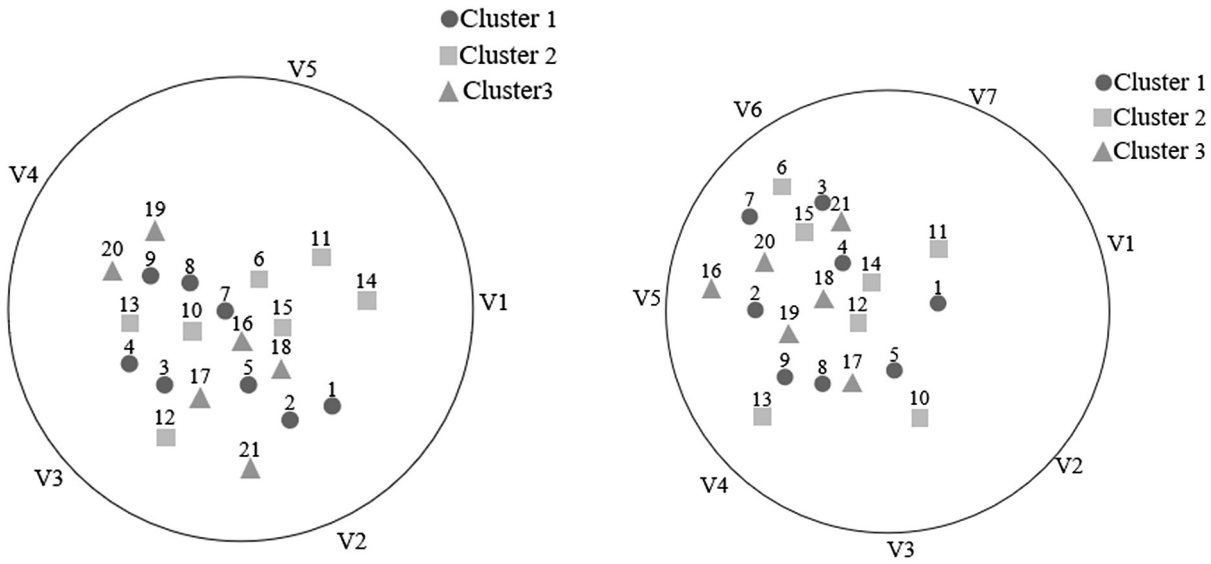
Figure 8. Visual display of intelligent agriculture inputs and outputs before and after dimensionality reduction.



(a) Before dimensionality reduction. *(b)* After dimensionality reduction.

Figure 9. Visual display of crop growth data of intelligent agriculture on RadViz plane before and after dimensionality reduction.

affecting the yield. Then, the data in 2014-2020 were taken as the test set to verify the effectiveness of our crop yield prediction model. The main factors were selected as follows: annual mean temperature, annual rainfall, irrigation volume, fertilizer application, organic matters, nitrogen, phosphorus, and potassium. Each of the above principal components can illustrate the original influencing factor, with no overlapping contents. The influence of the other variables is negligible. Table 1 shows the estimated values of these factors.



(a) *(b)*

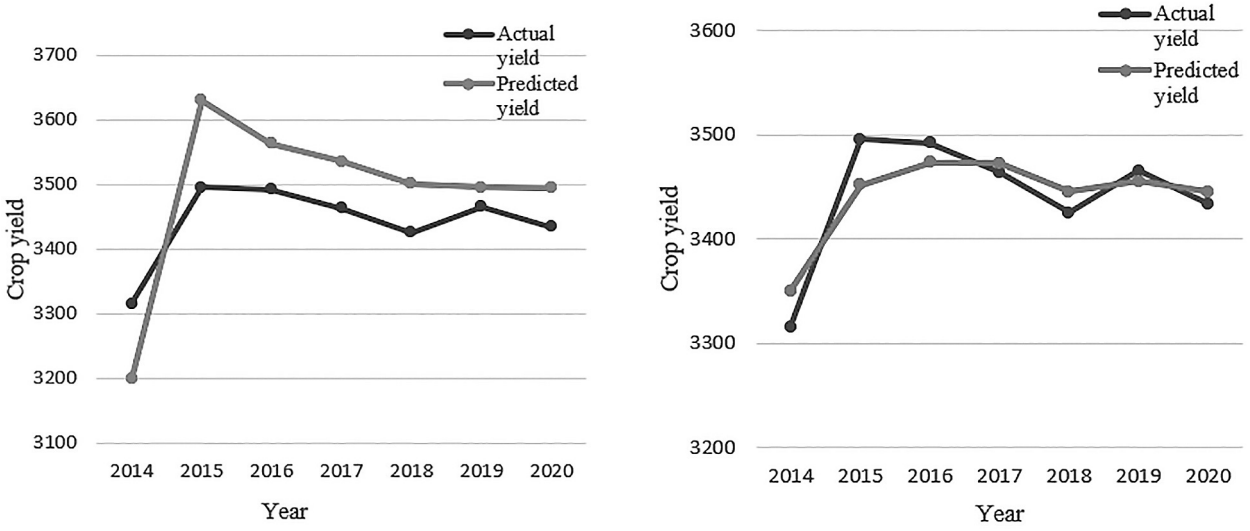
Figure 10. Visual display of clustered intelligent agriculture data after dimensionality reduction.

Table 1. Estimated values of factors affecting crop yield.

Year	Annual mean temperature	Annual rainfall	Irrigation volume	Fertilizer application	Organic matters	Nitrogen	Phosphorus	Potassium
2014	14.36	1,102.75	432	27.32	36.22	137.53	37.55	218.24
2015	14.35	1,008.51	452	28.17	34.35	132.65	38.21	220.72
2016	14.56	1,023.80	447	26.41	35.63	133.29	39.36	219.41
2017	14.92	1,067.31	435	27.63	31.91	144.51	37.28	226.37
2018	14.74	1,113.94	451	30.28	32.57	139.19	36.47	216.92
2019	14.37	1,095.37	449	31.39	28.59	136.05	35.90	218.15
2020	14.19	1,058.19	436	32.09	29.86	137.83	33.64	231.39

Next, contrastive experiments were carried out to evaluate the crop yield prediction effects of multiple regression algorithm 1 and residual principal component regression algorithm 2. Figure 11 compares the predicted results of each algorithm against the actual values. It can be seen that algorithm 2 achieved the better prediction performance, for its predicted values were closer to the actual values.

Finally, the prediction errors of the two models are compared in Table 2. The smaller the error, the better the prediction of crop yield. It can be seen that the MAE and MRE of multiple regression algorithm 1 stood at 135.306 and 4.16%, respectively. Residual principal component regression algorithm 2 lowered the MAE and MRE to 108.641 and 1.80%, respectively. Therefore, residual principal component regression algorithm is much more accurate than multiple regression algorithm in crop yield forecast.



(a) Multiple regression algorithm. *(b)* Residual principal component regression algorithm.

Figure 11. Comparison of crop yield predicted by different regression models.

Table 2. Comparison of prediction errors.

Year	MAE of algorithm 1	MRE of algorithm 1	MAE of algorithm 2	MRE of algorithm 2
2014	132.153	4.35%	136.729	3.71%
2015	235.251	6.32%	75.362	2.15%
2016	95.235	2.36%	76.753	1.35%
2017	96.456	3.67%	89.315	0.59%
2018	125.342	2.14%	127.256	1.35%
2019	167.354	4.96%	119.751	2.76%
2020	95.349	5.35%	135.324	0.67%
Mean error	135.306	4.16%	108.641	1.80%

Note: MAE and MRE are short for mean absolute error and mean relative error, respectively.

6. Conclusion

This paper heavily focuses on the application of data visualization and big data analysis in intelligent agriculture. Firstly, the authors established the framework of intelligent agriculture data visualization and detailed the RadViz data visualization for intelligent agriculture. Afterwards, PCA-based dimensionality reduction and KMC optimization were introduced in order to improve the visualization of intelligent agriculture data. Finally, the crop yield was predicted by multiple regression algorithm and residual principal component regression algorithm. Through experiments, the visualizations of intelligent agriculture inputs and outputs, crop growth data, and clustered data were compared before and after dimensionality reduction. The results show that the proposed model is effective in predicting crop yield, enabling the users to utilize the client conveniently, and obtain the information of the data samples intuitively. Finally, experiments were conducted to compare the two regression algorithms in crop yield prediction and verify their effectiveness.

Acknowledgment

The presented work was supported by Project of Shaanxi Province to improve public Science Quality (No.: 2021PSL39). Authors are grateful to the related departments for the financial supports to conduct this work.

References

[1] S. Mandal *et al.*, "IoT in Agriculture: Smart Farming Using MQTT Protocol Through Cost-Effective Heterogeneous Sensors", in *Proceedings of the International Conference on Frontiers in Computing and Systems, Springer, Singapore*, 2021, pp. 903–913.

[2] X.Y. Geng, "On the Regional Differences in Agricultural Water use Efficiency in China and Their Convergence", *International Journal of Design & Nature and Ecodynamics*, vol. 15, no. 2, pp. 189–196, 2020.

[3] Q. S. Zhou *et al.*, "Demand-Supply Balance of the Nitrogen Nutrients Converted from Regional Agricultural Organic Wastes for Agricultural Utilization", *International Journal of Design & Nature and Ecodynamics*, vol. 15, no. 6, pp. 813–819, 2020.

[4] J. Lin and J.L. Pan, "Farmers' Willingness As Well As Its Compensation Strategy to Prevent and Control Agricultural Non-Point Source Pollution In Water Source Areas-Based on A Comparative Analysis of Pesticide and Fertilizer Inputs", *International Journal of Sustainable Development and Planning*, vol. 15, no. 8, pp. 1303–1311, 2020.

[5] R. U. Onyeneke *et al.*, "Status of Climate-Smart Agriculture in Southeast Nigeria", *GeoJournal*, vol. 83, no. 2, pp. 333–346, 2018. https://doi.org/10.1007/s10708-017-9773-z

[6] S. S. S. Palakodati *et al.*, "Fresh and Rotten Fruits Classification using CNN and Transfer Learning", Revue d'Intelligence Artificielle, vol. 34, no. 5, pp. 617–622, 2020.

[7] K. Lakhwani *et al.*, "Development of IoT for Smart Agriculture a Review", in Emerging trends in expert applications and security, pp. 425–432, 2019.

[8] A. Sahour *et al.*, "Greenhouse Climate Controller by Using of Internet of Things Technology and Fuzzy Logic", *Instrumentation Mesure Métrologie*, vol. 20, no. 1, pp. 29–38, 2021. https://doi.org/10.18280/i2m.200105

[9] H. B. Fiehn *et al.*, "Smart Agriculture System Based on Deep Learning", in *Proceedings of the 2nd International Conference on Smart Digital Environment*, 2018, pp. 158–165.

[10] Y. Li *et al.*, "Studies in Informatics And Control: A Bibliometric Analysis from 2008 to 2019", *International Journal of Computers Communications & Control*, vol. 14, no. 6, pp. 633–652, 2019. https://doi.org/10.15837/ijccc.2019.6.3753

[11] Y. Xiao *et al.*, "A Multidimensional Information Fusion-Based Matching Decision Method for Manufacturing Service Resource", *IEEE Access*, vol. 9, pp. 39839–39851, 2021. https://doi.org/10.1109/ACCESS.2021.3063277

[12] I. Zyrianoff *et al.*, "Scalability of an Internet of Things Platform for Smart Water Management for Agriculture", in *Proc. of the 2018 23rd Conference of Open Innovations Association (FRUCT)*, 2018, pp. 432–439.

[13] T. A. Ahanger, "Defense Scheme to Protect Iot From Cyber Attacks Using AI principles", *International Journal of Computers Communications & Control*, vol. 13, no. 6, pp. 915–926, 2018. https://doi.org/10.15837/ijccc.2018.6.3356

[14] G. Branca *et al.*, "Assessing the Economic and Mitigation Benefits of Climate-Smart Agriculture and Its Implications for Political Economy: A case study in Southern Africa", *Journal of Cleaner Production*, vol. 285, p. 125161, 2021. https://doi.org/10.1016/j.jclepro.2020.125161

[15] M. E. Sahin *et al.*, "Design and Realization of a Hyperchaotic Memristive System for Communication System on FPGA", *Traitement du Signal*, vol. 37, no. 6, pp. 939–953, 2020. https://doi.org/10.18280/ts.370607

[16] A. Chandavale *et al.*, "Automated Systems for Smart Agriculture", in *Proc. of the 2019 IEEE Pune Section International Conference (PuneCon)*, 2019, pp. 1–6.

[17] G. Suciu *et al.*, "Secure Smart Agriculture Monitoring Technique Through Isolation", in *Proc. of the 2019 Global IoT Summit (GIoTS)*, pp. 1–5, 2019.

[18] A. N. Putri *et al.*, "Smart Agriculture Using Supply Chain Management Based on Hyperledger Blockchain", *IOP Conference Series: Earth and Environmental Science*, vol. 466, no. 1, p. 012007, 2020. https://doi.org/10.1088/1755-1315/466/1/012007

[19] D. Yadav *et al.*, "A Novel Convolutional Neural Network Based Model for Recognition and Classification of Apple Leaf Diseases", *Traitement du Signal*, vol. 37, no. 6, pp. 1093–1101, 2020. https://doi.org/10.18280/ts.370622

[20] S. Uppalapati, "Energy-Efficient Heterogeneous Optimization Routing Protocol for Wireless Sensor Network", *Instrumentation Mesure Métrologie*, vol. 19, no. 5, pp. 391–397, 2020. https://doi.org/10.18280/i2m.190510

[21] E. Symeonaki *et al.*, "Review on the Trends and Challenges of Cloud Computing Technology in Climate-Smart Agriculture", in HAICTA, pp. 66–78, 2017.

[22] M. Keivani *et al.*, "Automated Analysis of Leaf Shape, Texture, and Color Features for Plant Classification", *Traitement du Signal*, vol. 37, no. 1, pp. 17–28, 2020. https://doi.org/10.18280/ts.370103

[23] M. Manideep *et al.*, "Smart Agriculture Farming with Image Capturing Module", in *Proc. of the 2019 Global Conference for Advancement in Technology (GCAT)*, 2019, pp. 1–5.

[24] K. Saranya *et al.*, "IoT Based Pest Controlling System for Smart Agriculture", in *Proc. of the 2019 International Conference on Communication and Electronics Systems (ICCES)*, 2019, pp. 1548–1552.

[25] G. S. Nagaraja *et al.*, "Iot Based Smart Agriculture Management System", in *Proc. of the 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, vol. 4, pp. 1–5, 2019.

[26] P. Chen *et al.*, "Super-Durable, Low-Wear, and High-Performance Fur-Brush Triboelectric Nanogenerator for Wind and Water Energy Harvesting for Smart Agriculture", *Advanced Energy Materials*, vol. 11, no. 9, p. 2003066, 2021. https://doi.org/10.1002/aenm.202003066

[27] A. Khatri-Chhetri *et al.*, "Potential of Climate-Smart Agriculture in Reducing Women Farmers' Drudgery in High Climatic Risk Areas", Climatic Change, vol. 158, no. 1, pp. 29–42, 2020. https://doi.org/10.1007/s10584-018-2350-8

[28] F. Mazzetto *et al.*, "Reflections and Methodological Proposals to Treat the Concept of 'Information Precision' in Smart Agriculture Practices", *Sensors*, vol. 20, no. 10, p. 2847, 2020. https://doi.org/10.3390/s20102847

Contact addresses:
Wei Liu
School of Information Engineering
Shaanxi Xueqian Normal University
Xi'an
China
e-mail: yoyotianxia@163.com

Wei Liu was born in Sanyuan County, Shaanxi Province in 1976. She received her bachelor's degree in computer applications in 1997, master's degree in mechanical electronic engineering in 2009 and PhD in mechanical engineering in 2016. Since 2021, she has been a professor with the School of Information Engineering, Shaanxi Normal University. She authored three books, her research interests include computer application design and IoT application system design.

Received: April 2022
Revised: May 2022
Accepted: June 2022