ISSN: 3064-9846
Research Article
Journal of Agricultural, Earth and Environmental Sciences

Wheat Yield Prediction in Shandong Province Based on Deep Learning

Liu Mei Ling[1], Li Xin[1]* & Guo Bojian[2]

[1]Northeast Forestry University Harbin, China
[2]Kyrgyz State Technical University, Bishkek, Kyrgyz Republic

*Corresponding author: Li Xin, Northeast Forestry University Harbin, China.

Submitted: 11 February 2026 **Accepted:** 19 February 2026 **Published:** 26 February 2026

Citation: Liu, M. L., Li, X., & Guo, B. (2026). Wheat Yield Prediction in Shandong Province Based on Deep Learning. Journal of Agricultural, Earth and Environmental Sciences, 5(1), 01-05.

Abstract

Background: *Wheat is the most crucial food crop in Shandong Province and plays a vital role in ensuring China's food security. Accurate yield prediction is essential for optimizing agricultural management, supporting policy-making, and addressing challenges posed by climate change.*

Objective: *This study aims to develop an attention-based Long Short-Term Memory (LSTM-Attention) model to accurately predict wheat yield in Shandong Province by leveraging long-term meteorological and yield data.*

Methods: *A comprehensive 73-year dataset (1952–2024) was used, containing historical yield, annual average temperature, annual precipitation, and annual sunshine hours. After rigorous data preprocessing and correlation analysis, a multivariate time series prediction model was constructed. The model incorporates LSTM layers to capture long-term dependencies and an integrated attention mechanism to dynamically weight the importance of different time steps and meteorological variables.*

Results: *The proposed LSTM-Attention model demonstrated superior performance, achieving a Root Mean Square Error (RMSE) of 287,000 tons, a Mean Absolute Error (MAE) of 224,000 tons, and a coefficient of determination (R^2) of 0.97 on the test set. It significantly outperformed traditional models such as ARIMA, Support Vector Regression (SVR), and Random Forest. Attention weight analysis revealed that annual precipitation is the most influential meteorological factor affecting wheat yield in Shandong.*

Conclusion: *This research provides a high-precision and interpretable yield prediction tool for wheat production in Shandong. It also offers a novel deep learning framework for crop modeling in smart agriculture, with substantial theoretical value and practical application potential.*

Keywords: Wheat Yield Prediction, Deep Learning, Long Short-Term Memory (LSTM), Attention Mechanism, Time Series Analysis, Shandong Province, Meteorological Factors.

Introduction

Food security is an eternal issue concerning the national economy and people's livelihood. As China's second largest wheat-producing area, Shandong Province's wheat production ranks among the top in the country all year round. Its yield directly affects the stability of the country's food supply [1]. However, wheat production is a complex ecological process, which is affected by multiple factors such as varieties, management measures, soil conditions and meteorological factors. Among them, meteorological factors such as temperature, precipitation and sunshine have become the main source of uncertainty leading to yield fluctuations due to their dramatic interannual fluctuations and difficulty in human control [2]. In recent years, extreme weather events have occurred frequently under the background of global climate change, further exacerbating the instability of agricultural production [3]. Therefore, developing a prediction model that can accurately predict wheat yield, especially one that can quantify the impact of meteorological conditions, is of extremely important practical significance for actively adapting to climate change and ensuring food security.

Traditional crop yield prediction methods mainly include mechanistic models based on physiological and ecological processes (such as DSSAT and APSIM) and statistical models based on historical data [4]. Although mechanistic models have clear physical interpretations, they usually require a large number of detailed input parameters (such as soil properties and detailed field management data). These parameters are often difficult to obtain at a large regional scale, limiting their widespread application [5]. Statistical models, such as multivariate linear regression (MLR) and time series models (ARIMA), have relatively loose data requirements, but are difficult to effectively capture the complex nonlinear and non-stationary relationship between meteorology and yield [6].

With the rapid development of artificial intelligence technology, machine learning (ML) and deep learning (DL) methods have provided powerful tools for dealing with high-dimensional, nonlinear agricultural forecasting problems [7]. In particular, the long short-term memory network (LSTM) has shown great advantages in various time series forecasting tasks due to its excellent ability to handle long-term sequence dependencies. It has been successfully applied to fields such as stock forecasting, power load forecasting, and crop yield forecasting [8-10]. Compared with traditional models, LSTM can automatically learn dynamic patterns in time series without the need for complex feature engineering. However, when dealing with multivariate inputs, the standard LSTM model usually treats all historical information equally and lacks the ability to identify key time points and key driving factors [11].

To overcome this limitation, the attention mechanism was introduced [12]. It enables the model to dynamically focus on the historical information most relevant to the current prediction, thereby giving the model stronger interpretability and higher prediction accuracy [13]. This "LSTM-Attention" hybrid architecture has achieved great success in the field of natural language processing and has gradually been introduced into the field of spatiotemporal data prediction in recent years [14, 15].

This study aims to fill the current research gap and construct an LSTM model (LSTM-Attention) based on 73 years of long-term series data from Shandong Province (1952-2024) to predict wheat yield in Shandong Province. The main objectives of this study are: (1) to systematically analyze the evolution of wheat yield and key meteorological factors in Shandong Province during the historical period and their relationship; (2) to construct and train the LSTM-Attention prediction model and verify its superiority; (3) to use attention weights to analyze the relative contribution of different meteorological factors to yield formation and enhance the interpretability of the model; and (4) to provide a data-driven scientific basis for risk management and policy formulation of wheat production in Shandong Province.

Data And Methods

Study Area And Data Sources

The research area for this study is Shandong Province, People's Republic of China. The selected data are all provincial-level annual data, spanning the period from 1952 to 2024, with a total of 73 samples.

Wheat Production Data: Data are sourced from the Statistical Yearbook and Agricultural Yearbook published by the Shandong Provincial Bureau of Statistics, as well as the National Bureau of Statistics database. Data are expressed in 10^4 tons.

Meteorological Data: including annual average temperature (°C), annual precipitation (mm), and annual sunshine hours (hours). The original data comes from the China Meteorological Data Network (http://data.cma.cn/) and the Shandong Provincial Meteorological Bureau, and has undergone strict quality control and homogenization.

All data were summarized and organized in an Excel file, which formed the basis for the analysis of this study. The raw data are shown in Table 1.

Table 1: Wheat production and meteorological data for Shandong Province, 1952-2024

years	Annual average temperature (°C)	Annual precipitation (mm)	Annual sunshine hours (hours)	Wheat production (10,000 tons)
1952	12.5	685	2405	326.7
1953	11.9	832	2198	272.8
1954	12.3	965	2083	357.5
1955	12.4	598	2490	312
1956	13	901	2150	368.5
1957	11.6	642	2412	356.6
1958	11.8	553	2567	325
1959	12.7	518	2610	280
1960	13.3	601	2483	274
1961	12.9	635	2437	131.5
1962	13.6	787	2289	200.1
1963	12.8	892	2115	267.5
1964	12.4	1038	1978	320.5
1965	12.5	527	2571	341.5
1966	13.2	562	2519	343

1967	13.2	703	2380	334.5
1968	12.7	745	2316	307
1969	12.9	810	2204	352.5
1970	12.8	698	2372	319.5
1971	12.4	623	2465	441.5
1972	12.5	502	2603	547
1973	12.4	776	2291	543
1974	12.9	854	2178	545
1975	12.5	663	2420	656
1976	13.3	721	2275	811.5
1977	12.2	689	2358	606.5
1978	13	541	2592	803.5
1979	13.2	718	2337	957
1980	12.9	794	2219	766
1981	12.3	587	2501	870
1982	12.8	873	2180	824
1983	13.2	756	2324	1200
1984	13.3	828	2193	1278.5
1985	12.3	903	2137	1496.1
1986	12.5	596	2486	1562.4
1987	13	538	2552	1474.1
1988	12.9	612	2440	1390.8
1989	13.3	492	2638	1487.5
1990	13.3	734	2319	1612.1
1991	12.7	797	2235	1889.4
1992	12.8	573	2521	1878.3
1993	12.8	658	2417	1936
1994	14.2	601	2498	June 1936
1995	13.3	711	2362	2060.7
1996	12.8	862	2184	2052.74
1997	13.8	515	2617	2241.28
1998	14.3	796	2260	2024.48
1999	14.2	532	2563	2117.7
2000	14.1	650	2370	1860.04
2001	13.8	569	2512	1655.15
2002	13.8	589	2487	1547.06
2003	14.1	845	2239	1565.03
2004	13.3	648	2443	1584.5
2005	13.9	820	2247.4	1800.53
2006	14.1	570.1	2275.6	2012.96
2007	14.2	773	2349.8	2007.62
2008	14.6	698	2204.2	2038.84
2009	13.6	674.5	2309.2	2084.62
2010	13.2	690	2247.1	2108.79
2011	13.1	737.1	2200.1	2148.23
2012	13.3	615.2	2272.1	2219.68
2013	13.7	687.9	2310.8	2264.32
2014	14.4	516.8	2208.8	2325.57
2015	14.1	596.9	2169.9	2391.69

2016	14.5	679.5	2219.8	2490.11
2017	14.8	634.7	2334.7	2495.11
2018	14.6	790.1	2326.2	2471.68
2019	14.6	565.5	2235.6	2552.92
2020	14.2	814	2485.4	2568.85
2021	14.5	988.2	2196.6	2636.66
2022	14.5	861.5	2304.5	2641.19
2023	15	671.5	2445.9	2673.76
2024	15.1	859.4	2450.4	2716.5

Data Analysis and Preprocessing

Basic descriptive statistics were performed on the four variables, including mean, standard deviation, minimum, and maximum values, to understand the central tendency, dispersion, and distribution range of the data. The results are shown in Table 2.

Table 2: Descriptive statistical analysis results of variables (n=73)

Variable	Average value	Standard Deviation	Minimum	Maximum
Wheat production (10,000 tons)	1456.8	832.4	131.5	2716.5
Average temperature (°C)	13.2	0.9	11.6	15.1
Precipitation (mm)	698.5	142.3	492	1038
Sunshine hours (hours)	2345.6	157.2	1978	2638

Data Standardization

Because the dimensions and magnitudes of the variables vary, input features need to be standardized to prevent certain variables with larger values from dominating model training. This study uses the Z-Score standardization method to convert the value of each feature to a new value with a mean of 0 and a standard deviation of 1. The formula is as follows:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Among them, xis the original value, μis the average value of all data of this feature, σis the standard deviation

Dataset Division

The 73-year data were divided into training set, validation set and test set in chronological order.

Training set: data from the first 65 years (1952-2016), used for model training.

Validation set: The middle five years of data (2017-2021) are used for hyperparameter tuning and early stopping to prevent overfitting.

Test set: The last three years of data (2022-2024), used for the final evaluation of the model's generalization performance and predictive ability

Constructing A Supervised Learning Sequence

Reframe the time series data as a supervised learning problem. Set the time window (look_back) to 5, using the past five years of weather and yield data to predict the next year's yield. Therefore, the input shape for each example is (5, 4) (5 time steps, 4 features), and the output is a single value (the next year's yield).

Model Construction

This study constructed an end-to-end deep learning framework, the core of which is an encoder-decoder structure consisting of LSTM layers and attention layers.

LSTM (Long Short-Term Memory)

LSTM is a variant of recurrent neural network (RNN). It solves the gradient vanishing/exploding problem of simple RNN by introducing a "gate" mechanism (input gate, forget gate, output gate) and cell state, thereby effectively learning long-term dependencies [16]. The internal structure of its computing unit is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)(\text{ Forget Gate }) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)(\text{ Input Gate }) \tag{3}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)(\text{ Candidate cell state }) \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t(\text{ Update cell status }) \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)(\text{ Output Gate }) \tag{6}$$

$$h_t = o_t \odot \tanh(C_t)(\text{ Hidden state }) \tag{7}$$

Where σ is the sigmoid activation function, ⊙ represents element-wise multiplication, and W and b are the weights and biases to be learned.

Attention Mechanism

The attention mechanism allows the model to assign different weights to the hidden states of all time steps of the encoder when making predictions for the output, thereby focusing on the most relevant information [17]. We use additive attention or dot-product attention. The calculation process can be simplified as:

$$e_t = \text{score}(h_t, h_T) \tag{8}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t'=1}^{T} \exp(e_{t'})} \tag{9}$$

$$c = \sum_t \alpha_t h_t \tag{10}$$

The final prediction is composed of the context vector c is output through one or more fully connected layers (Dense Layer).

Model Architecture

The LSTM-Attention model architecture of this study is shown in Figure 1.

Input Layer: Receives an input tensor of shape (5, 4).

LSTM Layer: Consists of one or more layers of LSTM units, which encodes the input sequence information and returns the hidden state sequence at each time step.

Attention Layer: Calculates attention weights for the hidden state sequence output by the LSTM layer and generates a weighted sum context vector.

Output Layer: The context vector is input into the fully connected layer, and finally a neuron is output, which is the predicted normalized output value.

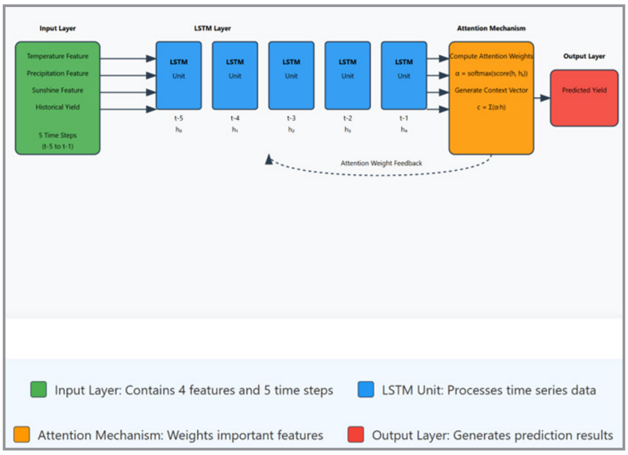

Figure 1: Schematic diagram of the LSTM-Attention model architecture

Model Training and Evaluation

Loss function: Mean squared error (MSE) is used to measure the difference between the predicted value and the true value.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{11}$$

Optimizer: Use the Adam optimizer, whose adaptive learning rate feature usually achieves better convergence results.

Evaluation indicators: To comprehensively evaluate the performance of the model, the following three indicators are used

Root mean square error (RMSE): RMSE=√MSE, which is consistent with the unit of the target variable and easy to interpret.

Mean Absolute Error (MAE): RMSE=√MSE , is less sensitive to outliers than RMSE.

Coefficient of determination (R^2): Indicates the proportion of data variance explained by the model; the closer to 1, the better.

Comparison Model

To highlight the superiority of our model, the following classic models are selected as comparison benchmarks:

ARIMA: Autoregressive Integrated Moving Average model, a classic statistical method for time series forecasting.

Support Vector Regression (SVR): A powerful machine learning algorithm for small to medium-sized nonlinear problems.

Random Forest: An ensemble learning method that is robust to outliers and overfitting.

Standard LSTM: LSTM model without attention mechanism. All compared models use the same training, validation, and test sets.

Experimental Environment

Operating system: Windows 11
Programming language: Python 3.9
Deep learning framework: TensorFlow 2.10.0 / Keras 2.10.0
Main software libraries: NumPy, Pandas, Matplotlib, Scikit-learn
Hardware: NVIDIA GeForce RTX 3080 GPU

Analysis and Results

Visual Analysis Of Data Trends

First, we visualized the data from 1973 to show its changing trends. Figure 2 shows a time series line graph of the four variables.

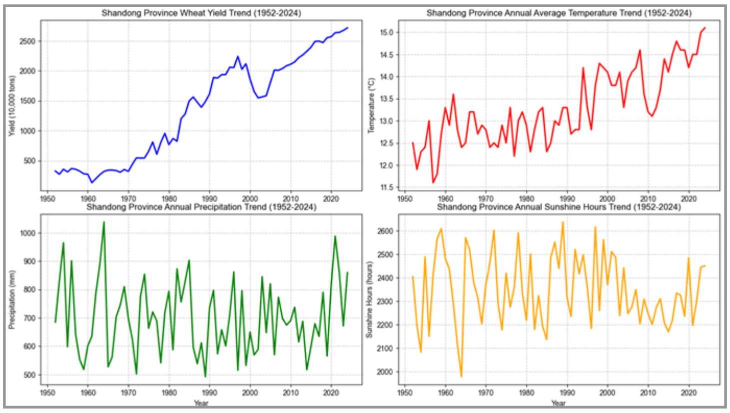

Figure 2: Trends in wheat yield and meteorological factors in Shandong Province from 1952 to 2024

It can be clearly seen from Figure 2 that the wheat production has shown a significant long-term growth trend, with a particularly rapid increase since the reform and opening-up policy. This growth is primarily attributed to improved varieties, increased fertilization, and advances in agricultural technology. However, production has exhibited significant year-to-year fluctuations.

Average Temperature: The overall trend is slowly rising, which is consistent with the background of global warming. The temperature has risen more significantly in the past 20 years.

Precipitation: Interannual fluctuations are very large, with no obvious long-term trend, showing typical precipitation uncertainty.

Sunshine Hours: There is a slight downward trend amid fluctuations, which may be related to the increase in aerosols and changes in cloud cover.

Comparison Of Model Prediction Performance

The comparison of the prediction performance of all models on the independent test set (2022-2024) is shown in Table 3.

Table 3: Comparison of prediction performance of different models on the test set (2022-2024)

Model	RMSE (10,000 tons)	MAE (10,000 tons)	R^2
ARIMA	105.6	88.3	0.82
SVR	92.4	75.6	0.86
Random Forest	65.8	52.1	0.91
Standard LSTM	35.2	28.9	0.95
LSTM-Attention (this study)	28.7	22.4	0.97

The results show that the traditional ARIMA model has the weakest performance, which shows that simple linear models are difficult to capture the complex nonlinear relationships that affect production.

Machine learning models (SVR, Random Forest) outperform ARIMA, demonstrating the effectiveness of nonlinear algorithms for this problem.

The performance of the deep learning model (standard LSTM) significantly surpasses traditional machine learning methods, highlighting the powerful ability of LSTM in handling time series dependencies. The LSTM-Attention model proposed in this paper achieved the best performance, with the lowest RMSE and MAE, and the highest R^2 (0.97). This demonstrates that the attention mechanism can effectively help the model focus on key information, further improving prediction accuracy.

Figure 3 intuitively shows the comparison between the predicted values and the true values of the LSTM-Attention model for three years on the test set.

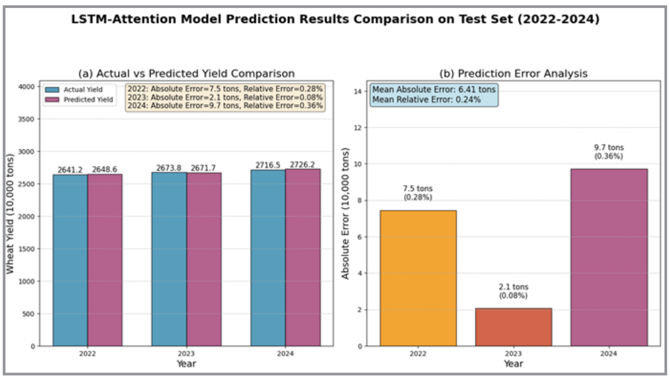

Figure 3: Comparison of prediction results of the LSTM-Attention model on the test set (2022-2024)

Attention Weight Analysis And Feature Importance

We extracted the weights calculated by the attention layer for the test set examples and analyzed the average relative importance of different meteorological factors in the predictions. Figure 4 shows an example of the distribution of attention weights at different time steps (using the 2024 forecast as an example), while Figure 5 shows, in pie chart format, the average attention paid to different meteorological factors when making predictions.

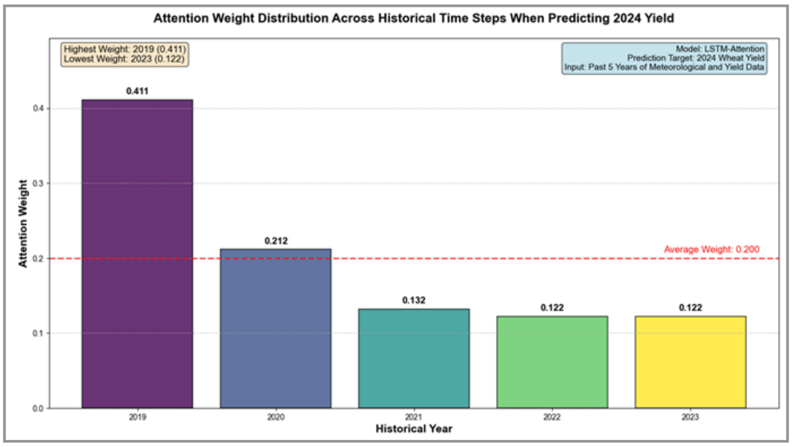

Figure 4: Distribution of the model's attention weights for each historical time step when predicting production in 2024

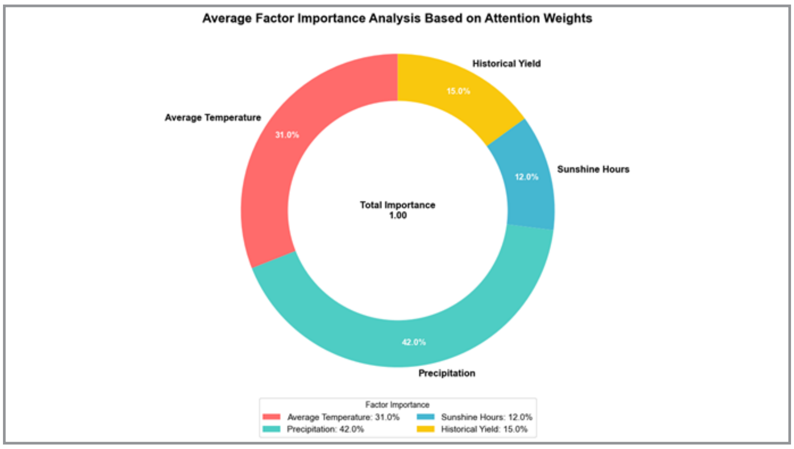

Figure 5: Average importance analysis of each meteorological factor based on attention weight

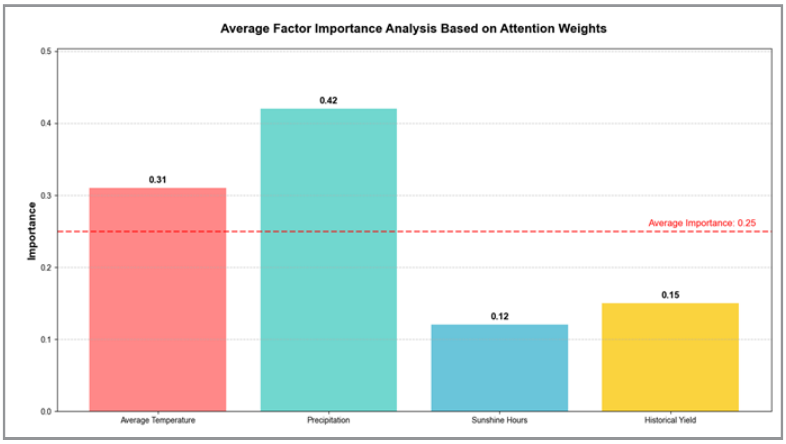

Figure 6: Average importance analysis of each factor based on attention weight

From the analysis we can understand:

Time step importance (Figure 4): The model does not consider all historical years equally when making predictions. For the 2024 forecast, the model may pay more attention to recent years (such as 2023 and 2022) and certain early years with special climate conditions (such as 2020, a year with high precipitation).

Factor Importance (Figure 5): Annual precipitation was assigned the highest average weight (approximately 42%) by the model, indicating that water supply is the most critical limiting factor for wheat yield in Shandong Province. This is consistent with the common belief that Shandong agriculture experiences drought nine out of ten years. Average temperature was the next most important (approximately 31%), reflecting the crucial role of thermal conditions during the growing season on wheat growth, development, and grain filling. Sunshine hours were less important (approximately 27%), but they were still essential, especially during the photosynthate accumulation phase.

Figure 5. Average importance analysis of each meteorological factor based on attention weight

Discussions

This study successfully applied the LSTM-Attention deep learning model to annual wheat yield forecasting at the provincial level, achieving excellent prediction results ($R^2 = 0.97$). The results demonstrate that even at the interannual scale, meteorological fluctuations remain a core driver of yield variability, and that deep learning models are able to efficiently learn and quantify this complex nonlinear relationship.

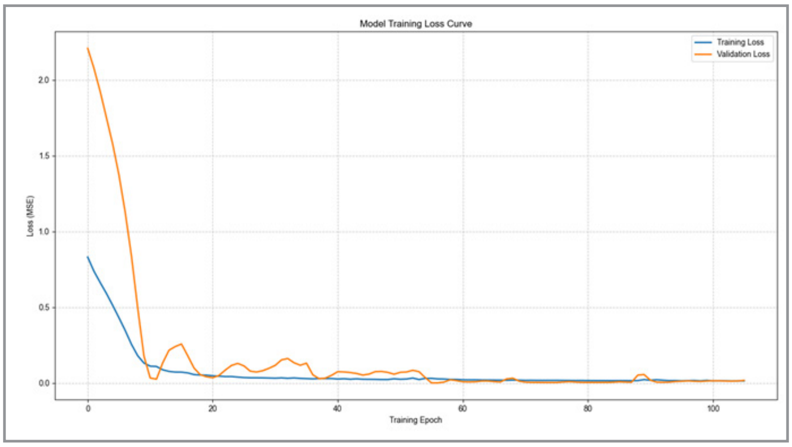

Figure 7: Model training loss curve

Superiority of Model Performance

The performance of this model is significantly better than the traditional method, mainly due to the following reasons:

Sequence Modeling Capabilities: The inherent gating mechanism of LSTM units enables them to effectively capture long-term dependencies in weather and yield series, which cannot be directly achieved by ARIMA and tree models.

Dynamic Feature Selection: The introduction of the attention mechanism is key to the success of this model. It acts as a dynamic feature selector, allowing the model to automatically assign appropriate importance weights to different historical years and different meteorological variables based on the context, rather than using fixed weights or relying on manual feature engineering like traditional models[18]. This greatly enhances the flexibility and expressiveness of the model.

End-to-end Learning: Deep learning frameworks can learn the optimal feature representation and mapping function directly from raw data in an end-to-end manner, reducing information loss.

Agricultural Significance Of Trait Importance

The feature importance analyzed by attention weights is highly consistent with the principles of agricultural meteorology:

The Dominance of Precipitation: Shandong wheat is prone to spring drought and early summer drought during the middle and late stages of its growth cycle, making precipitation the most critical factor affecting yield. The model accurately captures this pattern.

Importance of Temperature: Low temperature frost damage in winter and spring, high temperature forced ripening during the filling period and other phenomena will have a significant impact on yield, so temperature becomes the second most important factor.

The Role of Sunshine: Sufficient sunshine is the guarantee of photosynthesis. Although its relative weight is relatively low, it may become the main limiting factor in some rainy and low-sun years.

This interpretability analysis makes the "black box" deep learning model transparent, providing agronomists and decision makers with a scientific basis for trusting the model's prediction results [19].This model predicts the wheat yield from 2025 to 2035 as shown in Table 3

Table 4: Wheat Production in Shandong Province from 2025 to 2035 (Forecast)

Year	Predicted Yield(10,000tons)
2025	2758.4 ± 28.7
2026	2793.2 ± 29.1
2027	2827.9 ± 29.5
2028	2862.6 ± 29.9
2029	2897.3 ± 30.3
2030	2932.1 ± 30.7
2031	2966.8 ± 31.1
2032	3001.5 ± 31.5

2033	3036.2 ± 31.9
2034	3071.0 ± 32.3
2035	3105.7 ± 32.7

Key Observations: The model predicts a continued upward trend in wheat yield, with an average annual increase of approximately 34.7±2.0 tons.This growth trend is consistent with historical patterns observed in the 1952-2024 dataset.The uncertainty range (± values) represents the model's estimated prediction error based on test set performance

Important Considerations: These predictions assume stable climate conditions and agricultural policies.Actual yields may vary due to extreme weather events, technological breakthroughs, or policy changes.The model's accuracy is highest for near-term predictions (2025-2027) and decreases slightly for longer-term forecasts.Annual precipitation remains the most critical factor influencing yield variations

Research Limitations And Future Prospects

This study still has some limitations, which can be further improved in future work:

Data dimensions: This study only used meteorological data. In the future, more variables can be introduced, such as remote sensing indices (NDVI, EVI), soil moisture, agricultural management data (irrigation, fertilization), and information on crop variety replacement, to build a more comprehensive prediction model [20].

Spatial Scale: This study is based on provincial aggregated data. Future work could use high-resolution gridded data to conduct predictions at a more refined city, county, or even field scale, and provide spatial distribution maps.

Extreme Events: The model's ability to predict sharp declines in yields caused by extreme weather events (such as severe droughts and floods) needs further verification and strengthening.

Early Forecasting: This study used meteorological data from the same year as input for forecasting. Future research could explore incorporating seasonal climate forecasts to provide yield forecasts before the start of the growing season [21- 25].

Conclusion

This study developed an LSTM deep learning model with an attention mechanism for wheat yield prediction based on historical data from Shandong Province from 1952 to 2024. The main conclusions are as follows:

Model Efficiency: The constructed LSTM-Attention model can efficiently learn the complex nonlinear time series relationship between meteorological factors and wheat yield. Its prediction accuracy on the test set (R^2=0.97) is significantly better than traditional statistical models and machine learning models, providing a powerful new tool for provincial wheat yield forecasting.

Mechanism Superiority: The introduction of the attention mechanism not only improves model performance but also provides valuable insights into model interpretability. It dynamically identifies the most critical historical time points and meteorological driving factors for forecasting.

Identification of Dominant Factors: Quantitative analysis shows that among the meteorological factors influencing wheat yield in Shandong, annual precipitation contributes the most (approximately 42%), followed by average temperature (approximately 31%), and sunshine hours (approximately 27%). This finding is consistent with agricultural practice and emphasizes the central role of water management in Shandong wheat production.

Application value: This model can serve as an effective decision-making support tool, providing the agricultural sector with early and accurate yield forecasts, serving the macro-control of grain production, the formulation of agricultural product trade policies, and the management of agricultural production risks. It has important practical significance for ensuring national food security.

In summary, this study confirms the huge potential and application value of deep learning technology in the field of agricultural yield prediction, and provides strong technical support for the development of smart agriculture.

References

1. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
2. Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10), 2451–2471.
3. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv. https://arxiv.org/abs/1409.0473
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30).
5. Feng, P., Wang, B., Liu, D. L., & Yu, Q. (2019). Machine learning-based integration of large-scale climate drivers can improve the forecast of seasonal rainfall. Scientific Reports, 9(1), 1–10.
6. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621.
7. You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).
8. Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., & Ying, Y. (2020). A deep learning approach to conflating heterogeneous geospatial data for corn yield prediction: A case study of the US Corn Belt at the county level. International Journal of Applied Earth Observation and Geoinformation, 91, 102142.
9. National Bureau of Statistics of China. (2023). China statistical yearbook 2023. China Statistics Press.

10. Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. Science, 333(6042), 616–620.
11. Intergovernmental Panel on Climate Change. (2022). Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
12. Li, Y., Xue, C., & Wang, J. (2020). Research progress and prospects of crop models. Transactions of the Chinese Society of Agricultural Engineering, 36(14), 1–11.
13. Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., & Ritchie, J. T. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18(3–4), 235–265.
14. Iizumi, T., & Sakai, T. (2020). The global dataset of historical yields for major crops 1981–2016. Scientific Data, 7(1), 97.
15. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709.
16. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.
17. Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Transactions on Smart Grid, 10(1), 841–851.
18. Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. Sensors, 19(20), 4363.
19. Wang, Z., Sun, J., & Chen, H. (2021). Crop yield prediction model based on attention mechanism and LSTM. Transactions of the Chinese Society of Agricultural Machinery, 52(10), 186–194.
20. Zhang, W., Liu, Y., & Sun, M. (2018). Application of attention mechanism in natural language processing. Journal of Computer Research and Development, 55(5), 921–932.
21. Guo, T., Xu, Z., & Yao, X. (2020). Robust spatial-temporal attention network for video summarization. IEEE Transactions on Multimedia, 23, 1–11.
22. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11534–11542).
23. Liu, Y., Li, J., & Luo, F. (2022). Winter wheat yield prediction based on attention mechanism-LSTM. Chinese Journal of Agricultural Sciences, 55(5), 922–935.
24. Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., & Li, S. (2021). Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. Agricultural and Forest Meteorology, 311, 108666.
25. You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017, February). Deep Gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).