UDC 004.896
DOI: 10.15587/1729-4061.2023.292849

DEVELOPMENT OF A NEURAL NETWORK MODEL FOR TRAINING DATA ON THE EFFECTS OF PHOSPHORUS ON SPRING WHEAT GROWTH

Saltanat Sharipova
Doctoral Student*
Akerke Akanova
Corresponding author
PhD, Senior Lecture*
E-mail: akerkegansaj@mail.ru
Nazira Ospanova
Professor, Candidate of Pedagogical Sciences
Department of Computer Science
Toraighyrov University
Lomov str., 64, Pavlodar, Republic of Kazakhstan, 140008
*Department of Computer Engineering and Software
S. Seifullin Kazakh Agro Technical Research University
Zhenis ave., 62, Astana, Republic of Kazakhstan, 010011

A neural network was developed to predict the effect of phosphorus on spring wheat yields. The focus is on the neural network, including its structure, parameters, training methods and results related to phosphorus effects on spring wheat yields. An algorithm for developing a neural network model is also presented.

The study was conducted to address the critical need for developing a neural network to predict the effect of phosphorus on spring wheat yields in the Republic of Kazakhstan.

For data analysis, input data were used that cover the period from 2012 to 2022, including climatic indicators, regional features and phosphorus application. The target variable is spring wheat yield. To ensure the accuracy of the study, the data were preprocessed and standardized, and an outlier and variance analysis was performed. The developed neural network was trained and tested to obtain the best results. The mean squared error was used as a metric for evaluating the quality of forecasting. Additionally, indicators such as mean absolute error and coefficient of determination were considered.

The results of the study showed an MSE of 7.12, indicating that the model agrees well with the data and makes accurate predictions, which also suggests its practical relevance. The correlation analysis of the features showed that phosphorus application and spring wheat yield have a positive relationship. These results can be very useful for agriculture and farming enterprises, as they allow optimizing phosphorus application to the soil and increasing wheat yields

Keywords: neural network forecasting model, yield forecasting, phosphorus data, neural networks

Received date 30.08.2023
Accepted date 01.12.2023
Published date 29.12.2023

***How to Cite:** Sharipova, S., Akanova, A., Ospanova, N. (2023). Development of a neural network model for training data on the effects of phosphorus on spring wheat growth. Eastern-European Journal of Enterprise Technologies, 6 (4 (126)), 32–38. doi: https://doi.org/10.15587/1729-4061.2023.292849*

1. Introduction

The use of neural networks has a number of significant advantages [1]. First, neural networks are capable of automatically detecting complex and nonlinear relationships, which can be difficult to identify using traditional statistical methods [2]. Second, they are able to process large amounts of data, including many factors, which allows many variables to be taken into account in the analysis [3]. In addition, neural networks can be trained on data from various sources and adapt to changing conditions, making them a powerful tool for process modeling. Finally, the results obtained using neural networks can provide more accurate and customized recommendations [4]. The use of advanced technologies such as neural networks makes it possible to enhance the efficiency of agricultural systems, as well as effectively predict and adapt to variable conditions in agriculture.

Automation of agriculture has become an integral part of modern rural production, bringing many advantages and opportunities. However, in the context of agricultural automation, it is important to take into account changing natural conditions and climatic changes, which can have a serious impact on yields and production efficiency. Given these factors, forecasting becomes a key element in predicting and adapting to global changes in nature [5]. To organize such forecasting, various intelligent systems are being developed.

Intelligent systems have become a key tool for agricultural automation, but they also face challenges and problems [6]. One of the essential tasks is to accurately predict various aspects of agricultural production, ranging from yield levels to optimal fertilizer and resource management methods.

Anticipating future events and changes in agriculture is critical to enable early adaptation to changes in climate and natural conditions. This, in turn, helps to increase the sustainability of rural production and reduce risks associated with changing nature [7].

In this context, it is important to conduct research and develop intelligent systems that can predict the impact of various factors, such as phosphorus, on spring wheat yields. Accurate forecasts will allow agriculture to make informed decisions and adapt to changing conditions, ensuring food security and sustainable production [8]. Therefore, research on the development of a neural network model to predict the effect of phosphorus on spring wheat yields is relevant.

2. Literature review and problem statement

In modern scientific research, meteorological indicators and soil chemistry are most often considered as input data. Nitrogen, and, in rare cases, phosphorus, are common factors affecting crop yields. Crops such as wheat, corn and rice are often the subject of research, given their strategic importance for agriculture. In [9], a machine learning model was developed to predict winter wheat yields in Hokkaido at the municipal level based on meteorological data. In [10], a deep LSTM machine learning model was developed to estimate crop yields also based on climate data. The LSTM model showed outstanding performance in yield estimation compared to backpropagation and support vector machine (SVM) methods, providing stable performance under changing climate and soil conditions. However, predicting crop yields based on weather data alone is not enough, so, for example, in [11], the effect of nitrogen on yields was also predicted. As a result, it is concluded that increasing the dose of nitrogen fertilizer significantly improves the yield and efficiency of nutrient use in wheat crops. Similarly, in [12] using artificial neural networks, models were developed to estimate total nitrogen and mineralized nitrogen at the district level, as well as to predict wheat yields. The best model from this study showed that yields can be predicted based on precipitation, photothermal quotient and mineralized nitrogen.

It was the use of phosphorus that was considered in [13], where a neural network was developed that predicts changes in phosphorus levels in the soil during its cultivation. However, the effect of phosphorus on yields proved to be unpredictable due to limitations in the data considered. In [14], three multicriteria models were created to predict wheat yields using neural networks with the Multilinear Perceptron topology based on climate data and fertilizers (containing phosphorus). These models can predict and model yields for only three dates: April 15, May 31, and June 30. Also in [15], phosphorus, potassium, nitrogen application and climate data were used to predict rice yields in India. The study proposed a hybrid MLR-ANN (Multiple Linear Regression – Artificial Neural Network) model for more accurate forecasting.

In [16], two methods, the linear regression method (LASSO) and the backpropagation neural network method (BPNN), were compared to predict wheat yield at the field level based on climate data and soil characteristics. Another work [17] developed and evaluated a model for predicting winter wheat yield potential. The main input data for the prediction were estimated soil parameters (calcium, phosphorus, magnesium and others) using online soil spectroscopy with a prototype sensor. The proposed architecture provided visual information on factors affecting yield potential from online data on multilayer soil properties and crop growth characteristics based on satellite images. The work [18] investigated the possibility of predicting wheat yields in China using multidimensional data such as climate, remote sensing and soil data (physical and chemical soil properties) using machine learning methods. The results showed that models based on machine learning methods such as SVM, GPR and RF can give accurate yield forecasts with high accuracy, however, the listed works do not consider phosphorus application to the soil, but phosphorus already present in the soil.

To date, none of the papers reviewed has investigated or examined the effect of phosphorus on spring wheat yields. It is important to emphasize that such research will have significant relevance and potential practical implications for the agricultural sector. If a robust model is developed to predict the effects of phosphorus on spring wheat yields, this can be of great importance for agricultural producers, helping them optimize fertilizer use and increase yields.

All this suggests that it is advisable to conduct a study on the development of a neural network model to predict the effect of phosphorus on spring wheat yield.

3. The aim and objectives of the study

The aim of the study is to develop a neural network model for training data on phosphorus and predicting its impact on wheat yields. This will make it possible to effectively organize sowing and increase the yield of spring wheat.

To achieve the aim, the following objectives are accomplished:

– to prepare data for neural network training;

– to develop an algorithm and model for predicting the effect of phosphorus on spring wheat yield.

4. Materials and methods

The object of the study is the modeling of a neural network to predict the effect of phosphorus on spring wheat yields.

The hypothesis of the study suggests that the use of a neural network model will allow predicting the effect of phosphorus on spring wheat yields with high accuracy.

For training, input data from the National Bureau of Statistics of the Republic of Kazakhstan [19] and the National Hydrometeorological Service of the Republic of Kazakhstan were used [20]. The raw data included information on soil surface temperature, humidity, and precipitation over the past 10 years for various regions every three hours of every day of the year. The data provided also contain information on phosphorus application to the soil and spring wheat yield by region for each year. The presented data were processed, and the final dataset included such features as soil surface temperature, humidity, precipitation from April to September and phosphorus application. Spring wheat yield was chosen as the target variable.

Prior to training, preliminary data processing was carried out. This involved handling missing values and standardizing features. The missing values were filled in with the medians of the corresponding feature. The median value is the most suitable filling method for this study, as it divides the ordered dataset into two equal halves, and it can be more robust to outliers present in the dataset in question.

To eliminate differences in scale between the features used in the analysis, the feature standardization procedure was applied. Standardization is an important step in preparing data for training a neural network model and allows features to be comparable in size and scale.

Correlation between the parameters was also analyzed to determine the importance of each factor in the context of the impact on spring wheat yield. This correlation analysis helped to better understand the relationships between the various variables and their impact on the target.

In the final stage of data preparation, outliers and variance in the dataset were analyzed. The outlier and variance analysis is an important step in data processing, as it allows you to identify and account for anomalies or errors in data that can distort the results of the analysis.

Various statistical methods were used to identify potential outliers or anomalies in the data, including box plot analysis, value standardization, and outlier detection methods based on data distribution.

The mean squared error was used as a metric for evaluating the quality of forecasting. Assessing the quality of the model using MSE will allow us to judge how well the developed model is able to reproduce the effect of phosphorus on spring wheat yield. In addition, the training process will be carefully reviewed in order to minimize MSE and achieve the best forecasting results. To assess the quality of the model, additional metrics were applied – mean absolute error (MAE) and coefficient of determination (R^2). MAE is the mean of the absolute differences between the model's predictions and actual values. It is a measure of model accuracy, with a lower value indicating more accurate predictions.

The coefficient of determination (R^2) provides a measure of how well the model explains the variability of the dependent variable relative to independent variables. The R^2 value ranges from 0 to 1, with 1 indicating a perfect fit of the model to the data, and a value closer to 0 means that the model does not explain the variance in the dependent variable.

5. Results on predicting the effect of phosphorus on spring wheat yield

5. 1. Preparing data for neural network training

The existing neural network models did not consider the effect of phosphorus on wheat yield and focused more on meteorological data. Therefore, to model a neural network that will take into account not only weather data, but also phosphorus application in the Republic of Kazakhstan, it is important to perform a number of actions.

To achieve accuracy in forecasting, it is necessary to standardize the features. The feature standardization procedure was carried out using StandardScaler from the sklearn library in Python. Fig. 1 shows the data before and after standardization.

T_april	T_may	T_june	T_july	T_august
4.89583	20.3548	24.8250	27.9274	21.8145
8.91250	19.1492	24.1208	32.9839	26.0887
6.37500	19.1694	22.5375	26.6411	20.1532
6.52500	17.0605	22.8333	24.5323	21.8871
9.76250	21.9113	26.0417	27.2298	24.2097
...	...	...	...	...

T_april	T_may	T_june	T_july	T_august
-1.137759	0.154760	-0.319746	0.061732	-0.876836
-0.273208	-0.146687	-0.500296	1.280888	0.201226
-0.819381	-0.141637	-0.906238	-0.248403	-1.295858
-0.787095	-0.668945	-0.830398	-0.756848	-0.858524
-0.090253	0.543947	-0.007797	-0.106463	-0.272705
...	...	...	...	...

Fig. 1. Data before and after standardization

Each feature and target variable were also analyzed. Table 1 shows the relationships between the features and the target variable (correlation coefficient) and their characteristics.

The next step was to analyze outliers and variance in data processing. As can be seen in Fig. 2, there are outliers for some features. But within the framework of this study, such a feature as phosphorus application to the soil has more weight.

Table 1

Feature characteristics and correlation coefficient

Feature	Description	Unit of measurement	Correlation coefficient with the target variable
year	Year of sowing	–	0.203773
location	Place of sowing (location)	Id	–0.032777
T_april, T_may, T_june, T_july, T_august, T_september	Average soil surface temperature from April to September of the specified year	°C	–0.021887
			–0.098259
			–0.114572
			–0.084751
			–0.058412
			–0.038178
humidity_april, humidity_may, humidity_june, humidity_july, humidity_august, humidity_september	Average air humidity from April to September of the specified year	%	0.084636
			0.083512
			0.112556
			0.098496
			0.083570
			0.074636
precipitation_april, precipitation_may, precipitation_june, precipitation_july, precipitation_august, precipitation_september	Average precipitation from April to September of the specified year	mm	–0.021887
			–0.098259
			–0.114572
			–0.084751
			–0.058412
			–0.038178
phosphorus	Phosphorus application to the soil in the specified year	centner	0.432935

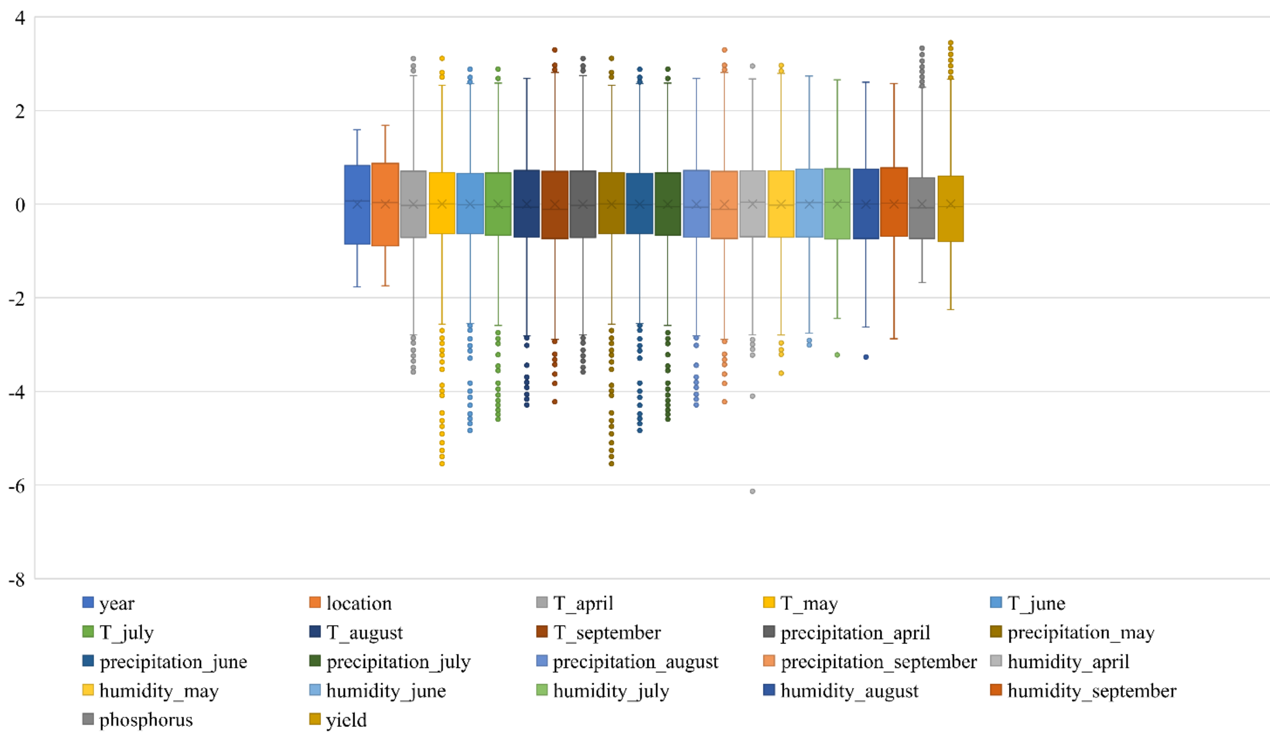

Fig. 2. Box plot for all columns

These outliers do not represent errors in the data, but are real anomalies, and therefore the possibility of leaving them was considered. Thus, the preprocessed data are ready to be fed into the developed neural network.

5. 2. Development of an algorithm and model for predicting the effect of phosphorus on spring wheat yields

The algorithm for developing a neural network model to predict the effect of phosphorus on spring wheat yield is shown in Fig. 3.

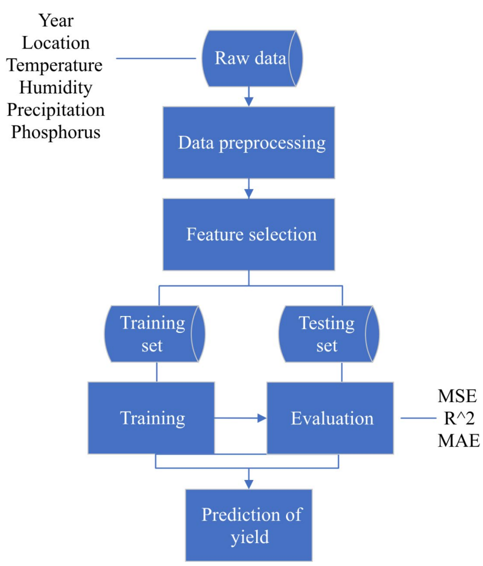

Fig. 3. Algorithm for developing a neural network forecasting model

As part of this study, a neural network was developed consisting of several layers, including input, hidden and output layers. The architecture (Fig. 4) was matched to the regression problem, and activation functions were optimized during the experiments. Data such as soil surface temperature, air humidity, precipitation from April to September, year and location, and phosphorus application are fed to the input of the neural network. And to the output – the yield of spring wheat.

The model was trained on a training dataset using the mean squared error as a loss function. The mean absolute error and the coefficient of determination were also considered. To speed up convergence, the Adam optimizer was used.

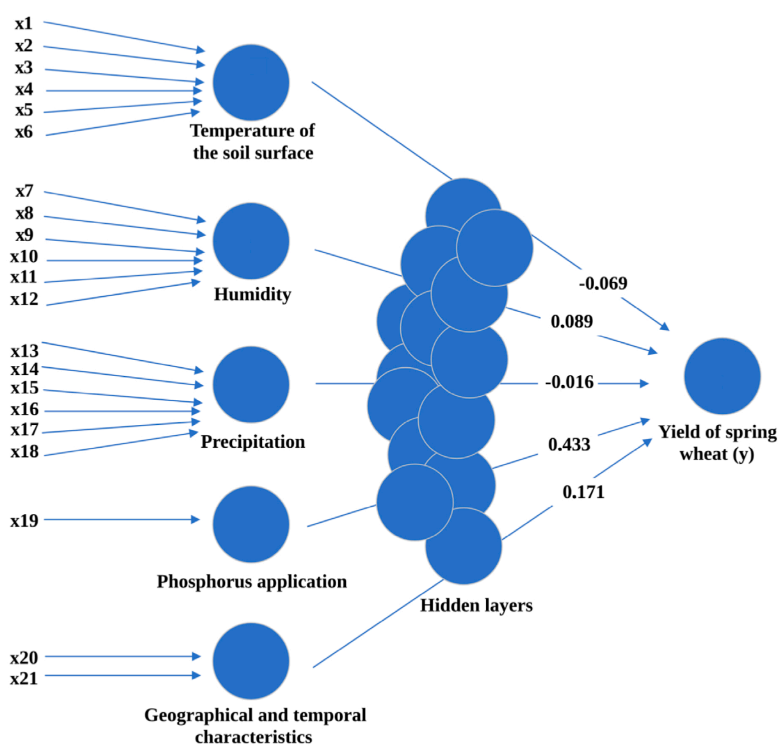

Fig. 4. Architecture of the forecasting neural network

Table 2 shows the forecasting results using the developed neural network.

The graph below (Fig. 5) shows the results of neural network predictions on the effect of phosphorus levels on spring wheat yield. The graph displays the predicted (in green) and actual values (in blue) of spring wheat yield.

Table 2

Forecasting results

Metric	Designation	Formula	Value
Mean squared error	MSE	$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}\right)^2$	7.12
Coefficient of determination	R^2	$R^2=\frac{\sum_{i=1}^{n}\left(\hat{y}-y_{mean}\right)^2}{\sum_{i=1}^{n}\left(y_i-y_{mean}\right)^2}$	0.87
Mean absolute error	MAE	$MAE=\frac{1}{n}\sum_{i=1}^{n}\left\lvert y_i-\hat{y}\right\rvert$	1.52

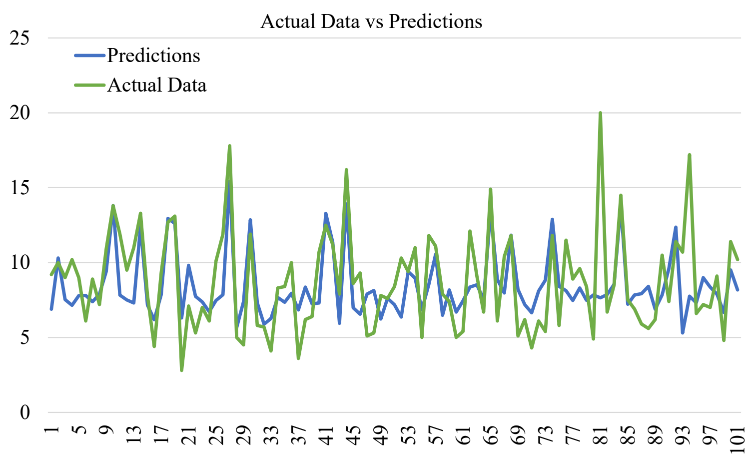

Fig. 5. Graph of actual and predicted data

From the analysis of the presented graph, it can be said that the actual and predicted values show similarity and subtle differences in their numerical expressions, which indicates the level of proximity between these sets. The values of actual and forecast data are in close proportion and do not show significant amplitude gaps or systematic discrepancies.

6. Discussion of the results of predicting the effect of phosphorus on spring wheat yield

The process of modeling a neural network for predicting the effects of phosphorus on spring wheat yields included data processing, the development of an appropriate algorithm, and the creation of a neural network. By the results of analyzing the input data, outliers were found (Fig. 2), and the values of the input variables had different scales. In this regard, standardization was carried out (Fig. 1), which made outliers less noticeable and equalized the scale of the features.

From Table 1, an important observation can be highlighted: among all the factors influencing wheat yield, the highest correlation coefficient (0.43) is observed in the case of phosphorus application to the soil. This indicates a clear relationship between phosphorus application to the soil and wheat yield, which can be considered a key factor for increasing yields.

According to the algorithm shown in Fig. 3, a neural network model with a hidden-layer architecture was built (Fig. 4). This neural network was trained using processed input data.

To assess the quality of the model, the mean absolute error (MAE) and mean squared error (MSE), as well as the coefficient of determination (R^2), were used. The model shows an MAE of 1.52 and an MSE of 7.12, indicating that the model generally agrees well with the data and makes accurate predictions.

Based on the results of forecasting the effect of phosphorus on spring wheat yield, the coefficient of determination is 0.87. This is a good indicator, showing a good fit of the model to the data, and is also higher than the results in a similar work [14].

In the graph (Fig. 5), you can see that the predicted values (in green) are similar to the actual values (in blue) over the entire time interval. This agreement between predictions and actual data confirms the reliability and efficiency of the algorithm.

A feature of the results obtained, compared to other works [13–15], is that hidden layers are used in which meteorological data and the amount of phosphorus added are processed. The study confirms the importance of phosphorus application to the soil to increase spring wheat yields, as there is a positive correlation between these parameters. The results of the model evaluation by metrics indicate the high efficiency of the developed model and its ability to accurately predict spring wheat yields when phosphorus is applied to the soil. The results are explained by the correct selection and processing of data, the adapted neural network model and adequate interpretation of the data. Besides, high rates of the coefficient of determination, low values of the mean squared and mean absolute errors demonstrate high accuracy of the model. These results have the potential for application in agriculture and allow farmers to make more informed decisions to increase wheat yields.

The result of data training can be limited by the boundaries of applicability within the Republic of Kazakhstan, since all data were collected regarding crops grown in the fields of this country. But the results of this study can be applied in other countries, provided that the climatic and soil conditions are similar to those of the Republic of Kazakhstan. Further steps include conducting additional research and experiments in various geographical areas to adapt the developed methods to local requirements and features.

The disadvantage of this development is the lack of an interface, which, accordingly, makes it impossible to supplement with new data, train and save them in the database. Further development of this research provides prospects for developing a system with improved access and interface. The creation of such a system can help increase the accessibility and clarity of the results obtained, as well as provide a wider range of users interested in this area.

7. Conclusions

1. When analyzing the input data for training the developed neural network, missing values were found, which were successfully filled with median values. To ensure the stability and consistency of the data, the features were standardized in order

to equalize their variance. Additionally, an analysis of outliers and correlations between the features and the target variable was carried out. As a result of all the measures taken, processed data were submitted at the input of the neural network.

2. As a result of the study, an algorithm for building a neural network was developed to predict the effect of phosphorus on spring wheat yield. A neural network model was built that successfully predicts results. The evaluation of the model quality showed that the mean squared error (MSE) was 7.12, the mean absolute error (MAE) was 1.52 and the coefficient of determination (R^2) was 0.87. The results confirm the reliability and accuracy of the algorithm and model.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was conducted without financial support.

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the presented work.

References

1. Nedic, V., Despotovic, D., Cvetanovic, S., Despotovic, M., Babic, S. (2014). Comparison of classical statistical methods and artificial neural network in traffic noise prediction. Environmental Impact Assessment Review, 49, 24–30. doi: https://doi.org/10.1016/j.eiar.2014.06.004
2. Hasanzadehshooiili, H., Lakirouhani, A., Medzvieckas, J. (2012). Superiority of artificial neural networks over statistical methods in prediction of the optimal length of rock bolts. Journal of Civil Engineering and Management, 18 (5), 655–661. doi: https://doi.org/10.3846/13923730.2012.724029
3. Akanova, A., Ospanova, N., Sharipova, S., Mauina, G., Abdugulova, Z. (2022). Development of a thematic and neural network model for data learning. Eastern-European Journal of Enterprise Technologies, 4 (2 (118)), 40–50. doi: https://doi.org/10.15587/1729-4061.2022.263421
4. Zhang, L., Huang, Z., Liu, W., Guo, Z., Zhang, Z. (2021). Weather radar echo prediction method based on convolution neural network and Long Short-Term memory networks for sustainable e-agriculture. Journal of Cleaner Production, 298, 126776. doi: https://doi.org/10.1016/j.jclepro.2021.126776
5. Rodríguez, S., Gualotuña, T., Grilo, C. (2017). A System for the Monitoring and Predicting of Data in Precision Agriculture in a Rose Greenhouse Based on Wireless Sensor Networks. Procedia Computer Science, 121, 306–313. doi: https://doi.org/10.1016/j.procs.2017.11.042
6. Aggarwal, P., Shirsath, P., Vyas, S., Arumugam, P., Goroshi, S., Aravind, S. et al. (2020). Application note: Crop-loss assessment monitor – A multi-model and multi-stage decision support system. Computers and Electronics in Agriculture, 175, 105619. doi: https://doi.org/10.1016/j.compag.2020.105619
7. Kansiime, M. K., Rwomushana, I., Mugambi, I., Makale, F., Lamontagne-Godwin, J., Chacha, D. et al. (2020). Crop losses and economic impact associated with papaya mealybug (Paracoccus marginatus) infestation in Kenya. International Journal of Pest Management, 69 (2), 150–163. doi: https://doi.org/10.1080/09670874.2020.1861363
8. Zhang, Q., Wang, K., Han, Y., Liu, Z., Yang, F., Wang, S. et al. (2022). A crop variety yield prediction system based on variety yield data compensation. Computers and Electronics in Agriculture, 203, 107460. doi: https://doi.org/10.1016/j.compag.2022.107460
9. Murakami, K., Shimoda, S., Kominami, Y., Nemoto, M., Inoue, S. (2021). Prediction of municipality-level winter wheat yield based on meteorological data using machine learning in Hokkaido, Japan. PLOS ONE, 16 (10), e0258677. doi: https://doi.org/10.1371/journal.pone.0258677
10. Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., Li, H. (2021). An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. Agricultural and Forest Meteorology, 310, 108629. doi: https://doi.org/10.1016/j.agrformet.2021.108629
11. Sharma, S., Kaur, G., Singh, P., Alamri, S., Kumar, R., Siddiqui, M. H. (2022). Nitrogen and potassium application effects on productivity, profitability and nutrient use efficiency of irrigated wheat (Triticum aestivum L.). PLOS ONE, 17 (5), e0264210. doi: https://doi.org/10.1371/journal.pone.0264210
12. Alvarez, R., De Paepe, J. L., Gimenez, A., Recondo, V., Pagnanini, F., Mendoza, M. R. et al. (2019). Using a nitrogen mineralization index will improve soil productivity rating by artificial neural networks. Archives of Agronomy and Soil Science, 66 (4), 517–531. doi: https://doi.org/10.1080/03650340.2019.1626984
13. Alvarez, R., Steinbach, H. S. (2017). Modeling Soil Test Phosphorus Changes under Fertilized and Unfertilized Managements Using Artificial Neural Networks. Agronomy Journal, 109 (5), 2278–2290. doi: https://doi.org/10.2134/agronj2017.01.0014

14. Niedbała, G., Nowakowski, K., Rudowicz-Nawrocka, J., Piekutowska, M., Weres, J., Tomczak, R. J. et al. (2019). Multicriteria Prediction and Simulation of Winter Wheat Yield Using Extended Qualitative and Quantitative Data Based on Artificial Neural Networks. Applied Sciences, 9 (14), 2773. doi: https://doi.org/10.3390/app9142773
15. Maya Gopal, P. S., Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. Computers and Electronics in Agriculture, 165, 104968. doi: https://doi.org/10.1016/j.compag.2019.104968
16. Tang, X., Liu, H., Feng, D., Zhang, W., Chang, J., Li, L., Yang, L. (2022). Prediction of field winter wheat yield using fewer parameters at middle growth stage by linear regression and the BP neural network method. European Journal of Agronomy, 141, 126621. doi: https://doi.org/10.1016/j.eja.2022.126621
17. Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture, 121, 57–65. doi: https://doi.org/10.1016/j.compag.2015.11.018
18. Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J. (2020). Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. Remote Sensing, 12 (2), 236. doi: https://doi.org/10.3390/rs12020236
19. Bureau of National Statistics. Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Available at: https://stat.gov.kz/en/
20. National Hydrometeorological Service of Kazakhstan. Available at: https://www.kazhydromet.kz/en/