Assessment of soil fertility indicators based on remote sensing data

Georgy Matyunin[1*], *Svetlana* Ogorodnikova[1], *Ekaterina* Murmantseva[1], *Vladislav* Rozanov[1], and *Ruslan* Palyga[1]

[1] Moscow Aviation Institute, Moscow, Russia

Abstract. The article discusses the possibility of using artificial intelligence and machine learning methods to predict soil fertility based on remote sensing data. The research object was 70 soil samples taken in the village of Vshchizh, Bryansk region. Multispectral imaging of the earth's surface from the Sentinel-2 satellite was also used to assess the soil condition. The study in this article is based on the indicators of humus (%), P2O5 (mg/kg), and K2O (mg/kg). The prediction is done using a neural network model based on the Rosenblatt perceptron. Data analysis is conducted using the statistical software environment RStudio. The results of the model show the values of the total mean square error (MSE): MSE=0.178 for humus prediction, MSE=0.138 for P_2O_5 prediction, MSE=0.171 for K_2O prediction. Additionally, the program calculated the correlation values between the predicted and calculated soil fertility. K(humus)=0.548, K(P_2O_5)=0.287, K(K_2O)=0.244. Thus, the neural network most accurately predicted soil fertility based on the humus content.

1 Introduction

Currently, remote sensing data (RS) are actively used for environmental monitoring. They are collected using satellites[1]. With their help, it is possible to study indicators such as the condition of vegetation cover[2], chemical composition of soils[3], and other equally important indicators. RS also allow monitoring land over time, providing opportunities to establish the causes of changes in territory conditions and take necessary measures in a timely manner[4]. However, one should not forget the aspect without which it is difficult, and sometimes impossible, to conduct research - artificial intelligence (AI). It is AI algorithms, in conjunction with remote sensing data, that ensure high accuracy and reliability of forecasts in the field of agroclimatic monitoring and land management[5]. In particular, artificial intelligence and machine learning methods allow predicting the yield of agricultural crops[6].

The task of developing algorithms predicting crop yield is relevant in our time[7]. This allows the farmer to determine whether there are enough nutrients in the soil for plants and to track the dynamics of changes in soil properties that affect yield and vegetation. This makes it possible to make effective management decisions, both in the field of national land use and in the field of economics[8].

* Corresponding author: medvedka_799rus@mail.ru

Thanks to such technologies, it is possible to reduce the amount of fertilizer applied, and as a result, reduce the cost of the final agricultural products.

In particular, in addition to the described capabilities, artificial intelligence and machine learning methods can be used to predict plant diseases and determine optimal methods of treatment [9]. By analyzing data on disease symptoms, environmental conditions, and previous cases of diseases, a model can be created to help predict the probability of occurrence of a specific disease and optimal methods to combat it.

The aim of this research is to test a computer program created based on artificial intelligence methods in the R software environment, allowing to forecast the yield of agricultural crops.

2 Materials and Methods

This section provides detailed information on the object of study, from the methods of collecting necessary data, on which machine learning and artificial intelligence will be based.

The object of the study is soil samples. Field studies were conducted to collect data [10], and 70 points were carefully selected for sampling. Soil samples were collected in the territory of the village of Vshizh in the Zhukovsky district of the Bryansk region. They cover various topographical and climatic conditions. Soil samples were analyzed using legislatively approved methods.

In this specific case, the study is based on indicators of humus (%), P2O5 (mg/kg), and K2O (mg/kg). The chemical composition of the soil was analyzed using laboratory methods that ensure high accuracy of the obtained values.

In addition, satellite images from 25.05.2023 with a resolution of 10 meters per 1 pixel obtained from the Copernicus image database were used to assess the fertility of the soil in this area. On this internet resource, information about the soil condition based on multispectral imaging of the earth's surface can be obtained [11]. The images were obtained using modern Sentinel-2 satellite observation systems.

Predicting soil fertility, based on data obtained during the chemical analysis of soil samples, will be carried out using a multilayer model of a neural network based on the Rosenblatt perceptron. It consists of an input layer, hidden layers, and an output layer. Neurons in this model are interconnected and transmit data to each other [12].

The principle of the perceptron's operation is to train it on data examples. At the beginning of training, neuron weights are initialized to random values. The perceptron then takes input data and outputs a prediction. If the prediction is incorrect, neuron weights are adjusted using the error backpropagation algorithm.

For our study, the Rosenblatt perceptron model is suitable, as the multilayer model can be trained based on the provided data. Thus, this model demonstrates high adaptability. This is important because the environment, its components, and in particular the chemical composition of soils change over time. Consequently, for further development, it is not necessary to change parameters, but simply to adjust the model weights [13].

From the perspective of model development and data analysis, the choice of programming language is an important aspect that affects the effectiveness, reliability, and convenience of implementing the research. When predicting soil fertility using neural networks, the cross-platform statistical programming environment R will be used.

Firstly, R is designed specifically for statistical data analysis, meaning it is originally intended to work with large amounts of input data. Secondly, this programming environment is ideal for working with different neural network models. For this and most other tasks, there are numerous ready-made libraries. Thirdly, R contains convenient interactive tools. One of them is Rstudio with syntax highlighting, program text navigation, table sorting, and displaying graphs in a separate window [14]. In this research, this is extremely necessary, as

the Rosenblatt perceptron visually demonstrates the workings of the neural network. Another undeniable advantage of the R language is its prevalence among scientists in various fields.

3 Results and discussions

The results of the RStudio program are shown in the figures below. The diagram shows the Rosenblatt perceptron model for three substances studied in the soil – humus, P_2O_5, and K_2O, respectively, and the values of the sum of squared errors (which is the sum of the squares of the differences between the predicted and actual values) and the number of training iterations are presented.

In addition, the program calculated the correlation values between the predicted and calculated soil fertility.

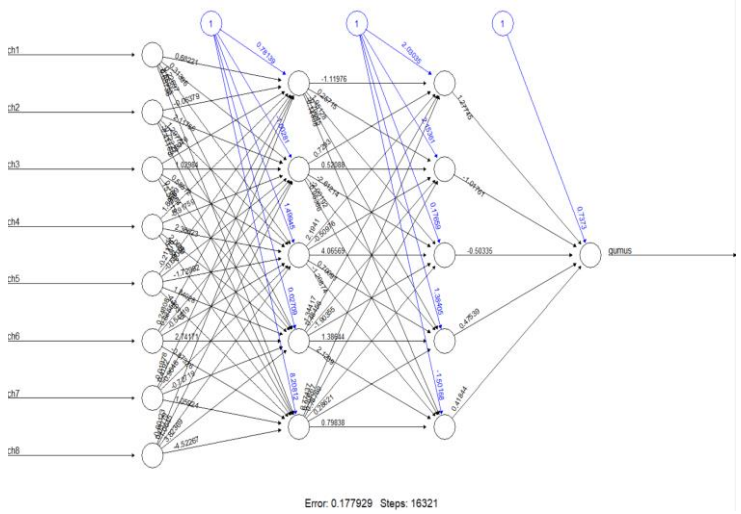

Fig. 1 The results of the neural network (input data – humus indicators)

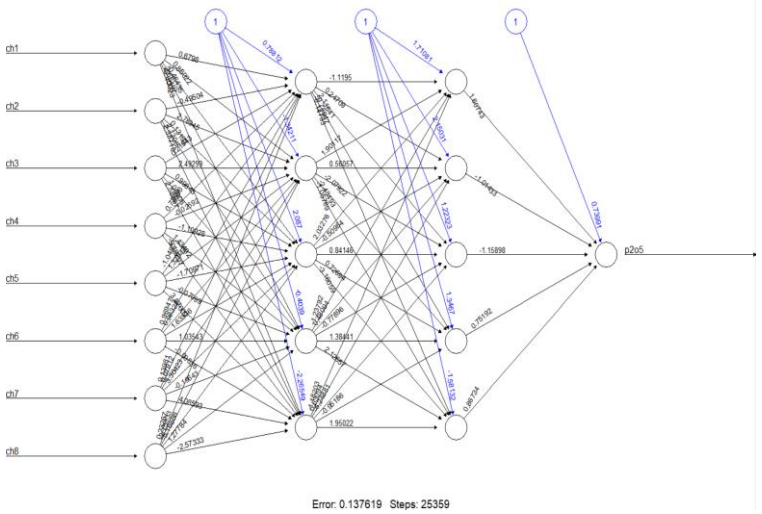

Fig.2 The results of the neural network (input data – P_2O_5 indicators)

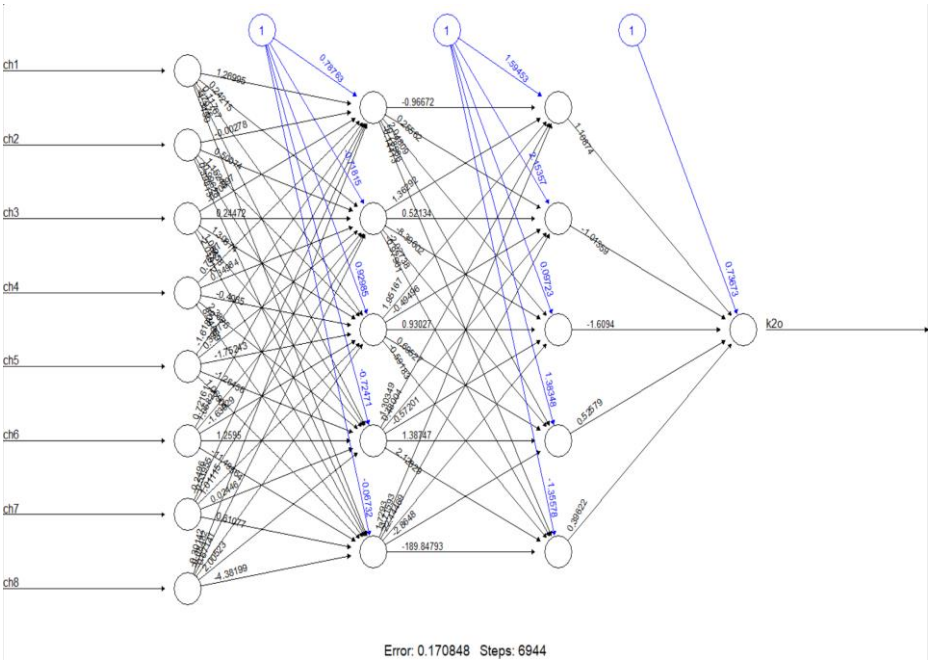

Fig.3 The results of the neural network (input data – K_2O indicators)

K(humus)=0,548;
K(P_2O_5)=0,287;
K(K_2O)=0,244;

According to the results of the study, the highest correlation value was obtained when predicting soil fertility based on the humus content, indicating the high accuracy of the neural network prediction. This is because humus has a distinct black and dark brown color that is clearly visible in satellite images. Thus, the neural network recognizes the values of color channels and based on them accurately predicts how fertile the soil is in a particular area. As for P_2O_5 and K_2O oxides, the forecasting efficiency is much lower due to the inability to accurately depict them in satellite images. These chemical substances can only be detected through field and laboratory research.

However, the effectiveness of the model's performance is better evaluated by the value of the total mean squared error. In this case, its value is very low. The lower the error value, the more effective the model. It can be used in cases where the studied factor can be precisely tracked using remote sensing data obtained through multispectral imaging. For example, in assessing land degradation [15], in studying plant productivity [16], and in evaluating the NDVI (Normalized Difference Vegetation Index).

The practical application of this method is as follows - the presented neural network model can help determine the optimal conditions for the growth of agricultural crops and thereby increase productivity. By analyzing data on soil composition and weather conditions, one can determine the optimal parameters for plant growth and development. This will help increase productivity, improve taste quality, and reduce the risk of plant diseases.

In addition to the areas described, the model can be used to forecast crop yields depending on a series of factors, including crop variety, fertilizers, irrigation, and climatic conditions. This can help agricultural enterprises to more effectively plan their activities, optimize resource usage, and increase profitability.

Therefore, using neural networks in agriculture can significantly improve the production process and make it more efficient and economically beneficial.

4 Conclusion

The study investigated the possibility of predicting soil fertility using neural networks. The process was based on remote sensing data obtained from satellite images. Seventy soil samples were analyzed in the laboratory for humus, P_2O_5, and K_2O content.

A neural network model based on the Rosenblatt perceptron was developed. Using this model, soil fertility was predicted based on the quantitative substance indicators described above using the statistical programming environment R. The model, based on remote sensing data, demonstrated a small magnitude of total mean square error values, indicating high accuracy of the predicted results.

At the same time, low correlation values were obtained for the research results for the P_2O_5 and K_2O indicators, unlike the humus indicators, for which the correlation value is average. Thus, the study established that for the most accurate analysis of soil fertility, it is necessary for the studied indicator to be clearly visible in multispectral satellite imagery, which is the basis of remote sensing data.

Neural networks play a huge role in agriculture. In general, the application of neural networks in agriculture can significantly improve plant growing processes, reduce costs for soil processing and fertilizers, as well as increase crop yield and product quality. Machine learning and data analysis methods are already actively used in agriculture and provide huge opportunities for the improvement of this industry.

References

1. Belyavskii A.E. Outer Loop Structure of a Spacecraft Thermal Control System with Heat Stores, **42(1)**, 60 – 62 (2022)
2. Lebedev, I., Ogorodnikov, S. Exploring the relationship between soil chemical composition and NDVI index using AI, **462**, 03041 (2023)
3. Sorokin A. E., Zhelonkin M. V., Ogorodnikov S. S. Monitoring of Agricultural Land Using Pilotless Aircraft, **42(12)**, 1315 – 1317 (2022)
4. Slavko V.D., Andreeva O.V. & Kust G.S. Assessment of the Dynamics of Land Cover at the Local Level to achieve Land Degradation Neutrality (for Desertified Lands of the Dry-Steppe Zavolzh'e), **13**, 50–58 (2023)
5. Sotnikova E.V., Sorokin A.E., Bulychev S.N. Assessing Predicting the Environmental Conditions of Megacities, **42 (1)**, 93–95 (2022)
6. Sishodia, R.P., Ray, R.L., Singh, S.K. Applications of remote sensing in precision agriculture: A review **12(19)**, 1 – 31 (2020)
7. J. Mohd, A. Haleem, I. H. Khan, R. Suman, Adv. Agrochem, **2**, 15-30 (2023)
8. Ogorodnikov S.S. Land Degradation Neutrality in the Tula region, **723(4)**, 042053 (2022)
9. Ali, Z.A., Yang, C., Israr, A., Zhu, Q. A Comprehensive Review of Scab Disease Detection on Rosaceae Family Fruits via UAV Imagery, **7(2)**, 97 (2023)
10. C. Jolivet, C. Ratié, D. Chavrit, N. Soler-Dominguez, A. Bispo, Encycl. Soils Environ. (Second Ed.), **4**, 551-567 (2023)
11. Copernicus // https://dataspace.copernicus.eu/ (date of the request: 13.02.2024)
12. Du, K.-L., Leung, C.-S., Mow, W.H., Swamy, M.N.S. Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era, **10(24)**, 4730 (2022)

13. de Zarzà, I., de Curtò, J., Calafate, C.T. Optimizing Neural Networks for Imbalanced Data, **12(12)**, 2674 (2023)
14. J. K. Kruschke, Doing Bayesian Data Anal. (Second Ed.), 33-70 (2015)
15. Ogorodnikov, S.S. The use of remote sensing data in the assessment of the Land Degradation Neutrality, **1010(1)**, 012040 (2022)
16. Ogorodnikov, S.S., Sorokin, A.E. Satellite Probing to Assess Soil Condition, **42(6)**, 639 – 641 (2022)