

# Farmers Assistance Program Based on Artificial Intelligence (AI) and Big Data

Qi Gao, Liling Zhang, Wanying Yang, Jingyi Xiao, Zhiying Xu

School of Information Engineering, Wuhan Business University, Wuhan 430056, China

## ABSTRACT

This farmers assistance program aims to build an efficient and intelligent agricultural management system through the deep integration of artificial intelligence (AI) and big data technologies, so as to improve agricultural production efficiency, ensure the quality of agricultural products, promote sustainable agricultural development, and ultimately achieve the goals of increasing farmers' income and prosperity in rural areas.

## KEYWORDS

Artificial intelligence; Big data; Farmers assistance

## 1. INTRODUCTION

With the rapid development of technology, the traditional agricultural production model has been difficult to meet the needs of agricultural workers and the market. With the in-depth development of technologies such as cloud computing, machine learning, and deep learning, artificial intelligence (AI) and big data technologies have been widely applied in the agricultural production process, injecting new vitality into traditional agriculture. These technologies can not only process and analyze massive amounts of agricultural data, provide accurate decision support, but also improve the work efficiency of agricultural workers and realize the modernization of agricultural production. Therefore, it is imperative to apply artificial intelligence and big data technologies to agricultural production activities.

## 2. AGRICULTURAL BIG DATA

From the perspective of agricultural market demand, agricultural big data can be used to guide agricultural production, predict the market demand for agricultural products, and assist agricultural decision-making, so as to achieve the expected goals of avoiding risks, increasing production and income, and transparent management. From the perspective of agricultural production links, agricultural big data can use sensors to collect climate and soil big data, provide farmers with the best cultivation management decisions, assist farmers in effectively managing their farmland, and allow farmers to extract the highest value from every seed and reduce agricultural costs.

From the overall trend of agriculture, by analyzing real-time environmental data, the current growth status of crops and plot information can be obtained; by using algorithm models, the future environmental trend can be predicted, and the precise future climate trend and pest and disease trends can be obtained; by analyzing the overall trend of environmental data, precise planting suggestions and management guidance can be obtained.

In short, the application of big data in agriculture can help farmers better manage their farmland, improve agricultural production efficiency, reduce costs, increase income, and also promote the sustainable development of agriculture.

### **3. FURTHER ANALYSIS USING ARTIFICIAL INTELLIGENCE**

#### **3.1. Agricultural Risk Prediction and Model Comparison Based on AI**

Risk prediction helps people take appropriate actions and preparation plans in advance to deal with risk factors to reduce losses and risks. At present, the application of artificial intelligence in risk prediction is mainly based on regression and classification algorithms, and clustering and dimensionality reduction algorithms are often used as data preprocessing methods to participate in risk prediction. Artificial intelligence realizes the fitting of nonlinear information and the learning of features by learning the mapping relationship from input variables to output variables. Therefore, the data type of the input agricultural risk variable itself has an important influence on the use of artificial intelligence for risk prediction. The following mainly starts from the early corn leaf spot disease prediction model, and through the comparison of fluorescence spectrum analysis and AI modeling, expounds the AI algorithm applicable to specific data and its application in agricultural risk prediction.

##### **3.1.1. Experimental design**

Three plots were set up as the experimental group, namely the normal growth (no lesion) area of 16 m×8 m, the potentially diseased (disease incubation period) area of 16 m×4 m, and the early diseased (early disease state) area of 16 m×4 m. To prevent the spread of germs between planting areas, an interval of 1.2 m was set as the standard. The artificial inoculation of spot disease spores method was used to achieve different degrees of corn spot disease infection. At the late jointing stage (13-leaf stage) of corn, when the plant height is about 1 m, the experiment begins. By spraying and inoculating, the conidial suspension of spot disease (from the Plant Pathology Laboratory of Jilin Agricultural University) was inoculated, with a spore density of  $1 \times 10^5 \sim 1 \times 10^6$  spores·mL<sup>-1</sup> and an inoculation amount of 5~10 mL·plant<sup>-1</sup>. The severity of corn spot disease is proportional to the amount of conidia inoculated. The potentially diseased leaves are inoculated once, the early diseased leaves are inoculated three times, and the normal non-diseased leaves are inoculated with an equal amount of distilled water. After 14 days, corn in the inoculated area began to develop spot diseases, and the disease severity was relatively consistent, which could meet the research observation requirements. Among them, for corn large and small spot diseases, refer to the national standard of the People's Republic of China GB/T2339.1-2009 "Technical Specifications for the Prevention and Control of Corn Large and Small Spot Diseases and Corn Borers". In the experiment, CK leaves are grade 0 with no lesions, potentially diseased leaves are grade 0~0.5, and early diseased leaves are grade 0.5~2.

##### **3.1.2. Dataset selection and division**

A total of 190 corn leaf samples were collected in the experiment. After removing 30 abnormal samples, they were reasonably divided according to the training set: prediction set of 3:1, which were 120 samples in the training set and 40 samples in the verification set. Among them, there are 60 samples in the training set and 20 samples in the verification set for CK leaves. There are 30 samples in the training set and 10 samples in the verification set for potentially diseased leaves. There are 30 samples in the training set and 10 samples in the verification set for early diseased leaves.

##### **3.1.3. Study on the response law of corn leaf physiological parameters to spot disease**

###### **(1) Study on the response law of corn leaf SPAD to spot disease**

The relative content of chlorophyll is an important indicator for evaluating the photosynthetic efficiency of plants and environmental stress, and it is also one of the sensitive factors for analyzing

spectral changes. Therefore, real-time and non-destructive monitoring of chlorophyll status is of great significance for agricultural condition monitoring and yield estimation of crops.

In this study, under different disease degrees, the measured values of corn leaf SPAD were obtained. The results are shown in Table 1. The research shows that as *Bipolaris zeicola* and *Bipolaris maydis* attach and reproduce on the leaves, the number of necrotic cells in the leaves increases, the area of oval or spindle-shaped lesions expands, the chlorophyll content drops significantly, the SPAD value of the leaves decreases, and the synthesis and decomposition of light by chlorophyll is inhibited. The photosynthesis of leaves is weakened. As the degree of spot disease increases, the wilting degree of leaves increases, and the chloroplast structure is damaged, especially more obvious after the heading period, which seriously hinders the normal growth, development and fruit formation of corn. The comprehensive analysis above shows that the SPAD value can be used to characterize the law of corn spot disease.

**Table 1.** Statistical table of measured values of SPAD under different disease degrees

Blade class	Maximum	Minimum	Mean	Standard deviation	DS/Number
Normal blade	69.30	50.50	59.60	4.61	80
Potentially diseased leaf	65.30	46.40	57.90	3.75	40
Early-onset leaf	63.80	46.10	55.73	6.22	40

(2) Study on the response law of Fv/Fm of corn leaves to leaf spot disease.

Fv/Fm is the maximum photochemical quantum yield of PSII or the potential maximum quantum yield of PSII. It reflects the quantum yield when the reaction centers of PSII are all in an open state and is generally constant between 0.80 and 0.85. When plants are subjected to photoinhibition, environmental stress, or certain gene mutations occur, the Fv/Fm value will change significantly. Therefore, the fluorescence parameter Fv/Fm is widely used in the research on plant photosynthesis, physiological biochemistry, and the effects of various adverse stress on plant photosynthesis. The degree of damage to the PSII complex caused by adverse stress can be measured by the reduction amplitude of Fv/Fm, reflecting the degree of damage to plant photosynthesis by adversity.

Under different disease degrees, the measured values of Fv/Fm of corn leaves were obtained. The results are shown in Table 2. The research shows that the reduction amplitude of Fv/Fm can be roughly used as an indicator of PSII photoinhibition damage. The average Fv/Fm of normal non-diseased leaves is 0.781, the average Fv/Fm of potentially diseased leaves is 0.751, and the average Fv/Fm of early diseased leaves is 0.726. In the diseased state, Fv/Fm significantly decreased ( $p < 0.05$ ). In the early diseased state, Fv/Fm decreased by 7.04%, and the Fv/Fm of potentially diseased leaves decreased by 3.84%. It can be seen that in the potentially diseased state and the early stage of disease, both the SPAD value and Fv/Fm of the leaves decreased significantly, and the decrease amplitude was relatively large. The results show that the reduction of Fv/Fm at the potential disease stage and the early disease stage of corn plants will lead to damage to the photosynthetic system of plant leaves and a decrease in photosynthetic capacity. The physiological parameter law of leaf spot disease can be characterized by the Fv/Fm parameter.

**Table 2.** Statistical table of measured values of Fv/Fm under different disease degrees

Blade class	Maximum	Minimum	Mean	Standard deviation	DS/Number
Normal blade	0.8651	0.7380	0.7867	0.0299	80
Potentially diseased leaf	0.7760	0.7220	0.7537	0.0147	40
Early-onset leaf	0.7680	0.6576	0.7310	0.0292	40

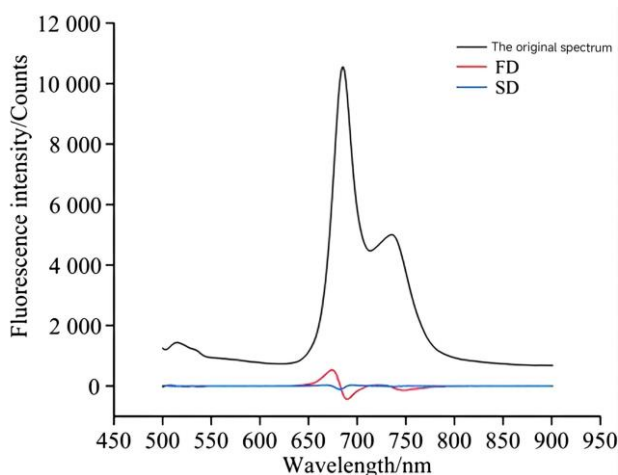
The comprehensive analysis of leaf SPAD and Fv/Fm under different degrees of leaf spot disease shows that as the severity of leaf spot disease increases, both leaf SPAD and Fv/Fm decrease to

varying degrees, which is consistent with the research conclusions of predecessors. This indicates that the more severe the leaf spot disease is, the faster the open ratio and electron transfer rate of the PSII reaction center in corn leaves decrease, and then the energy provided for photosynthesis is reduced, which is not conducive to the accumulation of dry matter in corn in the later stage.

### 3.2. Differential Analysis Based on AI

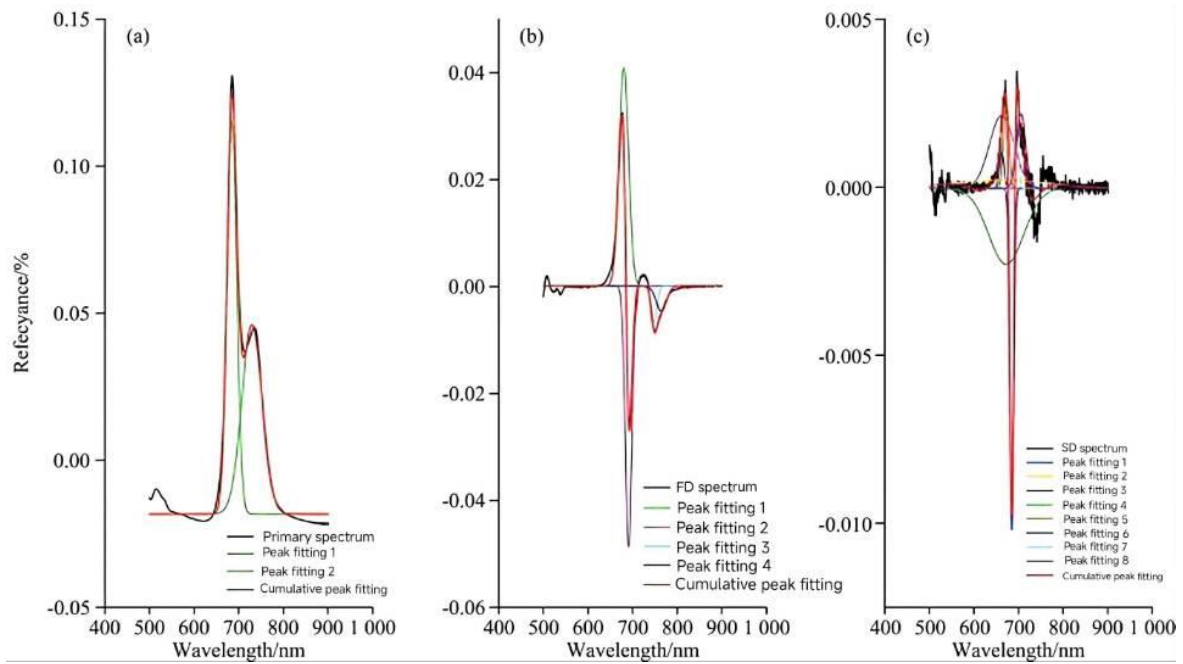
#### 3.2.1. Differential analysis of fluorescence spectra of corn leaves under different disease degrees

During the process of collecting chlorophyll fluorescence spectra, there will be a "jitter" phenomenon between different spectral bands, which will cause certain deviations and impacts on the spectral index calculation results and is not conducive to reflecting the real physiological conditions of plants. The laser transmitter of the chlorophyll fluorescence spectrometer used in the experiment emits a wavelength of 473.000 nm. The collected spectral band range is 331.010~1099.960 nm. Each spectrum has 1355 spectral information, and the data volume is redundant. The spectral region formed by the laser light source in the band 472.305~499.896 nm and the fluorescence spectrum between the band 900.000~1099.960 nm show a small amplitude fluctuation that tends to be stable and flat. After analyzing multiple spectra, it is found that the fluctuation range of this spectral band is small and has little impact on the detection of photosynthetic physiological information of corn leaves. Therefore, the above two bands can be regarded as invalid spectral bands in this experimental study and can be ignored. The original spectra, FD spectra and SD spectra in the range of 500.000~900.000 nm are shown in Figure 1.



**Figure 1.** Fluorescence spectral characteristics of leaves with different pretreatments.

Peak analysis was performed on the original spectrum, FD spectrum and SD spectrum in the range of 500.000–900.000 nm. The results are shown in Table 3. After the original spectrum is processed by FD and SD, multiple spectral curves show peak centers with regular changes. At this time, the number of sensitive bands that can reflect spectral properties increases, indicating an increase in characteristic variables, as shown in Figures 2 (a, b, c). The peak center of the original spectrum appears near 685.456 and 735.540 nm. The peak centers of the FD preprocessed spectrum appear near 676.896, 692.313, 748.573 and 750.271 nm. The peak centers of the SD preprocessed spectrum appear near 665.457, 668.318, 671.751, 685.465, 695.465, 698.015, 700.295 and 703.713 nm. The fluorescence intensities at the peak centers all show obvious fluctuations and differences, indicating that the original band has an extreme value at this band and can be used as a sensitive band for extracting characteristic wavelengths.



(a) : original spectral fitting diagram; (b): FD spectral fitting diagram; (c) : SD spectral fitting diagram

**Figure 2.** Spectral peak fitting diagram

**Table 3.** Peak fitting results

Spectral type	Peak type	Peak weighted mean center /nm	Global crest area of simulation curve	Percentage of crest area /%	Peak x value /nm	Maximum Peak height /nm	Peak width at half height (FWHM) /nm	Acorrelation	$R^2$ (COD)	After adjustment $R^2$
Primary spectrum	Gaussian	685.465	3.643 35	49.666 46	684.732 52	0.133 56	25.627 16	0.861 16	0.971 20	0.971 13
	Gaussian	735.540	3.692 28	50.333 54	729.835 59	0.064 37	53.887 05	0.816 31		
FD	Gaussian	676.896	1.088 19	49.217 59	680.235 66	0.040 73	25.096 41	0.949 35	0.989 85	0.989 67
	Gaussian	692.313	-0.879 66	-39.786 03	691.014 37	-0.048 75	16.952 2	0.959 00		
	Gaussian	748.573	-0.108 27	-4.897 06	748.778 09	-0.006 75	15.064 17	0.959 92		
	Gaussian	750.271	-0.134 86	-0.609 933	764.303 1	-0.004 71	26.917 19	0.959 91		
SD	Gaussian	665.457	0.013 93	2.040 09	662.655 29	0.001 05	12.429 88	0.959 89	0.976 82	0.976 00
	Gaussian	668.318	0.066 88	9.001 07	662.766 59	0.044 99	255.778 48	0.969 71		
	Gaussian	671.751	0.150 59	22.057 59	662.706 59	0.002 18	64.770 64	0.959 92		
	Gaussian	685.465	-0.221 96	-32.510 91	672.433 66	-0.002 26	92.422 7	0.959 76		
	Gaussian	695.465	0.038 43	5.628 58	672.437 74	0.002 58	13.995 63	0.959 80		
	Gaussian	698.015	-0.111 81	-10.376 16	684.923 79	-0.010 13	10.370 92	0.959 89		
	Gaussian	700.295	0.027 92	4.088 89	695.450 2	0.002 63	9.982 46	0.969 99		
	Gaussian	703.713	0.056 64	8.296 71	705.839 88	0.002 22	24.016 99	0.959 99		

Through the comparison and analysis of the original spectrum, FD spectrum, SD spectrum peak center, peak trough position and correlation coefficient, considering the obvious difference of fluorescence spectrum, the optimal characteristic wavelength (OCW) from the three methods is selected as the object of this experiment. The effective spectral band is 500.000~900.000 nm.

### 3.2.2. Modeling analysis of fluorescence spectrum based on different pretreatment methods

#### (1) Establishment of inversion model

With spectral sensitive band (OCW) as model input, multivariate scattering correction (MSC) and standard normal variable transformation (standard normal variable transformation) were adopted. SNV), polynomial smoothing (Savitzky-Golaay, S-G), FD spectrum first derivative preprocessing, SD spectrum second derivative preprocessing, etc. MSC-SG-FD, MSC-FD-SG, SNV-SG-FD, SNV-

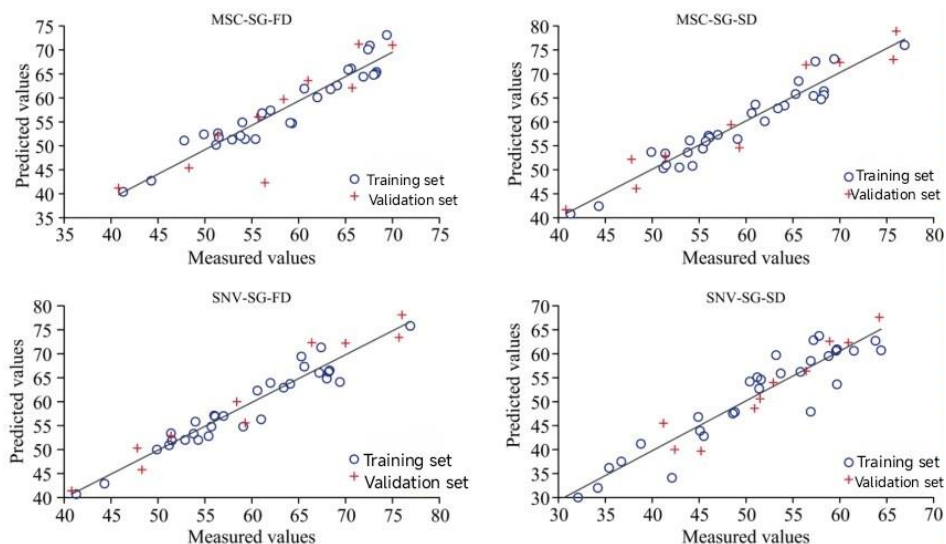
SG-SD and other four modeling combination methods were used to process the original spectrum, and the gradient method was used to select the correction set and prediction set, and the classical least squares of 500.000 ~ 900.000 nm band was established. Inversion model of fluorescence spectrum of SPAD and Fv/Fm (classic-least-squares).

### (2) Comparison of modeling methods and analysis of results

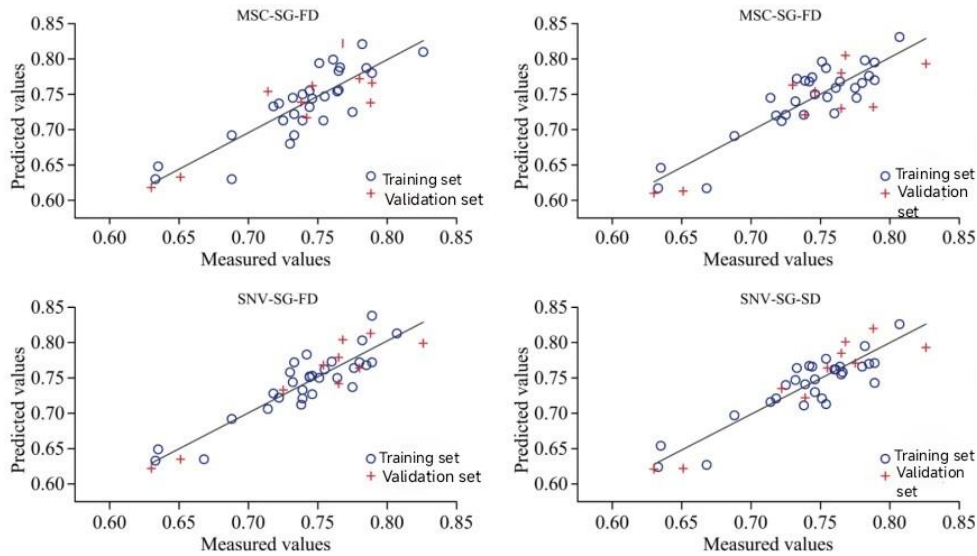
The correlation coefficients R and R, root mean square error RMSEC and RMSEP of model training set and validation set were used as the evaluation criteria of model effect. The larger the correlation coefficient, the smaller the root-mean-square error, the better the model effect. OCW preprocessing is the original feature spectrum selected to serve as a control against other preprocessing methods. The accuracy of estimating maize leaf physiological parameters by inversion using different models was compared and analyzed to determine the best modeling method.

### (3) Comparison of modeling methods for potentially diseased leaves

Fluorescence spectral inversion models with 5 single pretreatment methods and 4 combination methods were established for potentially diseased leaves. Correlation coefficients R and 2, root mean square error RMSEC and RMSEP of training set and prediction set in SPAD and Fv/Fm inversion models were obtained, and the results were shown in Table 4. The results showed that: in the inversion model of SPAD, the correlation coefficients of the training set of the five pretreatment methods were all above 0.937 2, and the root-mean-square error RMSEC is below 5.47. The training set correlation coefficient R of the four combined method models is above 0.9302, and the root-mean-square error RMSEC is below 3.56. The correlation between the measured values and the predicted values of the SPAD inversion models of the four combination methods is shown in Figure 3. In the Fv/Fm inversion model, the correlation coefficient R of the training set of the five single preprocessing method models is above 0.922 6, and the root mean square error RMSEC is below 7.45. The training set correlation coefficient R of the four combined methods is above 0.9447, and the root-mean-square error RMSEC is below 5.24. The correlation between the measured values and the predicted values of the Fv/Fm inversion models of the four combination methods is shown in Figure 4. The overall effect and performance of SPAD and Fv/Fm inversion models are better.



**Figure 3.** Correlation between measured and predicted values of SPAD inversion model for potentially diseased leaves



**Figure 4.** Correlation between measured and predicted values of Fv/Fm inversion model for potentially diseased leaves

**Table 4.** Comparison of modeling effects of different modeling methods for potentially diseased leaves

Modeling object	Degree of disease	Pretreatment method	$R_c$	RMSEC	$R_p$	RMSEP
SPAD	Potential disease	OCW	0.924 4	3.75	0.892 9	0.038 6
		MSC	0.937 2	4.28	0.921 3	0.025 6
		SNV	0.942 7	4.33	0.921 7	0.027 5
		SG	0.954 4	4.77	0.925 9	0.027 9
		FD	0.954 0	5.28	0.931 7	0.049 9
		SD	0.948 8	5.47	0.935 4	0.049 6
		MSC-SG-FD	0.930 2	3.56	0.926 3	0.028 7
		MSC-SG-SD	0.970 4	2.28	0.968 0	0.021 2
		SNV-SG-FD	0.985 2	1.59	0.975 4	0.015 0
		SNV-SG-SD	0.975 8	2.05	0.972 5	0.017 1
$F_v/F_m$	Potential disease	OCW	0.922 9	8.45	0.885 9	0.071 9
		MSC	0.927 3	7.42	0.920 7	0.063 0
		SNV	0.936 6	7.41	0.862 1	0.074 9
		SG	0.950 0	7.45	0.909 4	0.076 5
		FD	0.930 0	7.40	0.877 5	0.087 0
		SD	0.922 6	7.25	0.891 2	0.078 6
		MSC-SG-FD	0.944 7	5.24	0.927 5	0.036 1
		MSC-SG-SD	0.955 2	3.19	0.920 5	0.032 7
		SNV-SG-FD	0.976 8	2.85	0.961 5	0.023 0
		SNV-SG-SD	0.970 2	2.92	0.942 9	0.027 6

After establishing inversion models for potentially diseased leaves and comparing R and RMSEC of the model training set, SNV-SG-SD in SPAD inversion model had the best effect, with correlation coefficient R of 0.9852 and root mean square error RMSEC of 1.59. The SNV model had the worst effect, with R of 0.942.7 and RMSEC of 4.33. In the Fv/Fm inversion model, SNV-SG-FD has the best effect, its correlation coefficient R is 0.9768, and the RMSEC is 2.85. The FD model has the worst effect with R of 0.9300 and RMSEC of 7.40.

By comparing the predictions and results of different models, AI can help us build more appropriate prediction models. In the field of crop disease prediction, after calculating the predicted value, AI will further analyze the accuracy of the forecast results, and finally show the results and feasibility directly to the producers and give planting suggestions, including suitable crops and other information.

## **4. AGRICULTURAL AID RESULTS**

### **4.1. Precision Production - demand Reforecasting Market**

We often see or hear the news that farmers' agricultural products are unsalable, melons, fruits and vegetables are sold at a low price or rot in the fields. In fact, the reason is due to market supply and demand. At the same time, there will be “garlic you ruthless”, “ginger you army”, “beans you play” supply is less than the demand. In fact, if the data in the process of agricultural production can be summarized, it is not difficult to achieve “supply and demand balance” in order to produce reasonably.

For example, a farm in Anhui province produced a lot of grapes this year, but the local market demand was small. By increasing the demand Data collection found that the market demand for grapes in a certain place in Shandong is high, then the farm manager can contact the seller in Shandong as soon as possible to sell grapes to Shandong. In addition, farmers can collect consumer demand reports through the big data platform in advance, carry out market analysis, plan production in advance, reduce production risks, and help farmers make smarter decisions in farming.

### **4.2. Automated Production - Agricultural Environmental Monitoring**

Through agricultural big data, the production of crops can be sensed by detecting the production environment of crops through sensors. By collecting various index data in the growing environment of crops, and then placing the collected data in the localized or cloud data center, the historical data and real-time monitoring data of agricultural production can be analyzed, and the related monitoring ability of crop planting area, production schedule, agricultural product yield, weather conditions, temperature conditions, disaster intensity and soil moisture can be improved.

For example, the system monitors the insufficient soil moisture of the potatoes in greenhouse No. 1, then the soil moisture can be filled in time. For example, monitoring the light color of the peppers in greenhouse No.3, then the reason can be analyzed through the data. Imagine if in the growth process of crops, climate disasters can be avoided and scientific and effective prevention and control measures, planting methods can be effective guidance, then it will be followed by stable yield and even increase, from the source to improve agricultural production efficiency.

### **4.3. Supply Chain Tracking -- Achieve Traceability of Agricultural Products**

The agricultural big data technology platform can be used to track every process from field to table.

Big data is being used to improve all aspects, with agricultural producers, suppliers and transporters using iot sensor technology, scanning devices and analytical tools to monitor and collect relevant data from the supply chain. For example, the quality of agricultural products during production and transportation can be monitored in real time by GPS-enabled sensors, helping to prevent foodborne illness and reduce waste in the supply chain.

At the same time, agricultural "big data" helps to carry out monitoring and early warning of agricultural products, through in-depth mining and effective integration of agricultural production and circulation data scattered in the country's agricultural production areas, professional analysis and interpretation, for the production and circulation of agricultural products to provide efficient and high-quality information services, to improve the utilization of agricultural resources and circulation efficiency, from the source to ensure food safety.

## 5. CONCLUDING REMARKS

In conclusion, AI and big data have far-reaching and extensive significance in assisting agriculture. They have injected new impetus into the development of agricultural modernization by improving agricultural production efficiency and management level, optimizing logistics and inventory management, ensuring the quality of agricultural products, promoting market connectivity and enhancing farmers' skills. They are also gradually changing the traditional agricultural production methods and management models, and promoting the development of agriculture in a more efficient, greener and more sustainable direction.

## ACKNOWLEDGEMENTS

This paper was funded and supported by the 2024 School-level Innovation and Entrepreneurship Training Program of Wuhan Business University (No. 202411654192).

## REFERENCES

- [1] Lian Shiyou. Introduction to Artificial Intelligence Technology [M]. Xi 'an: Xidian University Press, 2007.
- [2] Bian Zhaoqi, Zhang Xuegong, Pattern Recognition [M]. Beijing: Tsinghua University Press, 2002.
- [3] Jiang Shuqin, Niu Huimin, Application Research of Artificial Intelligence in agricultural field under the background of Big Data [J]. Shanxi Agricultural Economics, 2019(19):77-78.
- [4] Sun Shulin, Ma Xinlu. Application of artificial intelligence in agricultural production [J]. Computer and Network, 2019, 46(08):46.
- [5] Ma Qianli. Analysis on the role and prospect of intelligent agriculture in agricultural production [J]. Southern Agricultural Machinery, 2019, 51(03):58.
- [6] Wang Yu, Zheng Zihui. Application direction and development path of artificial intelligence technology in agriculture [J]. Information and Communication Technology and Policy, 2019(06):29-31.
- [7] Chen Guifen, Li Jing, Chen Hang, et al. Research progress of artificial intelligence technology in agriculture in the era of big data [J]. Chinese Agricultural Abstracts--Agricultural Engineering, 2019, 31(01):12-16.
- [8] Li Liqin. Analysis on the application of artificial intelligence technology in modern agricultural production [J]. Agricultural technology and equipment, 2019(08):28-29.
- [9] Yu Qingze, Mao Weihui, Lu Xiuping. Analysis of practical application of artificial intelligence technology in modern agriculture based on foreign development trends. Henan Science and Technology, 2019(22):17-19.
- [10] Yang Yisheng, Xue Chunxu, Xu Ying, et al. Socio-economic characteristics, development logic and systematic interpretation of smart agriculture [J]. Journal of Jilin Agricultural University, 2021(2):146-152.
- [11] Zhang Hailan, Lei Guiping. Architecture design of smart agricultural greenhouse system based on Internet of Things [J]. Information Recording Materials, 2018, 19(2):63-64.
- [12] Ye Ziming, Dang Kaiwen, Sun Xinyu. Research on intelligent agricultural greenhouse system based on Internet of Things [J]. Guangdong Silkworm Industry, 2020, 54(2):15,17.
- [13] Zhang Chengnian. Research and implementation of smart agricultural greenhouse system based on Internet of Things [J]. Xinjiang Land Reclamation Science and Technology, 2016, 39(1):55-57.
- [14] Kong Fantao, Zhu Mengshuai, Sun Tan. Application analysis and suggestions of modern information technology in agriculture: Reflections on Internet enterprises' entry into agriculture [J]. Smart Agriculture, 2019(4):31-41.