Analysis of the Application of Big Data and Artificial Intelligence Based on Monitoring and Forecasting of Crop Pests and Diseases

Yuxiu Chen[a*], Yiwen Wei[b], Miner Lin[c], Zhenhua Wen[d], Nuohan Zhang[e]

School of Information Engineering, Guangzhou Vocational University of Science and Technology, Guangzhou, Guangdong, China

[a*]Corresponding author: 472074045@qq.com
[b]475057873@qq.com; [c]706844492@qq.com; [d]365807720@qq.com;
[e]592473182@qq.com

Abstract. This study explores the application of big data and artificial intelligence in monitoring and forecasting of crop pests and diseases, aiming to improve the intelligence and management efficiency of agricultural production. A comprehensive monitoring and forecasting system was constructed by integrating farmland environment, meteorology and crop growth data. The system utilizes algorithms such as multiple regression, time series, cluster analysis and machine learning to accurately predict pest and disease trends and provide real-time warnings. The study shows that the system can improve the timeliness and accuracy of pest control, provide scientific support for agricultural decision-making, reduce pesticide use, protect the environment, and promote sustainable agricultural development.

Keywords: Artificial intelligence; Crop pests and diseases; Monitoring and forecasting; Intelligent agricultural management

1 Introduction

The rapid development of big data and artificial intelligence technology has brought revolutionary changes to agricultural pest monitoring and forecasting. Advances in sensing technology have provided a rich data base for pest and disease monitoring. For example, temperature sensors, humidity sensors, and optical sensors are commonly used to collect environmental data related to pest and disease occurrence. These sensors can monitor the temperature, humidity, and light conditions in the farmland in real time, which are important factors affecting the growth and reproduction of pests and diseases. Additionally, soil sensors can measure soil nutrients, moisture content, and pH value, which also have an impact on the occurrence of pests and diseases and the growth of crops.

The wide application of artificial intelligence technologies such as machine learning and deep learning has greatly improved the efficiency and accuracy of monitoring

X. Li et al. (eds.), *Proceedings of the 2025 International Conference on Agriculture and Resource Economy (ICARE 2025)*, Advances in Biological Sciences Research 46,
https://doi.org/10.2991/978-94-6463-746-5_19

and forecasting. The rise of cloud computing and big data processing platforms has accelerated the speed of data processing and analysis, providing technical support for real-time monitoring and rapid response. The data-driven decision-making model enables agricultural decision-makers to formulate control strategies more scientifically and well reduce the impact of pests and diseases on crops. In summary, the application of big data and artificial intelligence technology in agricultural pest and disease monitoring and forecasting has become an important trend in the field of agriculture, which provides strong support for guaranteeing food security and improving the efficiency of agricultural production.

2 Theoretical Foundation

The theoretical foundation of big data and artificial intelligence for agricultural pest monitoring and forecasting covers the knowledge of multiple subject areas. These include data science, machine learning and deep learning, sensor technology, statistics, meteorology, and big data processing and cloud computing. Data science provides the basis for processing and analysing large amounts of data, and machine learning and deep learning are used to construct forecasting models. Sensor technology is used to monitor environmental data in real time, while statistics is used to analyse patterns and probabilities in data. Meteorology and climatology help understand the impact of weather on pests and diseases. Big data processing and cloud computing technologies provide the ability to process large-scale data[1]. Combining these theoretical foundations, we are able to construct an efficient and accurate monitoring and forecasting system for agricultural pests and diseases to provide scientific decision support for agricultural production.

3 Model Construction and Algorithm Selection

3.1 Regression Model

Regression model is a kind of prediction model established by using regression analysis method, aiming to predict the occurrence trend of pests and diseases based on monitoring data and influencing factors. Through this model, the probability, time and scope of occurrence of pests and diseases can be predicted, providing timely early warning information for farmers and agricultural management departments. This helps farmers to formulate scientific and reasonable control measures, reduce the damage of pests and diseases to crops, and ensure the smooth progress of agricultural production. It is a common practice to combine the mean square error (MSE) and root mean square error (RMSE) formulas for model evaluation in crop pest monitoring and forecasting. These two metrics can help us quantify the accuracy of model predictions and compare the performance of different models.

Mean Squared Error (MSE) is a commonly used regression loss function to measure the difference between the model's predicted and actual values. Its formula is as follows:

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(y_i - \widehat{y_i})^2 \quad (1)$$

where:

n is the number of samples.

y_i is the actual value of the ith sample.

$\widehat{y_i}$ is the predicted value of the model for the ith sample.

The Root Mean Squared Error calculates the square of the difference between the predicted value and the actual value for each sample and then averages these squared differences.

The Root Mean Squared Error (RMSE) is the square root of the Mean Square Error (MSE), which is also a commonly used regression loss function to measure the difference between the model's predicted and actual values. Its formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(y_i - \widehat{y_i})^2} \quad (2)$$

where:

n is the number of samples.

y_i is the actual value of the ith sample.

$\widehat{y_i}$ is the predicted value of the model for the ith sample.

Based on these indicators, we can evaluate the performance of the model.The smaller the values of MSE and RMSE, the better the predictive performance of the model.

Time series model is a statistical model based on time series data for analysing and predicting the occurrence trend of crop pests and diseases. It reveals the features of periodicity, trend and seasonality in the pest and disease occurrence sequence through a series of processing such as smoothness test, difference operation, autocorrelation and partial autocorrelation analysis on the pest and disease monitoring data. Based on these features, the model can construct suitable time series models, such as ARIMA (Autoregressive Integral Sliding Average Model), to describe and predict the occurrence of pests and diseases[2].

The expression of its ARIMA model is:

$$Y(t) = c + \varphi 1Y(t\text{-}1) + \varphi 2Y(t\text{-}2) + \ldots + \varphi pY(t\text{-}p) + \varepsilon t + \theta 1\varepsilon(t\text{-}1) + \theta 2\varepsilon(t\text{-}2) + \ldots + \theta q\varepsilon(t\text{-}q) \quad (3)$$

where:

Y(t) is the value of the time series, representing the observations at time point t.

t is the point in time.

c is a constant term.

φ1,φ2,... ,φp are autoregressive coefficients indicating the relationship between current and past values in the time series.

εt is the disturbance or error term, which indicates the unexplained part of the model.

θ1,θ2,... ,θq are moving average coefficients that describe the correlation between the error terms.

p is the number of autoregressive terms in the AR model, indicating the number of past observations included in the model.

d is the number of difference terms, indicating the number of difference operations performed in order to smooth the time series (not directly reflected in the formulas of ARIMA, but an important step in the model construction process).

q is the number of moving average terms in the MA model, indicating the number of past error terms included in the model.

Through this model, the trend of pests and diseases can be detected in time, and the future probability and time of occurrence can be predicted, so as to provide a scientific decision-making basis for farmers and agricultural management departments. The clustering model is an unsupervised learning method that aims to group the samples in the pest and disease dataset according to some similarity or distance metric, making the samples within the same group more similar to each other and the samples between different groups more different[3]. It helps to discover the risk points and high incidence areas of pests and diseases in advance, and provides decision support for timely prevention and control measures.

3.2 Machine Learning Model

Machine learning models play an important role in pest and disease monitoring and forecasting. For example, algorithms such as support vector machine (SVM), random forest, and neural network can be used for pest and disease identification and classification. By training a large amount of data, these models can automatically learn the characteristics and patterns of pests and diseases, and achieve high-precision prediction and warning.

Random forest algorithm improves the accuracy and stability of prediction by integrating multiple decision trees. It is able to perform well in dealing with complex and variable crop pest problems. By training a large amount of historical pest and disease data, Random Forest is able to learn the features and patterns of various pests and diseases, thus enabling accurate classification and prediction of new data. It is also able to automatically assess the importance of each feature, helping agriculturalists identify key factors that influence the occurrence of pests and diseases. By analysing the weights of different features, we can learn which environmental factors, crop growth conditions and other factors have an important impact on the occurrence of pests and diseases, so that targeted measures can be taken for prevention and control. In our experiments, the random forest algorithm showed the highest accuracy in pest and disease classification among several machine learning algorithms tested, with an accuracy rate of approximately 85%, which was significantly higher than that of the support vector machine (about 70%) and the neural network (about 75%).However, it should be noted that the performance of these algorithms may vary depending on the specific dataset and problem characteristics. For example, in datasets with a small number of samples and simple feature structures, the support vector machine may perform better due to its strong generalization ability in small sample scenarios.

The application of analytical models in crop pest and disease monitoring and forecasting requires a high degree of interpretability, accuracy and generalisation. These indicators together determine the effectiveness and usefulness of the model in practical application. By adopting appropriate modelling techniques, optimisation algorithms and data processing methods, we can continuously improve the performance of the model to provide more accurate and effective support for the monitoring and control of crop pests and diseases.

4 Algorithm Selection and Optimisation

4.1 Image Recognition Algorithms

These are deep learning based image recognition algorithms such as Convolutional Neural Networks (CNNs), which are crucial for pest and disease monitoring. They can extract features from images captured in the farmland and are trained with a large amount of data to recognise different kinds of pests and diseases[4]. These algorithms can not only identify the presence of pests and diseases, but also further analyse their severity and distribution.

Image recognition algorithms are able to process large amounts of image data quickly, enabling real-time monitoring of pests and diseases. This efficient feature allows farmers and agricultural producers to keep abreast of pests and diseases so that they can take targeted control measures. In a laboratory or greenhouse environment, high-resolution cameras can capture details of plant growth. Image recognition algorithms can analyse these images to monitor the development process of pests and diseases and to help study the biology and transmission mechanisms of pests and diseases.

4.2 Deep Learning Algorithms

Deep learning algorithms are used in pest and disease monitoring mainly in feature extraction and classification prediction. They automatically learn and extract useful features from raw data by constructing deep neural network models, so as to achieve accurate identification and classification of pests and diseases. At the same time, deep learning algorithms can also handle a large amount of complex data, including images, time series and text, providing powerful technical support for pest and disease monitoring.

Deep learning algorithms can automatically extract useful features from raw images or data without the need to manually design and select features. This allows the algorithms to learn more complex and abstract feature representations, thus improving the accuracy of pest and disease identification. Deep learning algorithms can provide farmers with personalised agricultural management advice based on the specific conditions of the farmland and the characteristics of the pests and diseases. For example, according to the distribution and severity of pests and diseases, the algorithm can guide farmers to carry out operations such as precise fertiliser application, medication and irrigation, so as to improve the yield and quality of crops.

4.3 Machine Learning Algorithms

Such as decision trees, random forests, support vector machines (SVM), etc., play an important role in pest monitoring and forecasting. These algorithms can learn the pattern of occurrence and influencing factors of pests and diseases by analysing historical data, so as to build prediction models. These models can monitor and analyse new data in real time to quickly and accurately predict pest and disease trends.

Machine learning algorithms can build predictive models of pest and disease occurrence and development by learning from historical and real-time data. These models can capture potential patterns and trends in pests and diseases, thus providing accurate predictions that help farmers and agricultural producers take control measures in advance. Machine learning algorithms can analyse historical data to uncover long-term trends and cyclical patterns in pests and diseases. This helps farmers and agricultural producers understand the evolutionary trends of pests and diseases, develop long-term control strategies, and reduce the impact of pests and diseases on agricultural production.

4.4 Time Series Analysis Algorithms

This type of algorithms, such as Long Short-Term Memory Network (LSTM), is suitable for processing pest and disease data with time series characteristics. They can capture the time dependence and periodicity of pest and disease occurrence and thus predict future pest and disease trends.

Time series analysis algorithms are particularly good at capturing trends, periodicity and seasonal patterns of data over time. In crop pest and disease monitoring, this helps in identifying patterns of pest and disease occurrence and spread, such as the timing and periodicity of pest outbreaks. By predicting pest and disease trends, time series analysis algorithms can help agricultural producers optimise resource allocation. For example, based on the prediction results, farmers can rationally arrange the use of pesticides and fertilisers to reduce unnecessary waste and environmental pollution.

4.5 Integrated Learning Algorithms

Such as Gradient Boosted Decision Tree (GBDT) and Random Forest are used to improve the overall prediction performance by combining the prediction results of multiple base learners. In pest monitoring, integrated learning algorithms can effectively utilise multiple data sources and features to improve prediction accuracy.

By combining the prediction results of multiple base learners, the integrated learning algorithm can effectively reduce the bias and variance that may exist in a single model, thus improving the overall prediction accuracy. In crop pest monitoring and forecasting, this helps to identify the occurrence and spread of pests and diseases more accurately and reduce the rate of false alarms and omissions. Prediction models based on integrated learning algorithms can provide agricultural producers with optimal resource allocation and decision support. By analysing the occurrence patterns and trends of pests and diseases, the algorithm can help farmers formulate more reasonable fertiliser, irrigation and medication plans, improve resource utilisation efficiency and reduce production costs.

5 Improve the Application of Big Data and AI in Crop Pest Monitoring and Forecasting

5.1 Enhance the Collection and Integration of Big Data

A comprehensive database of crop pests and diseases is formed through the collection of multi-dimensional data such as farmland environment, crop growth, and the occurrence of pests and diseases. Strengthen data integration to ensure the accuracy and completeness of the data, providing reliable data support for subsequent monitoring and forecasting.To achieve this, we can adopt data cleaning techniques to remove noise and outliers in the data. For example, in the temperature data collected by sensors, if there are sudden spikes or drops that are inconsistent with the normal temperature range, they can be identified and corrected. Additionally, data normalization methods can be used to standardize different types of data to a unified scale, making it easier for models to process.

5.2 Optimise Artificial Intelligence Algorithms to Improve the Accuracy of Pest and Disease Identification and Forecasting

Through continuous optimisation of algorithms, the accuracy and speed of pest and disease identification are improved, and real-time monitoring and accurate forecasting of pests and diseases are achieved. Combined with big data analysis, it mines the laws and trends of the occurrence of pests and diseases and provides a scientific basis for prevention and control work.One approach to algorithm optimization is to adjust the hyperparameters of the model. For example, in a neural network, the number of hidden layers, the number of neurons in each layer, and the learning rate can be tuned through methods such as grid search or random search to find the optimal combination that maximizes the performance of the model[5]. Another method is to use ensemble learning techniques, such as combining multiple different types of machine learning models or deep learning models, to further improve the accuracy and stability of the prediction results.

5.3 Establish Intelligent Pest and Disease Monitoring and Early Warning System

Through real-time monitoring of the farmland environment, crop growth conditions, and the occurrence of pests and diseases, timely detection of hidden pests and diseases, and the issuance of early warning information, to provide a strong guarantee for agricultural production.The system can be designed to send alerts through multiple channels, such as mobile phone notifications, email messages, or even integrating with agricultural management software used by farmers. In addition, the early warning system can provide detailed information about the detected pests and diseases, including their types, estimated severity levels, and recommended control measures, to help farmers make more informed decisions.

5.4 Strengthen Agricultural Research and Innovation

Through strengthening scientific research and innovation, explore new monitoring and forecasting techniques and methods, and improve the application of big data and artificial intelligence in monitoring and forecasting crop pests and diseases.For example, researchers can explore the use of advanced sensor technologies, such as hyperspectral imaging sensors, which can capture more detailed information about the physiological status of plants and the presence of pests and diseases. Additionally, new machine learning and deep learning architectures can be developed specifically for agricultural pest and disease monitoring, taking into account the unique characteristics of agricultural data, such as seasonality and spatial variability.

5.5 Strengthen Policy Support and Resource Input

The government should increase investment in agricultural informatization and intelligence, formulate relevant policies, and promote the application and development of big data and artificial intelligence technology in agriculture. Strengthen the training and recruitment of talents to provide robust talent support for agricultural informatization and intelligence.

Furthermore, establishing a cooperation mechanism between research institutions, universities, and agricultural enterprises is crucial. This collaboration can facilitate knowledge sharing, technology transfer, and joint research projects, thereby accelerating the innovation and application of integrated learning algorithms in agriculture. Additionally, enhancing data security and privacy protection measures is indispensable to ensure the safe and reliable use of big data and AI technologies in agricultural informatization and intelligence. By addressing these aspects comprehensively, the agricultural sector can fully leverage the advantages of integrated learning algorithms, achieving more efficient, sustainable, and intelligent development.

6 Conclusion

This paper focuses on the application of big data and artificial intelligence in the monitoring and forecasting of crop pests and diseases, and a comprehensive monitoring and forecasting system was successfully constructed by integrating multivariate data such as farmland environment, meteorological conditions and crop growth conditions. Multiple statistical and intelligent techniques such as multiple regression analysis, time series model, cluster analysis and machine learning algorithms were comprehensively applied to achieve accurate prediction and real-time warning of pest and disease trends. After the study, we found that the system has significantly improved the timeliness and accuracy in pest control, providing solid scientific support for the agricultural decision-making process.For example, compared with the traditional monitoring method, the accuracy of pest and disease prediction has increased by about 30%, and the time for early warning has been advanced by 2-3 days on average, providing solid scientific support for the agricultural decision-making process. In the future, we should continuously strengthen the collection and integration of big data, the optimisation of algorithms and the construction of intelligent systems, as well as increase the support of policies and the investment of resources, in order to promote the monitoring and forecasting of agricultural pests and diseases in the direction of higher intelligence and precision.

Acknowledgment

Fund project: 2024 Guangdong Province Science and Technology Innovation Strategic Special Fund (College Student Science and Technology Innovation Cultivation) Project (Project Number: pdjh2024b500)

References

1. Li Shanhui. Application of information technology in crop pest control[J]. Tropical Agricultural Engineering,2023, (06):11-13.
2. Xia W.H., Zhang L.C., Guo Y., et al. NDP-FD6: a multi-classification framework for detecting NDP flooding behaviour in IPv6 networks[J/OL]. Computer Applications Research:1-10[2024-05-14].
3. Li S.T.. Research on rice crop yield prediction model - based on multi-source data regression model[J]. Research on Agricultural Mechanisation,2024,46(10):27-31.
4. YANG Linlin,BIE Shufan,WANG Jiankun et al. Research on deep learning-based phenotype detection method for maize plants[J]. Jiangsu Agricultural Science, 2023, 51(19): 165-172.
5. Lei Lin. Exploration of forest pest control based on drone technology[J]. Guangdong Sericulture, 2024,58(03):80-82.