Forecast and Control of China's Grain Yield Based on Big Data and BP Neural Network in the Context of Sustainable Agriculture

Haoran Wang[1, a], Zhengxian Chen[2*]

[1]Computer Science Department New York University New York, NY 10012, USA
[2]Department of Mechanical Engineering Columbia University New York, NY 10027, USA
hw2650@nyu.edu
[*]Corresponding author: zc2545@columbia.edu

Abstract
In actual agricultural production, many factors affect grain yield. Among them, non-linear factors such as climate, capital, land directly affect farmers' enthusiasm for production and significantly impact production forecasts. Based on the BP neural network, this paper considers the influence of various non-linear factors on grain yield. It then establishes models through machine learning to realize grain yield forecasting and management by using different economic factors. It provides a helpful reference for formulating macroeconomic policies, sustainable agricultural policies, and regulating grain yield.

Keywords: *Agricultural Economics, Machine Learning, Neural Network, Grain Yield, Sustainable Agriculture, Grain Yield Forecasting*

1. INTRODUCTION

Grain yield is a crucial indicator for developing countries. In Lester R. Brown's book, he raised the famous question of "Who will feed China?" which has aroused the world's attention to China's food security issues. Today, China feeds itself and makes excellent contributions to solving world food security and reducing global hunger. For a developing country with a large population like China, it is imperative to track grain yield and adjust macroeconomic policies at any time according to the situation of grain yield. At the same time, in the context of sustainable agriculture, it also needs to monitor and predict the status of water resources and forestry resources to ensure that they can meet sustainable agriculture needs.

Grain yield is affected by a combination of natural and unnatural factors. So far, researchers have mainly classified these factors into three categories: social factors, meteorological factors, and accidental factors [1]. However, traditional economic models usually contain little prior information about functions and have significant potential heterogeneity among different observation units. Those models often have multiple outputs and only focus on fitting historical data [2].

In actual production, the factors that affect grain yield are complex, and there are interactions and non-linear relationships between grain yield and many influencing factors. Traditional forecasting methods have significant limitations in solving such highly non-linear problems. Although conventional methods allow us to deal with these problems, machine learning methods increase the flexibility of data and function forms and processing efficiency and open up other analytical approaches. This paper divides the data and innovatively proposes the classification standard: data available for prediction and data unavailable for prediction. Finally, through the model based on the BP neural network, the precise control of grain yield and the precise environment protection in line with the concept of sustainable agriculture are realized.

2. RELEVANT RESEARCH

2.1. The change of rural land system and the impact of other economic factors on grain yield

In recent years, with China's economic development and changes in the land system, many scholars have realized the critical impact of the land system on

S. A. R. Khan et al. (Eds.): ICMSSE 2022, AHCS 12, pp. 58–65, 2023.
https://doi.org/10.2991/978-94-6463-056-5_10

agricultural production [3]. China's agricultural land system can be roughly divided into the following periods: before the completion of the land reform (before 1958), the people's commune period (1958-1977), the household responsibility system (1978-2003), and the Administration of Circulation of Rural Land Contracted Management Right (2004-). The land system in different periods has had an enormous impact on agriculture. Determining, registering, and certifying land ownership has affected farmers' enthusiasm for production and the area of arable land. Since the long-term data has less reference and the possibility of policy regression is minor, all the data studied in this paper have been collected since 1999, covering the household responsibility system (1978-2003) and the Administration of Circulation of Rural Land Contracted Management Right (2004-). This paper uses the neural network to consider the impact of the land system on agricultural production to some extent. At the same time, the land system and other economic factors are also intuitively reflected through data, such as rural electricity consumption, the degree of agricultural mechanization, and so on.

2.2. Analysis of Influencing Factors of Grain Yield

Regarding the influencing factors of grain yield, many related papers mainly concentrated on several categories, such as social factors, meteorological factors, random errors, etc. On this basis, this paper innovatively divides the factors affecting grain yield and proposes the classification criteria for data available for prediction and data unavailable for prediction.

2.3. Research on the use of neural networks in grain yield forecasting

With the popularity of neural networks in recent years, many scholars have used neural networks in grain yield forecasting. Some scholars have applied machine learning in agricultural production forecasting [4]. However, few papers put forward management methods in combination with specific rural production and agricultural economics. At the same time, papers on the use of neural networks for sustainable agricultural regulation are rare. This paper will explore this part in detail.

3. PRINCIPLES OF GRAIN YIELD FORECAST AND CONTROL BASED ON BP NEURAL NETWORK

Agricultural production is equivalent to a neural network, as shown in Figure 1. Each influencing factor is analogous to a neuron in this neural network, and each influencing factor also affects the other. The fluctuation of one neuron will cause the linkage of other neurons. For example, if the area of arable land increases, the affected area, irrigated area, and even rural electricity consumption will change to a certain extent. This paper analyzes the factors that affect agricultural production, such as arable land scale, the affected area, the irrigation area, the total power of agricultural machinery, the consumption of chemical fertilizers, precipitation, average temperature, etc. Thus, this paper inputs characteristic value variables to describe the output variable value of agricultural grain yield. Plus, it uses machine learning to establish a forecast and analysis model for grain yield. Therefore, it realizes pre-production forecasting, in-production control, and post-production analysis of grain yield and sustainable agricultural parameters.

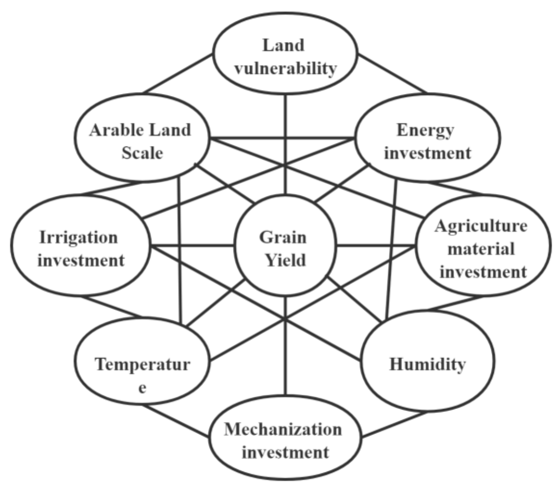

Figure 1. Neural network diagram

3.1. Grain yield neural network learning model

There is a linear or non-linear relationship between many agricultural production factors mentioned in the previous paper and grain yield. The mapping table of feature value variables and output value target variables is sorted out, and the neural network learning model is built, as shown in Figure 2.

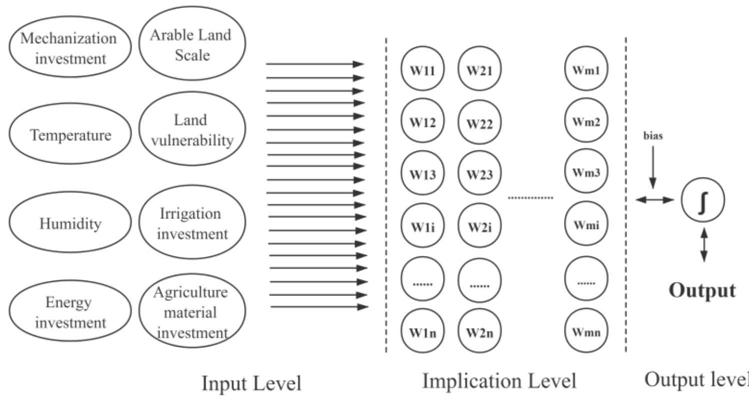

Figure 2. The neural network learning model

3.2. Data classification and variables table

This paper sorts out the data types. Due to space and time constraints, it selects eight representative data sets to construct the neural network's independent variables. The independent variables and dependent variables are shown in the following table.

TABLE I. VARIABLES TABLE

Sort	Variables	Indicators	Basis and expectation effect
Grain	Grain Yield	Total Grain Yield/10^4t	Use total yield instead of per capita grain yield since this paper focuses on the changing law of total grain yield, and therefore each explanatory variable corresponds to the total
Land	Arable Land Scale	Total Crop Area/10^4h ㎡	Arable land is the primary carrier of grain yield [5]. It assumes that the more Arable land area, the higher the food output.
	Land vulnerability	Affected area/10^3h m2	The affected area reflects climate change, farmland infrastructure construction, etc [6]; therefore, the expected effect is negative
Capital	Irrigation investment	Irrigated area/10^3hm2	Irrigation development is the main influencing factor of crops; therefore, the expected effect is positive [7].
	Agriculture material investment	Consumption of chemical fertilizers/10^4t	The input of agricultural materials investment represented by the number of chemical fertilizers can improve the soil fertility, which is the essential material input in current grain yield, and can promote the increase of grain yield; therefore, the expected effect is positive
	Mechanization investment	Total agricultural machinery power/10^4kW	Agricultural mechanization can reduce land abandonment, complete operational efficiencies, and benefits that cannot be achieved by human and animal power, and help increase grain production; therefore,

			the expected effect is positive [8].
	Energy investment	Rural electricity power consumption /10^4kW	Electricity is a modern energy source. The amount of electricity used in rural areas indicates the productivity of the rural regions. In agricultural production, electricity replaces human, animal power, coal, diesel, gasoline, etc., further liberating and improving productivity.
Climate	Temperature	Mean temperature (℃)	Temperature directly affects photosynthesis rate and respiration rate, the two main processes that determine grain productivity [9].
	Humidity	Annual precipitation(mm)	Appropriate precipitation can promote grain yield, and continuous rain will reduce the photosynthetic capacity of crops, reduce nutrient accumulation, reduce gain and quality. Being in a high humidity environment for a long time is conducive to the breeding of diseases. It will induce the spread of various diseases of autumn crops, which will seriously affect the expected growth of roots and be harmful to the growth of crops [10].

This table includes all the variables in the paper. At the same time, following the concept of neural networks, this paper proposes the ideas of data available for prediction and data unavailable for prediction. Available forecast data means that accurate data can be obtained in advance in production and appropriately used to forecast yield data. Relatively unavailable forecast data indicates that the information is not entirely suitable for forecasting output value; accurate data cannot be obtained before production or has little to do with grain yield. The semi-available forecast data is in between: part of the data can be obtained or used for forecasting under certain circumstances. The specific division is shown in the following table.

TABLE II. DATA CLASSIFICATION

Data available for prediction	**Date may not be used for prediction**	**Data can be used for prediction in some condition**
Arable Land Scale, Irrigation Investment, Mechanization Investment, agricultural population	GDP of the year, disposable income	Agriculture material investment, Land vulnerability, Temperature, Humidity

According to the data division, this paper selects some data as independent variables to ensure the scientificity of data prediction to the greatest extent.

3.3. Forecast, control, and analysis process of grain yield and sustainable agriculture indicators based on BP neural network

First, this paper compiles the mapping relationship table of feature value variables and outputs value target variables according to the neural network construction process, then sorts out the direct relationship between the independent variable and the dependent variable, as shown in the following table.

TABLE III. MAPPING TABLE

Input Variable	Output Variable	Cause Analysis of Difference
Arable Land Scale, Land vulnerability, Irrigation Investment, Agriculture material investment, mechanization Investment, Temperature, Humidity	Grain Yield	Some of the semi-available data are from the same period last year, leading to some variance in actual forecasts. At the same time, some errors in statistics will also lead to the decline of model accuracy.

In the actual forecast, some semi-available data, such as average temperature and precipitation, are predicted using the value of the previous year, which will cause specific errors. At the same time, the Total agricultural machinery power will also have certain changes during the forecast period, which will have a specific impact on the final result, but it is generally controllable. The overall process is shown in the figure below:

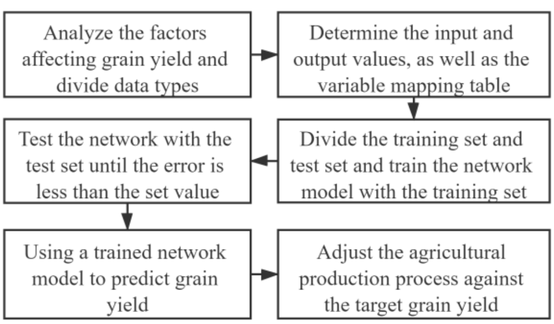

Figure 3. Neural network prediction flow chart

At the same time, the neural network will be well used in sustainable agriculture. The general process is shown in the figure below,

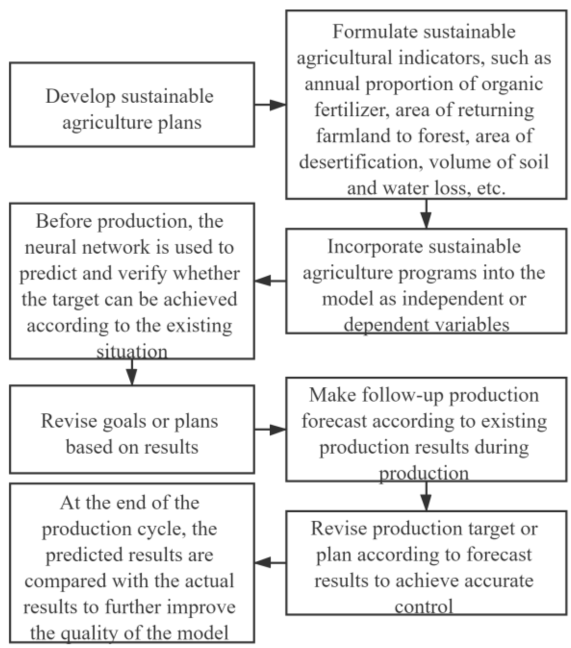

Figure 4. Flow chart of sustainable agriculture applications

Before agricultural production, agricultural managers can make plans for the production cycle and formulate quantitative indicators. At the same time, according to the indicator type, it is brought into the input or output data of the neural network model verification group.

For example, if the goal of returning farmland to forests is formulated, its influence will be reflected in the change in the area of arable land. Researchers or managers can first use the original arable land area to make predictions to obtain grain yield A and then use the arable land area after returning farmland to forests to predict and get grain yield B.

They could compare the two and make predictions again by adjusting other indicators, such as the amount of fertilizer used and the degree of agricultural mechanization represented by agricultural machinery power. Then what they need is to make sure to formulate the most scientific sustainable agriculture plan before production. Thus, under the guidance of this method, farmers can achieve not only sustainable development but also minimize the negative impact on agricultural output.

At the same time, the use of models in production can achieve precise management. The first is to regulate and control grain yield and forecast the total yield of the production cycle based on the currently completed production process. In this way, timely control of unfavorable factors can ensure that the predetermined outcome is reached at the end of the production cycle.

At the end of production, the previously predicted values are compared with the actual values to improve the neural network further.

4. TRAINING AND TEST RESULTS OF GRAIN YIELD PREDICTION AND SUSTAINABLE AGRICULTURE CONTROL MODEL BASED ON BP NEURAL NETWORK

In the context of big data, this paper extracts a large amount of China's agricultural statistics. The final input data is shown in the following table (2010, Hebei). The units in the table have been mentioned above.

TABLE IV. DATA FORMAT (SAMPLE)

Input Data	Arable Land Scale	Land vulnerability	Irrigation investment	Agriculture material investment
	8260.75	1527.4	4548.01	322.9
	Mechanization Investment	Temperature	Humidity	Others(Electricity Consumption)
	10151.3	10.1	10984	511.8
Output Data	Grain Yield			
	2975.9			

After the training, test with the test set, compare the prediction value and actual value.

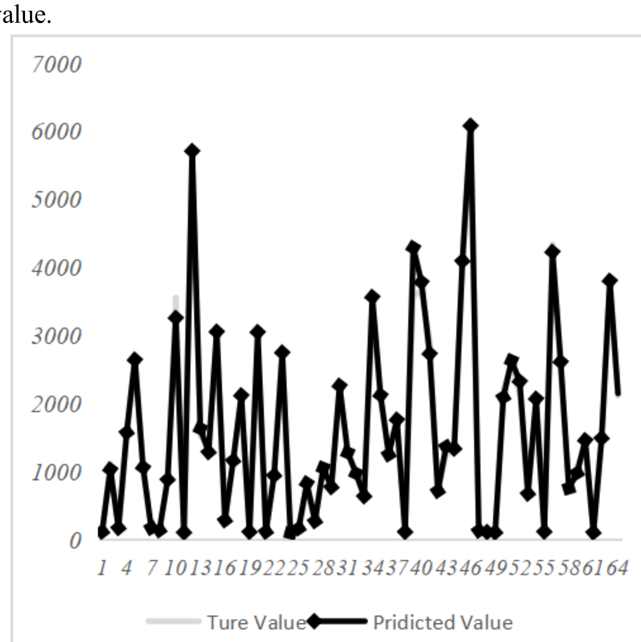

Figure 5. Comparison of actual and predicted values

The regression coefficient and Output chart is shown in figure x,

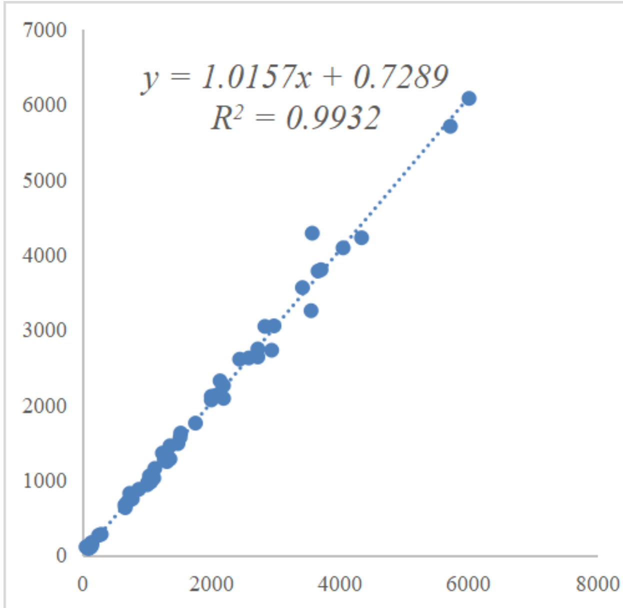

Figure 6. Regression coefficient

The coefficient of determination is 0.9932, very close to 1, which means the training results are almost ideal [11].

5. CONCLUSION

In the context of sustainable agriculture, this paper proposes a grain yield prediction model based on BP neural network. It analyzes how this model can be used in sustainable agricultural production. At the same time, the model considers the impact of many non-linear factors affecting grain yield. In the context of sustainable agriculture, the prediction results of grain yield obtained by the model proposed in this article have a helpful reference for formulating macroeconomic policies, sustainable agricultural policies, and regulating agricultural output value. While using in practical production, the model needs to consider some conditions. This model still has much room for improvement. However, due to paper space limitations, this paper does not elaborate on the impact of various independent variables on grain yield. Future research directions will also focus on improving the accuracy and scientificity of the model and studying the influence degree of each independent variable on grain yield.

REFERENCES

[1] GUO Liang, ANDREAS Wilkes, YU Hai-Ying, XU Jian-Chu. Analysis of Factors Influencing Yield Variability of Major Crops in China[J]. Plant Diversity and Resources, 2013, 35(4): 513-521. (in Chinese with English abstract)

[2] Storm H, Baylis K, Heckelei T. Machine learning in agricultural and applied economics[J]. European Review of Agricultural Economics, 2020, 47(3): 849-892.

[3] Research Group of the Shaanxi Survey Team of the National Bureau of Statistics. An Empirical Analysis of the Impact of Rural Land System Changes on Grain Production[J]. Survey World, 2015(4): 5.

[4] Cao J, Zhang Z, Tao F, et al. Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches[J]. Agricultural and Forest Meteorology, 2021, 297: 108275.

[5] Godfray H C J, Beddington J R, Haddad L R, et al. Food security: The challenge of feeding 9 billion people[J]. Science, 2010, 327(5967): 812—818.

[6] Simelton E, Fraser E D G, Termansen M, et al. The socioeconomics of food crop production and climate change vulnerability: A global-scale quantitative analysis of how grain crops are sensitive to drought[J]. Food Security, 2012, 4(2): 163—179.

[7] Deng Dacai. Food security: Arable land, trade, technology and conditions[J]. Research on Financial and Economic Issues, 2010(2): 3—7. (in Chinese with English abstract)

[8] Gu Lemin. Relative analysis of China's grain yield and influence factors based on criterion of least absolute deviation[J]. Transactions of the Chinese Society of

[9] Wheeler T R, Craufurd P Q, Ellis R H, et al. Temperature variability and the yield of annual crops[J]. Agriculture, Ecosystems & Environment, 2000, 82(1-3): 159-167.

[10] Holder R, Cockshull K E. Effects of humidity on the growth and yield of glasshouse tomatoes[J]. Journal of Horticultural Science, 1990, 65(1): 31-39.

[11] Junling Deng, Electricity cost management based on BP neural network under carbon neutral background, Commercial Accounting, 82-86, 2021(12).