International Journal of Information Systems and Supply Chain Management
Volume 17 • Issue 1

Application and Research of Key Technologies of Big Data for Agriculture

Lei Yang, Anhui University of Finance and Economics, China
Huijuan Ye, Jiangsu Agri-Animal Husbandry Vocational College, China*

ABSTRACT

With the rapid development of science and technology, advanced technical means such as information technology have been widely used in various fields of society, realizing the reform and innovation in different fields. Mage data technology, Internet, cloud computing technology and so on have changed people's production and lifestyle, and society has entered a new era of mage data. Combined with the current situation of agricultural development, we should create a mage data training base for enterprises, build a perfect training system, carry out cooperation with colleges and universities, vigorously introduce professional talents, and spare no effort to promote the effective application of mage data technology in the agricultural field. This paper analyzes the importance of the application of mage data technology in the agricultural field, mage data mining analysis provides decision support, and the specific application of mage data technology in the agricultural field. Finally, through k-means algorithm, the industrial scale can be increased to more than 75%.

KEYWORDS

Agricultural Field, Mage Data, Technology Application and Research

With the development of the times, the agricultural sector needs to adapt to new requirements and introduce information technology to ensure the healthy development of its sector. As the foundation of human cognition, image processing technology is gradually being applied in agricultural work and has significant implications for agricultural development. This article aims to analyze the importance of image data technology in the field of agriculture and explore its role in decision support and specific applications. Firstly, the authors will introduce the importance of image data technology in the field of agriculture. At the same time, through data mining techniques, the authors will fully utilize a large amount of accumulated agricultural information, find the inherent connections and patterns between various factors, and provide important guidance for agricultural production, especially for the high yield and high quality of crops. Secondly, this paper will explore the specific applications of image data technology in the field of agriculture. By applying the k-means algorithm, the noisy data in the environmental dataset can be effectively processed, and stable and accurate clustering results can be obtained in a short period of time. Meanwhile, through image data analysis and processing, the authors can continuously adjust and improve agricultural production, thereby improving the accuracy of agricultural condition monitoring. This article improves the scale of agricultural industry

DOI: 10.4018/IJISSCM.344038
*Corresponding Author

through the application of the k-means algorithm and explores the specific application of image data technology in the field of agriculture, providing feasible solutions for the adjustment and improvement of agricultural production. In future research, the authors will continue to explore the development of image processing technology to promote sustainable development in the agricultural field.

The contribution of this research and innovation lies in:

1. The k-means algorithm is used to study the image analysis of data in the agricultural field. It can effectively process the noise data in the environmental data set, and the running time is short. The clustering results are stable and accurate. The data mining technology is applied to the classification of agricultural economic types, which greatly improves the previous classification methods.
2. The big data mining method is used to collect agricultural environmental monitoring data, and the k-means algorithm is used to optimize and adjust the cluster center of environmental monitoring management information clustering to achieve the optimization of classification management. The system has good information classification ability and strong retrieval ability.
3. Based on image data, agricultural production can be continuously adjusted and improved by using the analysis and processing of the data processing platform so as to improve the accuracy of agricultural status monitoring.

The first part is the introduction, which mainly summarizes the research background of key technologies of agricultural image data visual recognition. The second section is the current related research, analysis of image data, and intelligent agricultural management related research. The following section is the map/reduce data Hadoop model platform and k-means algorithm. The parameters and operation process of the algorithm are described in detail. The next section describes the importance of image data technology in the agricultural field. In the development of modern agriculture, image data technology is an indispensable and important technical means, which can accurately grasp the basic laws of agricultural production, improve the utilization of agricultural resources, and enhance the efficiency of agricultural production. Using data mining technology to mine a large amount of accumulated agricultural information can overcome the phenomenon of "rich data and poor knowledge," find the internal relations and laws of various factors, guide agricultural production, and have important significance for high yield and quality of crops. The final section is the conclusion. It is found that the map/reduce data processing platform combined with the k-means algorithm can increase the scale of agricultural industry to more than 75%.

LITERATURE REVIEW

In recent years, with the rapid development of science and technology, advanced technical means such as information technology have been widely used in various fields of society, realizing reform and innovation in these fields. Mage data technology is an inevitable product in the context of the information age. Mage data technology, Internet, cloud computing technology have changed people's production and lifestyle, and society has entered a new era of image data. Due to limited accessibility outside of the scientific community, hyperspectral images have not been widely used in precision agriculture. Compared with multispectral imaging, hyperspectral imaging is a more advanced technology, which can obtain the detailed spectral response of target features (De Alwis et al., 2022). It is a useful tool to monitor the temporal and spatial changes of crop morphology and physiological status and support precision agriculture practice (Lu et al., 2020). As a basic resource, data plays a key role in the development of various industries. China is a large agricultural country with diverse agricultural data and huge data systems, which brings great challenges to data analysis, management, storage, and so on. Agricultural remote sensing is one of the core technologies of precision agriculture. It takes into account the variability of the field to manage specific locations, rather than unified management in traditional agriculture. As a general remote sensing data, agricultural remote sensing data has all the characteristics of big data. The collection, processing, storage, analysis, and visualization of agricultural remote sensing big data are the keys to the success of precision agriculture. In the past decades, the use of remote sensing technology for precision agriculture (PA) has increased rapidly. The unprecedented high-resolution (spatial, spectral, and temporal) satellite images have promoted the application of remote sensing in many PA applications, including crop monitoring, irrigation management, nutrient application (Huang et al., 2018), pest management, and yield prediction (Sishodia et al., 2020). The combination of big data technology and modern agriculture can add rich experience to agricultural projects and greatly improve project execution ability and data analysis ability, which shows that modern agriculture can fully exert its technical application advantages to meet the requirements of centralized data processing and analysis. In the field of farming, with the continuous in-depth development of information construction and Agricultural Internet of Things (IoT) technology, as well as the continuous development of electronic information technology in recent years, the diversification and cheapness of electronic collection equipment have led to the rapid progress and development of Agricultural IoT. In addition, the continuous investment of the state in the construction of agricultural science and technology informatization, as well as the vigorous development of smart farming, fine farming, Agricultural IoT technology, and other projects, have made the level of agricultural science and technology achieve rapid development.

Farming has become another key area of mage data application. With the development of computer vision, this technology has been widely used in the field of agricultural automation and plays a key role in its development. Computer vision technology will be combined with intelligent technologies such as deep learning technology to apply to all aspects of agricultural production management based on large-scale data sets. It should be more widely used to solve current agricultural problems and better improve the economy. The ability to automatically monitor farmland is an important capability of precision agriculture, which helps to achieve more sustainable agriculture. Accurate and high-resolution monitoring is a key prerequisite for targeted intervention (Tian et al., 2020) and selective application of agrochemicals (Sa et al., 2018). Based on powerful algorithms and processing systems, image recognition technology can reduce human consumption, so it is widely used in agricultural production. At present, image recognition technology is widely used in the fields of weed recognition, crop pest recognition, picking mature crops, path planning, etc. Ampatidis and Partel (2019) developed a data acquisition and image processing technology using small unmanned aerial vehicles, multispectral imaging, and depth learning convolutional neural networks to evaluate the phenotypic characteristics of citrus crops. This low-cost, automated, high-throughput phenotype technology uses artificial intelligence (AI) and machine learning to provide a consistent, more direct, cost-effective, and fast approach (Ampatzidis & Partel, 2019). Deep learning technology improves the performance of hyperspectral image analysis. Compared with traditional machine learning, the deep learning architecture uses the spatial and spectral information of hyperspectral image analysis. Wang et al. (2021) provided insights and identified potential research directions for deep learning in agricultural hyperspectral image analysis.

With the substantial improvement of social informatization, the agricultural field produces a large amount of relevant data every day, but this data has not been effectively used. Therefore, it is particularly important and critical to research and develop a suitable agricultural mage data application system platform, which will play a vital role in improving the effective utilization of mage data in China's agricultural field. Agriculture has always been an important economic and social sector of humans. Therefore, the use of innovative technologies is crucial for the agro food sector. Naranjo-Torres et al. (2020) introduced the basic principles, tools, and two examples of convolutional neural networks (CNNs) for fruit classification and quality control. The application of CNN in fruit recognition has greatly improved the ability to obtain excellent results (Naranjo-Torres et al., 2020). Agricultural mage data is a data set that combines agricultural regionality, seasonality, diversity, periodicity, and other characteristics. It has a wide range of sources, diverse types, complex structure, and potential value and is difficult to be processed and analyzed by ordinary methods. Agricultural mage data retains the basic characteristics of mage data, such as huge volume, diverse types, low value density, fast processing speeds, high accuracy, and complexity and extends and deepens the information flow within farming. Agricultural mage data is the practice of mage data concepts, technologies, and methods in farming. Agricultural mage data involves farmland, sowing, fertilization, pest control, harvesting, storage, breeding, and other links. Therefore, in order to make effective use of agricultural mage data, all relevant units and individuals are developing and studying the key technologies of agricultural mage data platforms.

The country has made full use of the innovation and changes brought by mage data to the agricultural field and has greatly reduced the trial-and-error cost of agricultural development through distributed databases, cloud storage, and virtualization technology, realizing the transformation to precision farming. At present, the important force of agricultural production is still small farmers who mainly operate at home (Naranjo-Torres et al., 2020). The decentralized management mode has the characteristics of extensive, seasonal, and regional, and the agricultural information generated is difficult to be effectively applied to production practice. Therefore, the development of agricultural mage data should take into account the needs of large-scale production while accelerating the modernization of small farmers and should be based on serving the whole process of agricultural production. In the era of mage data, data resources are growing explosively. The development of mage data in farming has achieved multidimensional, full sample, and multitype collection of data resources between different regions and departments.

Obstacles to practical use include the lack of understanding of synthetic aperture radar (SAR) data by agricultural end-user agencies. Dingle Robertson et al. (2020) reviewed the operation sequence of SAR data processing and how sequence selection affects processing time and classification results (Rakhmanov & Wiseman, 2023). This is an important consideration when designing and delivering operating systems, especially for large areas that require hundreds of SAR images. These findings will encourage national, regional, and global food monitoring initiatives to consider SAR sensors as an important data source for agricultural production maps (Dingle Robertson et al., 2020). Horng et al. (2019) proposed a harvesting system based on IoT technology and intelligent image recognition. Agricultural decision-making requires rich experience; with the proposed system, crop maturity can be determined by training the target detection of the neural network model, and then the robot arm can be used to harvest mature crops (Horng et al., 2019). The automation of agricultural activities based IoT can transform the agricultural sector from static and manual to dynamic and intelligent, thereby improving productivity and reducing manpower. PA and wireless sensor network (WSN) are the main drivers of agricultural automation. Shafi et al. (2019) uses a low altitude remote sensing platform to obtain multispectral images and further process them to classify healthy and unhealthy crops. Wang et al. (2018) reviewed various applications of terahertz imaging in agriculture and food engineering, including food safety and quality testing, as well as agricultural fields, such as seed testing, water content evaluation of plant leaves, etc.

The increasing availability and quantity of remote sensing data (such as Landsat images) enable multidimensional analysis of land use/land cover change. Viana et al. (2019) applied long-term land use and land cover (LULC) analysis in rural areas. The experiment was conducted in a rural area characterized by a mixed environment of agriculture and animal husbandry. It considers the seasonality of the LULC type and has realized the overall accuracy value of classification based on the time series (Viana et al., 2019). Unmanned aerial vehicles (UAVs) are becoming an interesting acquisition system for weed location and management, because they can obtain images of the entire farmland with very high spatial resolution and low cost. Bah et al. (2018) proposed a new fully automatic learning method, which uses CNNs and unsupervised training data sets to detect weeds from UAV images. Viljanen et al. (2018) evaluated a new machine learning technology for estimating grassland canopy height and biomass using multispectral photogrammetric camera data. It shows that the proposed multispectral photogrammetry method can provide an accurate estimation of grassland biomass and can be developed as a low-cost tool for practical agricultural applications (Viljanen et al., 2018). Precision agriculture depends on obtaining accurate knowledge of crop phenotypic traits at the field level. Marani et al. (2021) studied the use of the depth learning framework to automatically segment grape clusters in the color images of consumer red, green, blue-depth (RGB-D) cameras placed on agricultural vehicles. Field test results show that the proposed strategy improves the average segmentation accuracy of the four depth neural networks in the range of 2.10 to 8.04% (Marani et al., 2021). Toscano et al. (2019) analyzed the performance of estimating the yield of durum wheat using different techniques and data processing methods. The potential use of Sentinel-2 and Landsat-8 images for field production variability in precision agriculture was evaluated. Remote sensing data analyzed using these methods can be used to assess durum wheat yields and most importantly to describe spatial variability in order to adopt site-specific management, improve productivity, save time, and provide potential alternatives to traditional farming practices (Toscano et al., 2019).

Nitrogen plays an important role in the growth and productivity of rice plants. Sethy et al. (2020) proposed a prediction method of rice nitrogen deficiency based on a CNN. The accuracy score of the CNN classification model is compared with other traditional image classification models (such as feature bag and color feature) (Sethy et al., 2020). Zhong et al. (2018) designed and implemented a vision based fly counting and classification system. The structure of the system is as follows: First, a yellow sticky trap is installed in the monitoring area to capture flying insects, and then a camera is set to capture images in real time (Zhong et al., 2018).

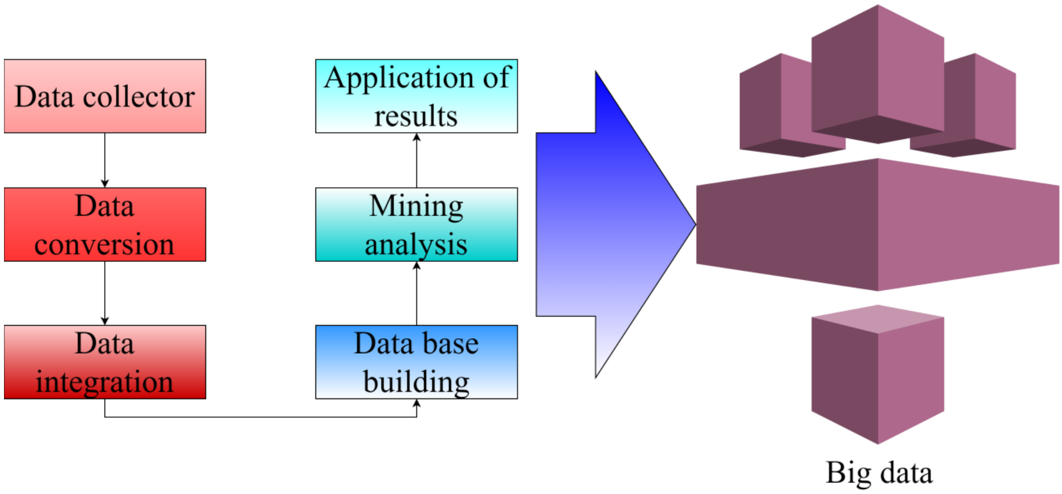
The development of mage data not only is the evolution of technology, but also breaks the traditional data statistics mode and realizes the unstructured mining of multilevel and multichannel data. This is fully proved. Mage data greatly expands the data sources used in various fields and stores the seemingly disordered and jumbled massive data into the public and open data cloud space through high-precision advertising analysis models. Image remote sensing is expected to become an important new technology to help farmers carry out precision agriculture, especially crop nutrition management. Most farmers may not have the technology to benefit from the management plan, which will bring maximum environmental and economic benefits to planned crop management operations (Hunt & Daughtry, 2018). The combination of mage data technology and farming overcomes the weaknesses of small-scale agricultural production and decentralized operations, such as broad data, complex collection, and difficult decision-making, and enhances the ability of data integration and data sharing of the entire agricultural industry chain. It can be said that farming itself is an ideal application field of mage data. The current application of mage data in farming includes but is not limited to agricultural production, agricultural services, and agricultural management. It is very important for agricultural management and sustainable agricultural development to obtain large-scale, high-resolution crop type distribution maps quickly and accurately. Due to the limitation of remote sensing image quality and data processing capability, large-scale crop classification is still challenging (Chong et al., 2021).

The information provided by crops can be transformed into profitable decisions only after being effectively managed. The current progress in data management has led to an exponential growth of intelligent agriculture because data has become a key element in modern agriculture to help producers make key decisions. Access to objective information through sensors to maximize productivity and sustainability will bring valuable advantages (Saiz-Rubio & Rovira-Más, 2020).

MATERIALS AND METHODS

At present, the scattered construction of scientific and technological data resources in the field of farming, the wide variety of data, and the diversity of data structures have brought great challenges to data management and utilization. In view of this situation, it is necessary to solve the management and control of data resources from the root. In addition to establishing an effective data management

Figure 1. Framework of Agricultural Mage Data Application System



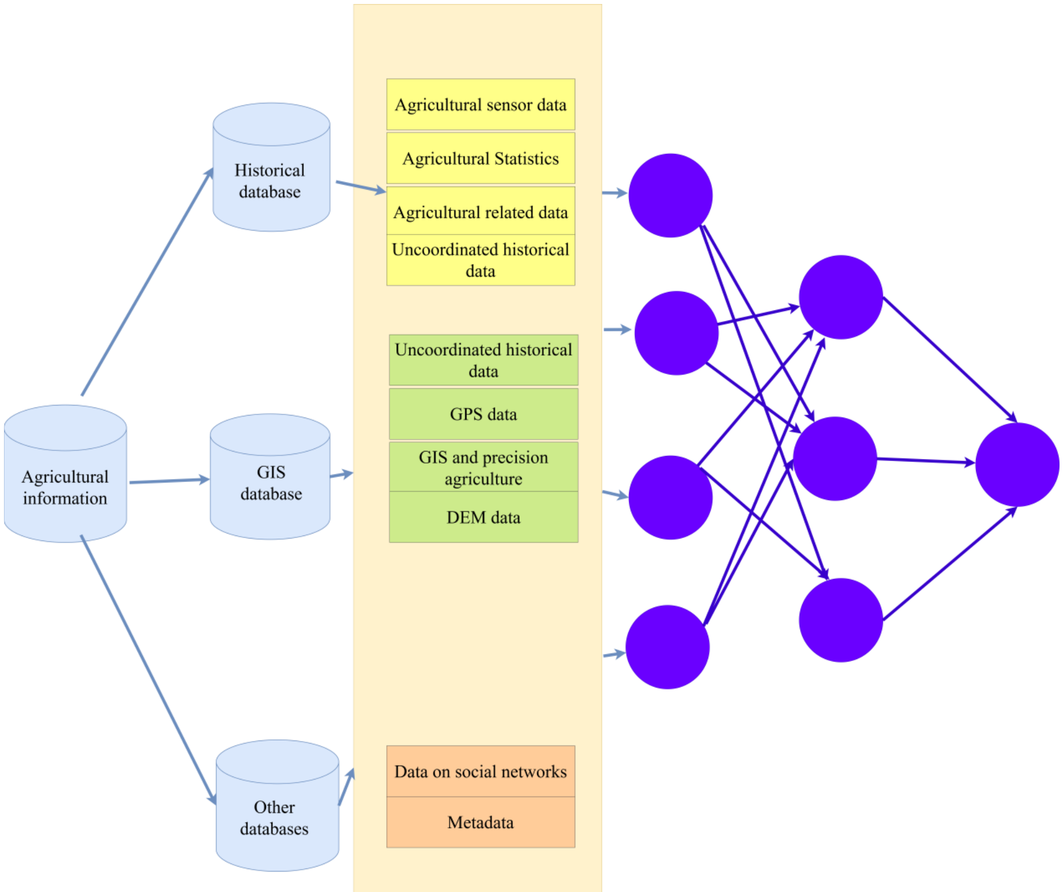
mechanism, it is also necessary to explore the integration and control of scientific and technological data resources in the agricultural field from the perspective of the life cycle of data and effectively monitor and manage agricultural scientific and technological mage data so as to meet the various needs of users in the agricultural field for information and knowledge in the era of mage data and make full use of data resources and fully reflect the value of data (Saiz-Rubio & Rovira-Más, 2020). The application of mage data technology to the agricultural field is an inevitable development trend, which has brought new changes to the production activities in the agricultural field and realized the transmission, transformation, integration, and purification of the explosive growth of agricultural data (Rawat & Kumar, 2015). Combined with agricultural production, business needs, and practical application requirements, comprehensively analyze massive data resources, sort out and refine agricultural data according to the relevant standards of information resource planning, and build the effective information resources scattered in each information island into a complete data center system (Singh & Misra, 2017), which can provide practical support for the unified management and decision-making of farming. The use of mage data technology combined with other advanced technical means plays a decisive role in data analysis, data processing, data mining, and related decisions in the field of farming, ensuring the accuracy of various decisions in the process of modern agricultural development, minimizing cost investment and improving the ultimate benefits of agricultural production.

The application of image data in the agricultural field requires the support of a sound theoretical system and application framework. This paper proposes a system framework that combines image data with farming. The framework is mainly composed of four parts, namely, the agricultural data definition part, sensor part, data collection part, and data mining analysis part (including data conversion, data integration, data database establishment, and mining analysis). Connecting these four parts are three main data streams, namely, the organizational representation of agricultural data, sensor data, and mining analysis result feedback. The foundation of building database and mage data mining analysis is distributed architecture and the cloud platform, as shown in Figure 1.

Mage Data Mining Analysis Provides Decision Support

Predictive data mining is when GIS analysis technology is used to simplify a large amount of data into a single prediction or score through data analysis. In the combined application of GIS and agricultural mage data, GIS can combine the above agricultural data with other relevant statistical data and establish predictive models based on these data to evaluate potential agricultural regions, agricultural

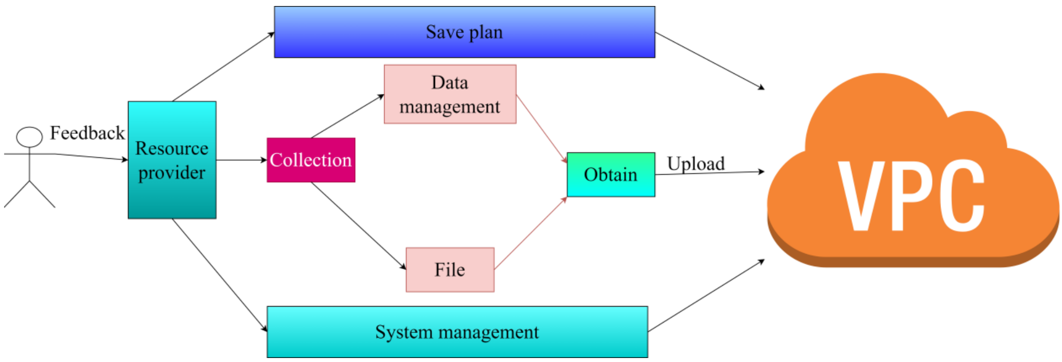
Figure 2. Training Diagram of Agricultural Mage Data Model



markets, agricultural planning projects, and other similar agricultural applications. The agricultural information data source required by mage data mining is mainly divided into three parts. The first part is historical data, which is a batch of data collected and recorded according to time series, including the sensor data mentioned above, statistical data in the agricultural field, cross domain agricultural related data, as well as some uncluttered historical materials and books. The second part is GIS data, including various agricultural thematic maps, GPS data, the application of GIS in precision farming, and digital elevation model data. The third part is other data, including video, audio, pictures, text, and metadata of various data on social networks. After importing the above data into mathematical model mining and analysis, the simulation results are obtained, as shown in Figure 2.

For a long time, China's agricultural informatization research has always focused on the accumulation of agricultural data. Up to now, the scale of agricultural databases has begun to take shape, and most of the stored data information is structured data. In a word, modern agricultural big data technology is the core guiding ideology for expanding the scale of the agricultural industry and can accelerate the transformation and upgrading of the agricultural industry. The long-term preservation of agricultural data decision-making resources is a complex and systematic work, which involves a series of issues such as preservation formats and standards, preservation strategies and methods, preservation metadata, intellectual property rights, preservation costs, etc. The theoretical research on the long-term preservation of agricultural data decision-making resources is rich in content, including not only preservation policies and preservation motives but also preservation subjects. As

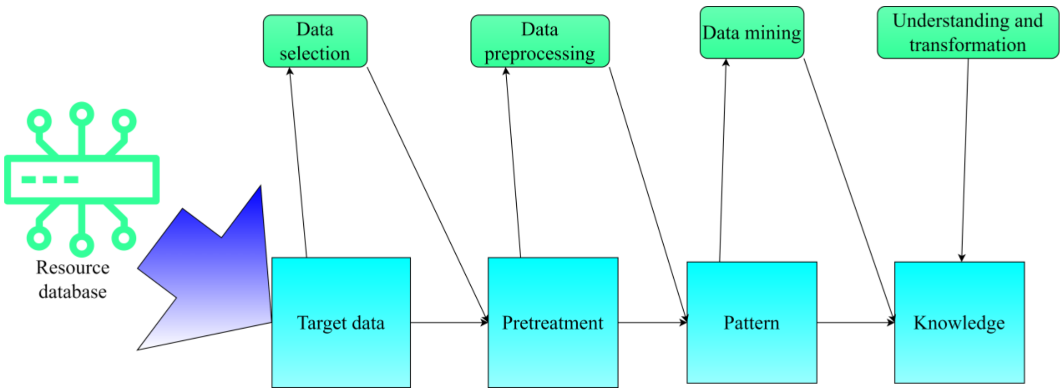
Figure 3. OAIS Reference Model



well as the research on the long-term preservation mechanism of data such as the cooperation mode between entities, it also involves the research on the value and cost of long-term preservation and the technical tools supporting long-term preservation, such as metadata preservation. For how to carry out long-term preservation activities, the model is not based on a specific preservation subject or preservation object. It has wide applicability and can be used as a reference for other preservation metadata projects. The open archival information system (OAIS) reference model is a digital resource archiving system that reflects the interaction process between resource providers, resource managers, and users. OAIS aims to establish a reference model and basic conceptual framework for information systems based on long-term preservation so as to maintain the long-term protection and accessibility of digital information in information systems. It defines three information packages and six functional modules. The three information packages are called submission information package (SIP), archive information package (AIP), and distribution information package (DIP). The six functional modules include collection, storage plan, data management, archiving, system management, and acquisition. This model establishes an overall framework for the long-term preservation of digital information; conceptualizes the environment, functional modules, and information objects related to the digital archive system; and provides a basic reference for the establishment of other long-term preservation models and the implementation of long-term preservation projects, as shown in Figure 3.

At present, there are many methods of processing agricultural information data, and data mining technology is the most widely used. Data mining generally refers to the process of discovering relationships in data, so it is usually called knowledge discovery in the database. One of the main purposes of data mining is to obtain knowledge from data. It is a process of extracting hidden knowledge and information from a large number of noisy, random, incomplete practical data. The emergence of data mining has broken the situation of data islands, making much previously incomprehensible data well used. According to the actual needs of agricultural information services in China, the specific design ideas of agricultural website data mining are as follows. First, analyze a large number of agricultural websites as data sources; second, in order to efficiently extract the useful information hidden in the data source website, it is necessary to select a reasonable agricultural website data extraction technology; third, based on the selected extraction technology, the corresponding extraction program is compiled to extract the hidden useful data from the target agricultural website; fourth, data processing is carried out through a standardized approach, and then data mining algorithms are used to mine the above agricultural information data, and finally relevant information is obtained. The process of data mining is shown in the Figure 4, which defines and understands the data in related fields, preprocesses the obtained data to remove the noise data, selects the appropriate data mining algorithm, and transforms the mining results into knowledge that users can understand.

Figure 4. Data Mining Process



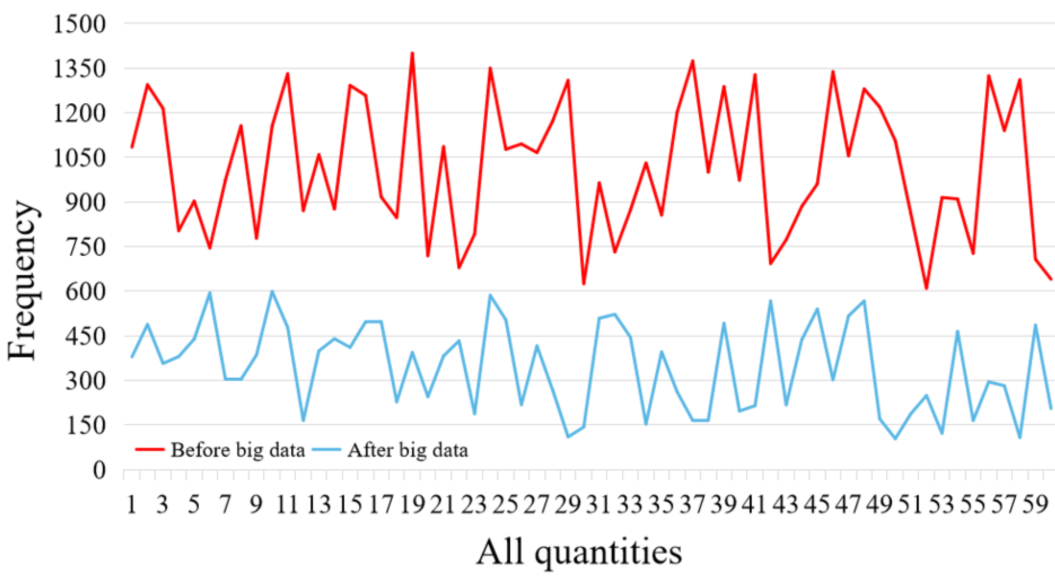
Mage data technology is widely used in the agricultural field, which is reflected in the production links of farming. In recent years, more and more rural people go out to work, resulting in the decreasing number of agricultural populations. In the process of agricultural planting, it requires considerable manpower and material resources if traditional planting methods are adopted. The application of big data, IoT, and other related technologies in agricultural production will significantly improve agricultural planting efficiency. For example, the intelligent control system design of agricultural greenhouses is realized by using mage data technology, including mage data service centers, control terminals, control systems, advanced monitoring systems, etc. Mage data technology can accurately control and manage various mechanical equipment used in the process of agricultural production, can minimize costs, and can improve the quality and efficiency of agricultural production. Problems such as pests and diseases and meteorological disasters may occur in agricultural production, which will have a serious impact on crop growth.

Using mage data technology in agricultural production can monitor the agricultural situation, predict relevant disasters in advance, and adjust and improve the agricultural situation monitoring system after analysis. It can also research the data processing platform. After collecting and analyzing meteorological data, mage data technology can comprehensively improve the prediction accuracy of natural disasters by combining land analysis, plant conditions, and meteorological simulation. In addition, mage data technology can also use advanced remote sensing technology and crop simulation technology to monitor crop growth, master crop production dynamics, and estimate crop yield. Remote sensing satellite monitoring can display the production data of crops as a whole and use the crop growth simulation technology to comprehensively monitor the crop growth environment, simulate the whole process of crop growth and development, and further improve the effect of agricultural prediction. Shown in Figure 5 is the number of problems encountered before and after the application of mage data.

Map/Reduce Data Hadoop Model Platform and K-Means Algorithm

The map/reduce data processing platform is a platform specially launched for processing a large number of data calculations. It adopts an almost completely different solution from the traditional data processing architecture. It is based on the idea of divide and conquer, which divides a large data set into many small data sets and distributes them to multiple nodes for calculation at the same time. Each node will periodically return its own calculation results and working status to the central node at intervals. The central node will calculate the return time of each node. If it exceeds a certain time, it will be considered that the node is wrong, and its tasks will be transferred to other nodes, so

Figure 5. Number of Problems Encountered Before and After Mage Data Application



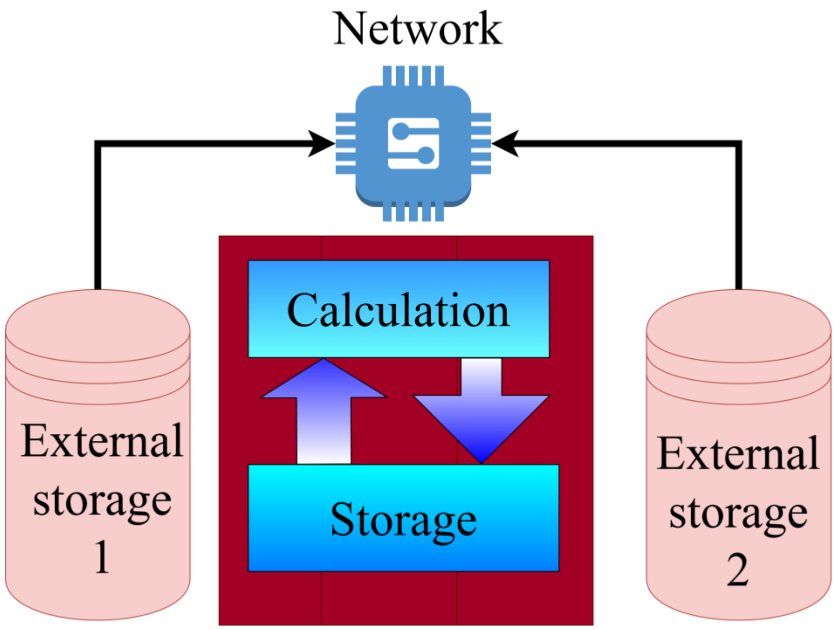
as to ensure the reliability of the whole calculation. The map/reduce programming model provides users with an easy-to-use programming interface. Users only need to write the functions related to the map and reduce processing stages to realize a distributed application, while other complex tasks, such as node communication, node failure, data segmentation, etc. are all completed by the running environment of the map/reduce platform. In order to reduce i/o consumption, the map/reduce model will try to use the data calculation on its own node. This strategy is also called data localization. When there is no required data locally, it will find the data from other nodes and try to find the required data from the local rack, so as to reduce the communication time. These features of the map/reduce processing platform can well solve the problem of processing efficiency faced by agricultural mage data, as shown in Figure 6.

Big data can be processed faster through MapReduce, a parallel processing technology. MapReduce is designed to realize parallel processing of big data through a large number of cheap servers. It has low requirements for data consistency. Its outstanding advantages are scalability and availability, and it is particularly suitable for the mixed processing of massive structured, semi-structured, and unstructured data.

The k-means clustering algorithm is an iterative clustering algorithm. The step is to pre-divide the data into k groups, and then randomly select k objects as the initial cluster center. Then one calculates the distance between each object and each sub-cluster center. Each object is assigned to the nearest cluster center, and the cluster center and the object assigned to it represent a cluster. The clustering center of the k-means clustering algorithm is determined by calculating the mean value of all data object attributes in a cluster. Therefore, the k-means algorithm is usually used to deal with numerical attributes. For example, suppose a cluster is $T_1 = (t_1, t_2, \ldots, t_n)$, then the calculation of its mean value is shown in Eq. (1):

$$x_i = \frac{1}{m}\sum t_n \tag{1}$$

Figure 6. Platform Architecture



In Eq. (1), the authors assume that each tuple has only one numeric attribute value. The k-means clustering algorithm stipulates that the mean value of each cluster must exist, but it can be different from the definition of Eq. (2). The idea of the k-means clustering algorithm is to select randomly from the data set first objects as the initial clustering center. Then one can calculate the distance between each data object and the mean center of each cluster and divide each data object to the nearest center point to form a new cluster. Then one recalculates the mean value of each cluster as a new center point. One repeats the above operations until the criterion function converges. The criterion function of the k-means algorithm is defined as Eq. (2):

$$E = \sum_{i-1} | x - x_i | \tag{2}$$

The first step of the spectral clustering algorithm is to determine the similarity matrix between data pairs in the data set; the second step is to calculate the eigenvector of the similarity matrix; and finally cluster the eigenvector to reflect the relationship between the corresponding data points. From the operation principle and process of the spectral clustering algorithm, it can apply spectral analysis to the data matrix, derive another feature of clustering data after analysis, and then cluster this feature. The proposed algorithm is based on spectral graph theory and adopts the optimal division idea of graph in theory. The spectral clustering algorithm imagines the data set as an undirected graph partition of multidimensional space and then integrates the concept of graph optimal segmentation to complete clustering calculation. In this way, the clustering problem is cleverly transformed into a pair of graphs G. The similarity calculation of the spectral clustering algorithm generally uses the Gaussian kernel function to calculate the similarity matrix S Elements as shown in Eq. (3):

Table 1. Algorithm Example Data Table

Serial number	Attribute 1	Attribute 2
1	$\mathrm{a}+b$	a_1+b_1
2	$a+c$	a_1+c_1
3	$\mathrm{a}+d$	a_1+d_1

$$S_{JJ}=Exp(-\frac{\|x_1-x_2\|}{2q^2}) \tag{3}$$

The processed matrix is defined as the connection matrix. Then the processed degree matrix obtained through matrix transformation needs to be processed. The formula is shown in Eqs. (4) and (5):

$$D=(d_i)=\sum_{j=1} w_{jj} \tag{4}$$

$$X=|x_i|\subseteq R=1,2\ldots,n \tag{5}$$

The k-means algorithm adopts the iterative method. In the iterative process, the data object is continuously assigned to the nearest cluster until the clustering criterion function converges. The input and output of the algorithm are shown in Table 1.

The first iteration: suppose two data objects, No. 1 and No. 2, are randomly selected as the initial cluster center and then the distance between each remaining data object and the two cluster centers is calculated. This distance is used to measure the similarity of objects, and the standardized European distance is used here. The standardized Euclidean distance formula is obtained from the Euclidean distance formula and standardized formula, as shown in Eqs. (6)–(8):

$$\mathrm{a}+b=\sqrt{|x_i+x_y|^2} \tag{6}$$

$$X=\frac{X-M}{S} \tag{7}$$

$$D(X,Y)\sqrt{|\sum I-1\frac{X_I}{S_I}|^2} \tag{8}$$

Table 2. K-Means Instance Clustering Process and Results

Iterations	1 Mean value	2 Mean value	New value
1	(0, 1)	(2, 2.5)	(1, 1.5)
2	(1, 1.5)	(2.5, 3)	(1.5, 2)
3	(1.5, 2)	(3, 3.5)	(2, 2.5)

The third iteration: according to the new cluster center calculated in the second iteration, the data objects are re-divided and the final new cluster is $(1,2,3,4)$ and $(5,6,7,8)$. This shows that the allocation of data objects is no longer changed in the clustering process, and the criterion function converges and the program terminates. The process and results of the change of cluster center and the new cluster generated in the whole clustering process are shown in Table 2.

Calculated by matrix k Eigenvalues and eigenvectors are obtained. According to the obtained eigenvectors, the k-means algorithm is used to cluster the eigenvectors k Manifold structure. The algorithm steps are described in Eq. (9).

$$X =| X_I \mid (C_1, C_2, \ldots C_N) \tag{9}$$

The k-means algorithm is widely used because of its simplicity and efficiency. But the algorithm also has some shortcomings. First, the k-means algorithm can be used only when the average value of clusters is meaningful. If the data object is name data, then this method cannot be used. Second, the algorithm is easily affected by the initial clustering center. The randomness of the initial clustering center may affect the clustering results. Third, the k-means algorithm has poor performance in finding non-convex clusters or clusters with large size differences. Not only that, the algorithm is also sensitive to "noise" and outlier data. A small amount of "noise" data or outlier data will have a great impact on the calculation of the clustering center, and in reality, "noise" and outlier data will be widespread; this is one reason that limits the scope of use of the k-means algorithm. During the execution of the algorithm, the minimum distance from all points that are not cluster centers to the cluster center set is updated, that is, every remaining point in the entire data set except the data points that have been marked as cluster centers is calculated X to cluster center set Y. The calculation of the shortest distance is shown in Eqs. (10)–(13):

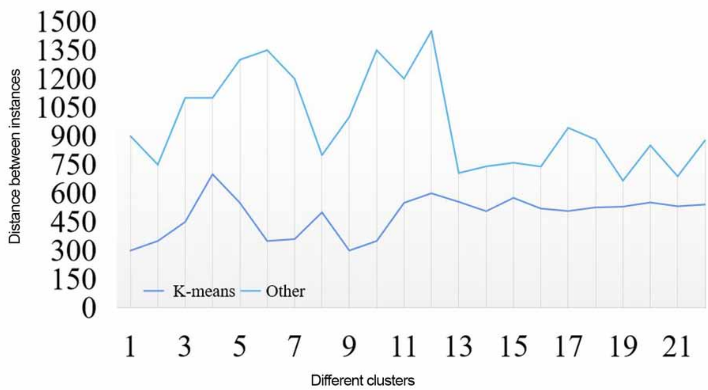
$$D(X,Y) = \min \sum (x,y)\{Y \in R\} \tag{10}$$

$$(t+1) = 1 - pt_{jj}(t) + \Delta j \tag{11}$$

$$\Delta \mathrm{t} = \sum \Delta k_j \tag{12}$$

$$\Delta \mathrm{t}_{ij} = q \setminus L_k \tag{13}$$

Figure 7. Distance Between Instances in the Same Cluster



After one completion, a feasible solution represented by the particle traversal order will be formed. In order to improve the convergence speed of the algorithm, it is necessary to construct the fitness function through the path cost and diversity index and screen the feasible solutions with low fitness value through the roulette selection operator. Eqs. (14) and (15) show that, in order to improve the diversity of the feasible solution set and avoid the algorithm falling into local optimal solution.

$$H_r = h_q^p / m \tag{14}$$

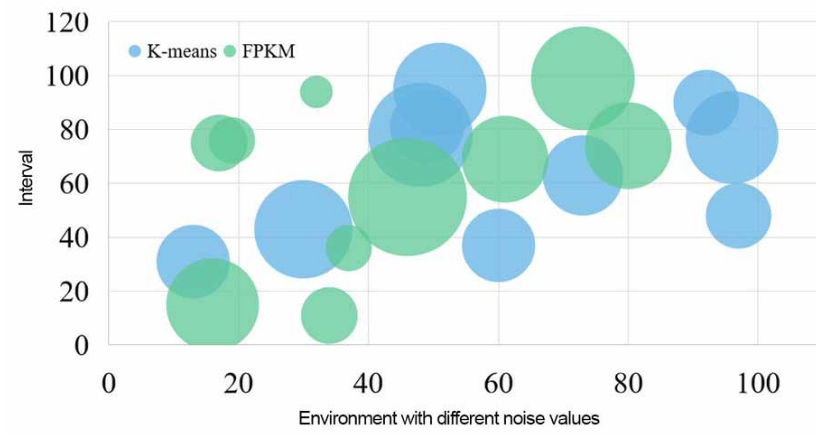
$$L_R = \sum_{Q=1}^{N-1} (X_Q - X_{Q+1}) \tag{15}$$

Since the purpose of clustering is to "maximize the difference between clusters and minimize the difference within clusters," the clustering result is represented by the sum of squares of errors within clusters (that is, the distance from the instance in each cluster to the center of the cluster). The more the sum of squares of errors within clusters is, the smaller the distance between instances in the same cluster is and the better the clustering result is, as shown in Figure 7.

The sum of squares of intra cluster errors of the k-means algorithm is much smaller than that of the unimproved fragments per kilobase per million mapped fragments (FPKM) algorithm, and the more obvious the sum of squares of intra cluster errors is, the better the clustering result is. Therefore, combined with the above two aspects of analysis, the k-means algorithm can effectively process the noise data in the environmental data set, and the running time is short, the clustering result is stable and accurate, and the clustering effect is significantly improved compared with that of the FPKM algorithm, as shown in Figure 8.

The big data mining method is used to collect the data of agricultural environmental monitoring, and the autocorrelation fusion feature scheduling method is used to classify and retrieve the environmental monitoring management information. According to its distribution characteristics, fuzzy clustering processing is carried out, and the k-means algorithm is used to optimize and adjust the clustering center of environmental monitoring management information clustering to achieve the optimization of classification management. The results show that the system has good information classification ability and strong retrieval ability.

Figure 8. Noise Comparison Between K-Means Algorithm and FPKM Algorithm



SPECIFIC APPLICATION OF MAGE DATA TECHNOLOGY IN THE AGRICULTURAL FIELD

The design of intelligent control systems of agricultural greenhouse. Actively adopting advanced monitoring systems, control systems, control terminals, and mage data service centers can accurately control various mechanical equipment, reduce production costs, and improve production efficiency and quality. The agricultural output in all regions showed an overall upward trend, especially in the central and southwestern regions, as shown in Figure 9.

Table 3 shows the proportion of agricultural output in different regions.

Effectively monitor agricultural conditions. In the process of agricultural production, people will encounter various meteorological disasters, insect disasters, and other phenomena that endanger agricultural production and crop growth. Scientific and reasonable use of mage data technology to monitor agricultural production can predict related disasters in advance. Based on mage data, agricultural production can be continuously adjusted and improved by using the analysis and processing of data processing platforms so as to improve the accuracy of agricultural situation monitoring. Monitoring agricultural production to measure natural disasters mainly analyzes the

Figure 9. Agricultural Output in Different Regions

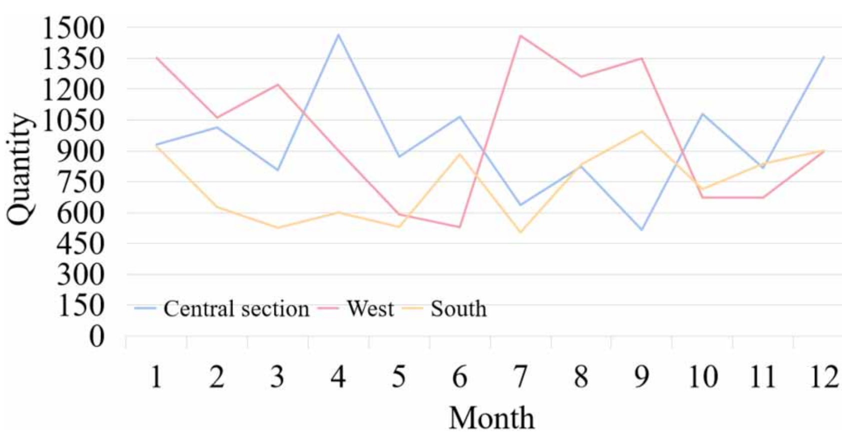


Table 3. Proportion of Agricultural Output in Each Region

Unit%	The first period	The second period	The third period
East Asia	7.08%	7.5%	9.07%
Middle East	12.8%	14.4%	13.01%
Western Region	28.61%	30.9%	30.9%
Western Europe	19.85%	15.44%	15.44%
North	26.66%	24.79%	24.94%

collected meteorological data. According to various factors such as agricultural production simulation, land analysis, and plant root conditions, agricultural production has comprehensively improved the accuracy of natural disaster prediction. Monitoring crop yield estimation and growth dynamics and monitoring crop growth mainly use advanced remote sensing technology of agricultural yield and crop simulation technology. At present, using image recognition technology to identify weeds in the field can be divided into four parts, image acquisition, image processing, accurate positioning, and signal transmission, so as to obtain a prescription map or dynamic image of farmland. Then, one can input this prescription map into the UAV or robot, and the UAV or robot can spray corresponding pesticides and fertilizers to the farmland according to the conditions marked in the prescription map. Remote sensing satellite monitoring can display the growth data of crops as a whole from agricultural output. Crop growth simulation technology mainly monitors the growth environment of crops so as to simulate the whole process of their growth and development. Combining the two aspects of agricultural output can make agricultural prediction more comprehensive and accurate and provide services in the quality and safety management of agricultural output products.

Nowadays, the quality of agricultural products is the focus of the masses. Although the prices of pollution-free products, green agricultural products, and organic agricultural products are high, they still occupy a large share in the market. Some bad businesses put fake and shoddy products into the market for the sake of interests(Bah et al., 2018). However, the use of mage data technology can effectively identify the authenticity of agricultural products. Computer vision technology has gradually been introduced into the detection of the internal quality of agricultural products, including conventional internal quality such as the acidity, sugar content, soluble particles of fruits, protein content of cereals and oils, internal moth eaten, etc., as well as some key indicators of food safety, such as food additives, melamine, aflatoxin. Secondly, mage data technology is used in agricultural materials services. Using mage data technology to collect the life and sales data of agricultural products in agricultural means of supply services, comprehensively analyze various factors, and scientifically and reasonably guide production personnel to plant. Using mage data technology can effectively collect and analyze historical agricultural industry and sales data so one can adopt targeted agricultural technology and improve agricultural production efficiency and levelly.

Using data mining technology to mine a large amount of accumulated agricultural information can overcome the phenomenon of “rich data and poor knowledge” and find the internal relations and laws of various factors to guide agricultural production, which is of great significance to the high yield and high quality of crops. Meteorology plays an important role in agricultural production, which will directly affect the yield of crops. Through the mining of meteorological data in previous years, meteorological conditions in the next few days can be predicted to guide agricultural production. Data mining plays a very important role in agricultural environmental monitoring and environmental protection. In the database of soil nutrients, the evaluation rules of soil fertility are mined to guide the scientific fertilization of agricultural production. From the data of farmland soil environmental quality and the knowledge of crop growth conditions, the quality status of agricultural products can be mined, and the possible environmental or ecological reasons for the deterioration of agricultural products can be analyzed. This paper introduces the technical background of agricultural mage data, data mining, and data mining systems and then introduces the requirements of agricultural mage data for data mining technology and the characteristics of the map/reduce data processing platform. Finally, through the k-means algorithm, the industrial scale can be increased to more than 75%.

CONCLUSION

In the process of agricultural development, researchers need to meet the requirements of the times to join information technology so as to ensure the healthy development of agriculture. As the basis of human understanding, image processing is gradually being applied to agricultural work, which is of great significance for agricultural development. This paper analyzes the importance of image data technology in the agricultural field. Finally, the industrial scale can be increased to more than 75% through the k-means algorithm. Build a mage data training base for enterprises, build a perfect training system, cooperate with universities, vigorously introduce professionals, and fully promote the effective application of data technology in the agricultural field. Researchers should increase the infrastructure construction of mage data, pay attention to the construction activities of telecom enterprises in rural areas, maximize the network coverage, and ensure the stable application of mage data technology in agriculture. The agricultural field lacks image data technology talents. Scientific and reasonable use of image data technology to monitor agricultural production can predict related disasters in advance.

Based on image data, agricultural production can be continuously adjusted and improved by using the analysis and processing of the data processing platform so as to improve the accuracy of agricultural status monitoring. Image processing is composed of many disciplines and has very important application values. The development of image processing in the future will be directly related to people's life and will have a great impact on human life. This article lacks exploration of integration with other related technologies, such as the IoT and big data analysis. Subsequent work can consider exploring the integration of image data technology with other technologies, such as artificial intelligence and machine learning, to further improve the accuracy and efficiency of agricultural decision support systems and enhance the benefits of agricultural development.

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those whose techniques have contributed to this research.

DATA AVAILABILITY

The figures and tables used to support the findings of this study are included in the article.

STATEMENT OF COMPETITIVE INTEREST

The authors declares that they do not have competitive interests.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This research was supported by the Short Text User-Generated Content Classification Algorithm and Its Application, Natural Science Research Project of Anhui Universities in 2021 [grant number KJ2021A0476].

PROCESS DATES

Received: 1/2/2024, Revision: 1/10/2024, Accepted: 1/11/2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Huijuan Ye (yehuijuana@163.com)

REFERENCES

Ampatzidis, Y., & Partel, V. (2019). UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing (Basel)*, *11*(4), 410. doi:10.3390/rs11040410

Bah, M. D., Hafiane, A., & Canals, R. (2018). Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sensing (Basel)*, *10*(11), 1690. doi:10.3390/rs10111690

Chong, L. U. O., Liu, H. J., Lu, L. P., Liu, Z. R., Kong, F. C., & Zhang, X. L. (2021). Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. *Journal of Integrative Agriculture*, *20*(7), 1944–1957. doi:10.1016/S2095-3119(20)63329-9

De Alwis, S., Hou, Z., Zhang, Y., Na, M. H., Ofoghi, B., & Sajjanhar, A. (2022). A survey on smart farming data, applications and techniques. *Computers in Industry*, *138*, 103624. doi:10.1016/j.compind.2022.103624

Dingle Robertson, L., Davidson, A., McNairn, H., Hosseini, M., Mitchell, S., De Abelleyra, D., Verón, S., & Cosh, M. H. (2020). Synthetic Aperture Radar (SAR) image processing for operational space-based agriculture mapping. *International Journal of Remote Sensing*, *41*(18), 7112–7144. doi:10.1080/01431161.2020.1754494

Horng, G. J., Liu, M. X., & Chen, C. C. (2019). The smart image recognition mechanism for crop harvesting system in intelligent agriculture. *IEEE Sensors Journal*, *20*(5), 2766–2781. doi:10.1109/JSEN.2019.2954287

Huang, Y., Chen, Z. X., Tao, Y. U., Huang, X. Z., & Gu, X. F. (2018). Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, *17*(9), 1915–1931. doi:10.1016/S2095-3119(17)61859-8

Hunt, E. R. Jr, & Daughtry, C. S. (2018). What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *International Journal of Remote Sensing*, *39*(15-16), 5345–5376. doi:10.1080/01431161.2017.1410300

Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing (Basel)*, *12*(16), 2659. doi:10.3390/rs12162659

Marani, R., Milella, A., Petitti, A., & Reina, G. (2021). Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. *Precision Agriculture*, *22*(2), 387–413. doi:10.1007/s11119-020-09736-0

Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A. (2020). A review of convolutional neural network applied to fruit image processing. *Applied Sciences (Basel, Switzerland)*, *10*(10), 3443. doi:10.3390/app10103443

Rakhmanov, A., & Wiseman, Y. (2023). Compression of GNSS data with the aim of speeding up communication to autonomous vehicles. *Remote Sensing (Basel)*, *15*(8), 2165. doi:10.3390/rs15082165

Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Sciences*, *18*(1), 77–84. doi:10.1016/j.ejrs.2015.02.002

Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., & Siegwart, R. (2018). WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing (Basel)*, *10*(9), 1423. doi:10.3390/rs10091423

Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. *Agronomy (Basel)*, *10*(2), 207. doi:10.3390/agronomy10020207

Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Nitrogen deficiency prediction of rice crop based on convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 5703–5711. doi:10.1007/s12652-020-01938-8

Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). Precision agriculture techniques and practices: From considerations to applications. *Sensors (Basel)*, *19*(17), 3796. doi:10.3390/s19173796 PMID:31480709

Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, *4*(1), 41–49. doi:10.1016/j.inpa.2016.10.005

Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sensing (Basel)*, *12*(19), 3136. doi:10.3390/rs12193136

Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation—A review. *Information Processing in Agriculture*, *7*(1), 1–19. doi:10.1016/j.inpa.2019.09.006

Toscano, P., Castrignanò, A., Di Gennaro, S. F., Vonella, A. V., Ventrella, D., & Matese, A. (2019). A precision agriculture approach for durum wheat yield assessment using remote sensing data and yield mapping. *Agronomy (Basel)*, *9*(8), 437. doi:10.3390/agronomy9080437

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sensing (Basel)*, *11*(9), 1104. doi:10.3390/rs11091104

Viljanen, N., Honkavaara, E., Näsi, R., Hakala, T., Niemeläinen, O., & Kaivosoja, J. (2018). A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture*, *8*(5), 70. doi:10.3390/agriculture8050070

Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, *54*(7), 5205–5253. doi:10.1007/s10462-021-10018-y

Wang, C., Qin, J., Xu, W., Chen, M., Xie, L., & Ying, Y. (2018). Terahertz imaging applications in agriculture and food engineering: A review. *Transactions of the ASABE*, *61*(2), 411–424. doi:10.13031/trans.12201

Zhong, Y., Gao, J., Lei, Q., & Zhou, Y. (2018). A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors (Basel)*, *18*(5), 1489. doi:10.3390/s18051489 PMID:29747429