Exploring the relationship between soil chemical composition and NDVI index using AI

Ivan Lebedev[1*] and *Sergey* Ogorodnikov[1]

[1] Moscow Aviation Institute (National Research University), 125993, Russia, Moscow, Volokolamskoe shosse, 4

Abstract. This scientific article presents the results of research focused on developing a method for predicting the Normalized Difference Vegetation Index (NDVI) based on soil chemical composition using a multilayer artificial intelligence (AI) model. This method aims to improve the accuracy and predictive capability of land resource assessment, as well as the impact of chemical factors on vegetation. The study involved collecting soil chemical composition data in various conditions, providing a wide range of information for analysis. For NDVI assessment, a key indicator of vegetation condition, data from modern Earth observation satellite systems were used. The central aspect of the research is the multilayer AI model based on the Rosenblatt perceptron, capable of detecting complex nonlinear relationships between soil chemical parameters and NDVI. The training algorithm was tuned for maximum accuracy and generalization of results. The results show that the developed model provides high accuracy in NDVI predictions, making it an important tool for agriculture, ecology, and sustainable land use. These findings highlight the potential of using AI and soil data to optimize agricultural production, monitor ecosystems, and manage land resources.

1 Introduction

Modern agricultural production and ecological modeling face the need for adaptation to extensive changes in the environment. However, in all its diversity, efficient management of land resources and agriculture is challenged not only by stochastic external factors but also by changes in internal factors, including the physical and chemical properties of the soil.

Changes in the chemical composition of the soil and physical processes associated with the impact of anthropogenic and natural factors can significantly affect the state of soil fertility and, consequently, agricultural production. In this context, the need for advanced methods of monitoring and assessing land resources becomes an integral component of sustainable development.

Therefore, the integration of modern technological innovations, such as artificial intelligence (AI), for determining the chemical composition of the soil and its relationship with the Normalized Difference Vegetation Index (NDVI), represents a relevant research direction [1]. Based on data obtained from contemporary satellite observation systems,

* Corresponding author: lebedev.ivan.ig@yandex.ru

artificial intelligence has the potential to revolutionize the analysis of soil parameters and predict their impact on vegetation cover.

This article presents the results of research dedicated to the development of a method based on the use of artificial intelligence for accurate NDVI assessment based on soil chemical composition data. This approach focuses on the analysis of multidimensional data to improve prediction accuracy and deepen the understanding of the relationships between soil properties and vegetation cover. The results of this research can have important practical applications in agriculture, ecology, and sustainable land use.

2 Methodology

The central aspect of this research is an enhanced methodological approach developed for the comprehensive analysis of soil chemical composition and the computation of the Normalized Difference Vegetation Index (NDVI) using a multilayer artificial intelligence (AI) model with a root algorithm based on the Rosenblatt perceptron.

To ensure the accuracy and reliability of the research, random soil samples covering diverse topographical and climatic conditions were collected. Soil chemical composition analysis, with a focus on macro- and microelements, was carried out using laboratory methods that ensure a high degree of precision.

To assess NDVI, a key indicator of vegetation cover health, images obtained from modern satellite observation systems were used. Numerical values were computed by calculating the ratios between the difference in intensity of reflected light in the red and infrared ranges and their sum for each pixel of the image.

The research's focal point is the developed multilayer artificial intelligence model based on the Rosenblatt perceptron. This model includes several hidden layers capable of detecting and exploring complex, nonlinear dependencies in the data. The training algorithm was carefully tuned using training and test datasets to achieve maximum accuracy and generalization of results for new datasets while avoiding overfitting.

The results of the obtained model were validated using independent test data and cross-validation methods, allowing for an assessment of its adequacy and effectiveness. Statistical metrics and comparisons with existing NDVI assessment methods were used to determine the reliability and applicability of the developed methodology.

3 Data acquisition

In the context of this research, the accumulation of necessary data represents a fundamental stage that ensures scientific accuracy in further research. This section is dedicated to the methods of data collection and preparation, forming the basis for AI training.

Field research began with the careful selection of locations and subsequent soil sample collection [2]. The random sampling process allowed for a high degree of data representativeness, considering the diversity of natural and climatic conditions. Each sample underwent detailed individual characterization, resulting in a total of 104 samples.

The primary stage of the research involved the analysis of the soil's chemical composition. During the analysis, soil acidity, humus content, phosphorus oxide, potassium oxide, calcium, magnesium, nitrogen, ammonium, and nitrates were determined. The analysis itself was conducted in strict accordance with industry standards.

To reproduce the Normalized Difference Vegetation Index (NDVI), a crucial indicator of vegetation health, satellite data and images were used. However, before these data could be included in the analysis, they underwent comprehensive preprocessing [3]. This stage

included not only radiometric and geometric correction but also artifact removal, following modern data processing methodologies.

In the end, a data table for AI training was obtained, containing the following columns: NDVI, pHKCl, Humus, P2O5, K2O, Ca, Mg, N_NH4, N_NO3 – a complete set comprising input and output data.

4 Model development

To analyze the relationships between soil chemical composition and the Normalized Difference Vegetation Index (NDVI), a decision was made to select a multilayer artificial intelligence (AI) model based on the Rosenblatt perceptron [4]. This choice is driven by both theoretical and practical considerations and represents an optimal strategy for achieving the research objectives.

4.1 Rationale for model choice

Nonlinear regression approach: The Rosenblatt perceptron model excels at modeling complex, nonlinear dependencies in data. Given the numerous nonlinear relationships between soil composition and NDVI, such an approach seems logical.

Adaptability and learning: Multilayer models have the ability to learn from provided data, making them adaptable to changes and environmental dynamics. In the context of ecosystem monitoring and changes in soil chemical composition, this adaptability is crucial, as it only requires model weight adjustments rather than parameter changes for further development.

4.2 Practical implementation

In terms of model development and data analysis, the choice of programming language is an important aspect that affects the efficiency, reliability, and convenience of research implementation [5]. In this research, the decision was made to use the R programming language [6].

Firstly, R boasts a rich ecosystem of third-party libraries and packages specifically designed for data analysis and statistical computation. These packages provide researchers with powerful tools for data manipulation, including handling large datasets, visualization, statistical analysis, and machine learning [7].

Furthermore, R offers high flexibility and scalability. It is an interpreted language, allowing for easy iteration over data analysis and the development of complex algorithms. Additionally, R is capable of processing large datasets and dealing with high data dimensionality, which aligns with the requirements of this research, where the volumes of soil chemical composition and NDVI data can be significant.

An additional advantage of R is its academic and professional support. R is widely used in academic settings and among researchers, providing access to a rich and diverse source of documentation and educational materials. There is also an active professional user community for R, facilitating knowledge exchange and the dissemination of best practices in data analysis.

Finally, R provides powerful tools for creating high-quality graphics and data visualization, which is an important aspect when analyzing data and presenting research results.

4.3 Comparison with alternative methods

A comparative analysis was conducted using alternative data analysis methods such as linear regression and Support Vector Machine (SVM).

Unlike the Rosenblatt perceptron model, linear regression assumes linear relationships between variables, making it insufficiently flexible for modeling complex nonlinear dependencies characteristic of this task.

SVM has its advantages, but when dealing with large datasets and high feature dimensionality, which is typical for soil and NDVI data analysis, a more complex model may be required to capture the nonlinear relationships that linear methods might miss.

4.4 Conclusion

The choice of a multilayer model based on the Rosenblatt perceptron for analyzing soil chemical composition and NDVI data was justified both theoretically and practically. This choice allows for more effective modeling of complex nonlinear relationships, which is a crucial factor in the context of the current task. Compared to alternative methods such as linear regression and SVM, the model based on the Rosenblatt perceptron proves to be more adaptive and suitable for complex datasets, making it the preferred choice for this research. Figure 1 displays the resulting neural network parameters.

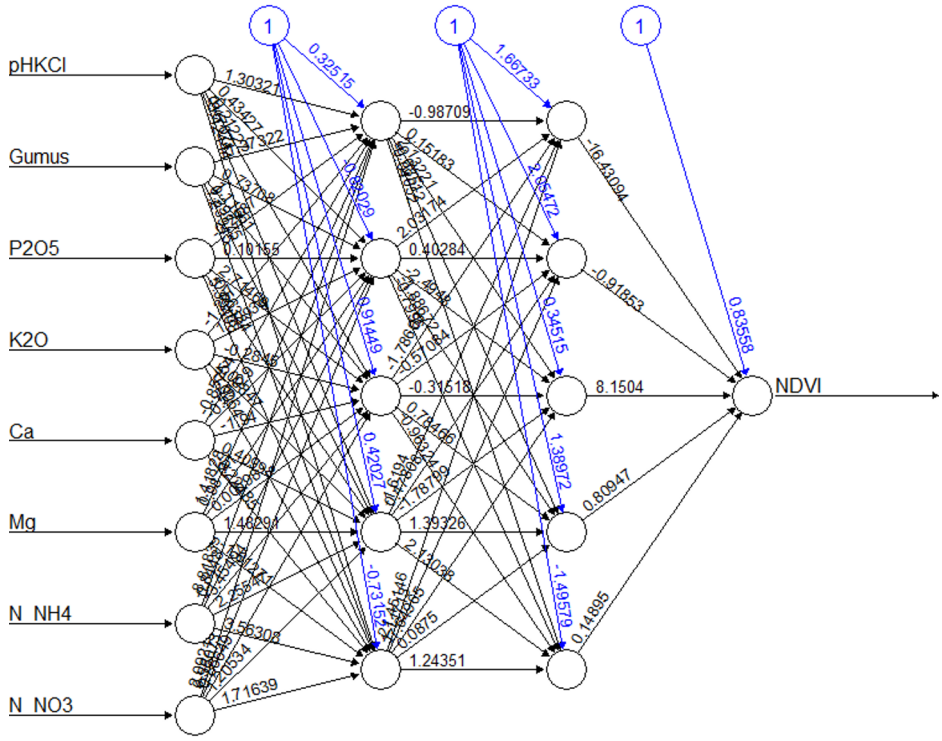

Fig. 1. The resulting neural network

As shown in the presented image (Figure 1), the number of perceptrons in the input layer matches the number of input parameters (similarly for the output layer). Such model architecture, along with two hidden layers, yielded the best results and responsiveness to changes in input data.

5 Practical application

In agriculture, predicting NDVI based on soil chemical properties allows for the optimization of fertilizer and resource use, leading to increased crop yields and reduced costs. This helps improve the sustainability and efficiency of agricultural production.

In an ecological context, the model enables the assessment of the impact of chemical factors in the soil on ecosystem health and vegetation. This is important for ecological monitoring, nature conservation, and the identification of potential environmental threats.

Predicting risks such as droughts, wildfires, or soil waterlogging becomes more accurate and effective with NDVI predictions based on soil chemical data.

Land resource management becomes more efficient and sustainable through the analysis of soil chemical composition and related NDVI data. Optimized land use contributes to biodiversity conservation and soil erosion prevention.

In urban environments, NDVI analysis can be used for planning green areas [8], parks, and gardens, improving the quality of urban life and addressing environmental issues.

The research results can also be applied in commercial and investment projects related to land use and property valuation for real estate investments [9].

Additionally, this model can be used for scientific research and educational purposes to delve deeper into the relationships between soil and plant processes.

6 Results

As a result of this research, a multilayer artificial intelligence model based on the Rosenblatt perceptron was developed, specifically aimed at predicting the Normalized Difference Vegetation Index (NDVI) based on provided soil chemical composition data. It demonstrated high accuracy in predicting NDVI in this chemical context, considering the complex and nonlinear relationships present between soil composition and vegetation condition.

It is important to note that multilayer artificial intelligence models prove to be reliable tools for data analysis, capable of learning from provided data and capturing complex patterns, which is critical in tasks requiring high accuracy and predictive capability. In this case, the model not only accurately predicted NDVI but also explored the relationships between specific chemical elements in the soil and this index.

As a result, not only a highly valuable practical tool was obtained but also valuable scientific data on how various chemical factors affect the health of the plant world [10]. This is of great importance for various fields, such as agriculture, where predicting soil and vegetation conditions can assist in resource management and production optimization.

Comparison of the developed model with alternative data analysis methods, such as linear regression and Support Vector Machine, confirmed the superiority of the chosen model, thanks to its ability to identify complex nonlinear dependencies in the data. Thus, the choice of a multilayer model based on the Rosenblatt perceptron proved to be well-founded and demonstrated its effectiveness.

7 Future research directions

Future research directions in the field of predicting NDVI based on soil chemical properties using multi-layer artificial intelligence models offer numerous potential avenues for development. First and foremost, there is potential for improving the models themselves, which includes further research into various neural network architectures and learning algorithms. This will allow for achieving higher accuracy and generalization when predicting NDVI.

Another important direction is the integration of additional data, such as climate and geographical parameters, which can enhance the predictive capabilities of the models. This will expand the understanding of the influence of different factors on NDVI and its long-term trends.

Additionally, research should focus on scalability and the transferability of results and models so that they can be applied in different regions and ecosystems.

Furthermore, with the data obtained, ongoing research in the field of ecosystem sciences and the development of practical projects in land resource management, agriculture, and environmental monitoring can be pursued. Research can contribute to the development of geospatial technologies for more accurate and sensitive NDVI data analysis over large territories.

8 Acknowledgements

The research was carried out at the expense of the grant of the Russian Science Foundation No. 23-74-01050, https://rscf.ru/en/project/23-74-01050

References

1. J. Mohd, A. Haleem, I. H. Khan, R. Suman, Adv. Agrochem, **2**, 15-30 (2023)
2. C. Jolivet, C. Ratié, D. Chavrit, N. Soler-Dominguez, A. Bispo, Encycl. Soils Environ. (Second Ed.), **4**, 551-567 (2023)
3. K. Vasiliev, V. Dementiev, N. Andriyanov, Procedia Comput. Sci., **126**, 49-58, (2018)
4. A. Patel, A. Kethavath, N.L. Kushwaha, A. Naorem, M. Jagadale, K.R. Sheetal, P.S. Renjith, Eng. Appl. Artif. Intell., **123**, part B (2023)
5. Y. Chen, L. Li, W. Li, Q. Guo, Z. Du, Z. Xu, AI Comput. Syst., 271-377 (2024)
6. J. Sun, M. Tang, Syst. Med., **1**, 1-8 (2021)
7. J. K. Kruschke, Doing Bayesian Data Anal. (Second Ed.), 33-70 (2015)
8. T.K. Khoshtaria, N.T. Chachava, Ann. Agrar. Sci., **15**, 217-223 (2017)
9. G. Debrunner, D. Kaufmann, Land Use Policy, **132** (2023)
10. J. Yan, L. Lou, W. Bai, S. Zhang, N. Zhang, **900** (2023)