BIO Web of Conferences **141**, 01027 (2024) https://doi.org/10.1051/bioconf/202414101027
AGRICULTURAL SCIENCE 2024

Intelligent approaches to wheat grain classification using neural networks in the agricultural sector

Anastasia Kozlova[1,2*], *Elena* Khudyakova[3], *Vadim* Tynchenko[1,2], *Marina* Stepantsevich[3], and *Mikhail* Nikanorov[3]

[1]Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia
[2]Bauman Moscow State Technical University, 105005 Moscow, Russia
[3]Moscow Timiryazev Agricultural Academy, Russian State Agrarian University, 127550 Moscow, Russia

Abstract. This work is dedicated to conducting a comprehensive analysis of a wheat dataset to identify significant attributes for accurate grain classification. The initial dataset contains various parameters of wheat grains, such as length, perimeter, area, compactness, and asymmetry coefficient. The focus of the study is on analyzing the relationships between these attributes and their impact on the classification target field. Initially, data normalization was performed to eliminate the influence of scale differences between variables. This was followed by a correlation analysis, which revealed several key relationships between attributes. Specifically, it was found that the asymmetry coefficient has a moderate positive correlation with the classification target, while the attributes area, compactness, perimeter, and width exhibit a moderate negative correlation. Length, on the other hand, shows a weak negative correlation with the target attribute. To gain a deeper understanding of the data structure, Kohonen Self-Organizing Maps were used, which helped to identify three clusters. The analysis revealed that the most significant attributes for clustering are compactness, width, perimeter, and area, while the asymmetry coefficient was found to be the least significant. In conclusion, two classification models were built and evaluated. The first model included all attributes from the dataset and demonstrated an accuracy of 0.97. The second model used a subset of attributes excluding the asymmetry coefficient and showed a slightly higher accuracy of 0.98. These results confirm that excluding less significant attributes can lead to a minor but noticeable improvement in model accuracy. Overall, the work highlights the importance of selecting the right attributes to enhance the effectiveness of classification models.

1 Introduction

Wheat grain classification is an integral part of modern agriculture, significantly impacting all aspects of cereal production and management. In the context of global agricultural sector

* Corresponding author: ankoz9@yandex.ru

globalization and diverse climatic conditions, classification becomes a key tool for achieving efficiency and sustainability in wheat production. This system not only simplifies interactions between producers, processors, and consumers but also plays a central role in optimizing agronomic practices and enhancing overall productivity.

Wheat, being one of the primary food crops, holds immense importance for food security and economic stability in many countries. However, effective management of its cultivation and processing requires consideration of not only the total amount of produced grain but also its quality. Wheat grain classification allows for the detailed differentiation of various varieties and types based on parameters such as protein content, gluten, grain size and shape, as well as other important characteristics [1-4].

For agriculture, wheat classification is a crucial tool for selecting optimal varieties that match specific growing conditions, including climate, soil type, and agronomic practices. This is particularly relevant in the context of climate change and resistance to diseases and pests. Proper variety selection can significantly impact yield, resilience to adverse conditions, and, consequently, the economic efficiency of agricultural production [5-7].

Furthermore, wheat classification facilitates more accurate and fair grain trading on international markets. Standardized classification criteria simplify export and import processes, reducing the risk of conflicts between trading partners and ensuring transaction transparency [8-10]. This is vital for maintaining price stability and predictability, which, in turn, affects the economy of rural areas and food security.

2 Materials and methods

Correlation analysis is a statistical method used to determine and assess the degree of relationship between two or more variables. It helps to identify whether a connection exists, its strength, and whether it is positive or negative. In this study, correlation analysis was employed to explore relationships between wheat grain parameters, aiming to determine the significance and interrelationship of attributes in the selected dataset [11-14].

To support and refine the results from correlation analysis, Kohonen Self-Organizing Maps (SOM) were used. SOM is a clustering tool that helps uncover structures and relationships in multidimensional data by visualizing it in a two-dimensional grid of neurons. Each neuron represents a cluster, and the proximity of neurons reflects the similarity of the data they represent [15].

The next step involved modeling using neural networks, a powerful tool for analyzing data and solving complex problems. Neural networks, inspired by the human brain, learn from data to perform tasks such as classification, regression, and clustering. They consist of artificial neurons organized into layers, with different types of networks suited to various tasks—feedforward networks for classification, convolutional networks (CNN) for image processing, and recurrent networks (RNN) for sequential data [16-19].

Neural network modeling requires substantial data and computational resources but can achieve high accuracy and efficiency, particularly in tasks where traditional methods might fall short. Neural networks excel in pattern recognition, automatic translation, text generation, and more. Thus, they are crucial for modern data analysis and machine learning, enabling the detection of complex dependencies and precise predictions [20].

3 Data structure

The dataset contains 8 attributes:

1) area – the cross-sectional area of wheat grain, measured in square units, which helps determine grain size.

2) perimeter – the boundary length of wheat grain, measured in linear units, useful for assessing grain shape.

3) compactness – a measure calculated as the ratio of the area of the grain to the square of its perimeter, indicating how close the grain is to a perfect shape, like a circle. a high compactness coefficient suggests a more rounded grain shape.

4) length – the length of wheat grain, measured in linear units, important for evaluating grain size and quality.

5) width – the width of wheat grain, measured in linear units, used alongside length to determine overall grain shape.

6) asymmetry coefficient – measures the degree to which the grain deviates from a symmetrical shape. high values indicate more pronounced irregularities or defects.

7) groove length – the length of grooves or indentations on the grain's surface, which can be significant for evaluating grain type, quality, and processing.

8) category – classification of wheat grains based on criteria such as variety, quality, or intended use.

The data structure is illustrated in Figure 1.

№	Name	Histogram	Min	Max	Average	Gaps
1	9.0 area		10,59	21,18	14,847...	0
2	9.0 perimeter		12,41	17,25	14,559...	0
3	9.0 compactness		0,8081	0,9183	0,8709...	0
4	9.0 length		4,899	6,675	5,6285...	0
5	9.0 width		2,63	4,033	3,2586...	0
6	9.0 asymmetry co...		0,7651	8,456	3,7002...	0
7	9.0 groove length		4,519	6,55	5,4080...	0
8	9.0 category		1	3	2	0

Fig. 1. Data structure

4 Results

The first step was data normalization, which helps ensure that data is structured and organized in a specific format. This avoids redundant and excessive data, simplifying processing and analysis. Structured data also aids in more accurate pattern and trend identification, making it easier to detect anomalies and improving the quality of analytical conclusions [21-24]. Thus, normalization makes data more orderly and easier to analyze, ultimately enhancing the accuracy and reliability of analytical results. Figure 2 shows the correlation table reflecting the results of the correlation analysis [25-27].

	area_n	asym...	category	comp...	groov...	length...	perim...	width_n
area_n	1,00	-0,20	-0,33	0,61	0,86	0,94	0,98	0,96
asymmetry...	-0,20	1,00	0,56	-0,31	0,01	-0,15	-0,20	-0,23
category	-0,33	0,56	1,00	-0,53	0,03	-0,26	-0,33	-0,42
compactne...	0,61	-0,31	-0,53	1,00	0,23	0,36	0,53	0,76
groove_len...	0,86	0,01	0,03	0,23	1,00	0,91	0,88	0,73
length_n	0,94	-0,15	-0,26	0,36	0,91	1,00	0,96	0,84
perimeter_n	0,98	-0,20	-0,33	0,53	0,88	0,96	1,00	0,93
width_n	0,96	-0,23	-0,42	0,76	0,73	0,84	0,93	1,00

Fig. 2. Correlation table

The analysis revealed that most attributes have significant positive or negative correlations with each other. Regarding relationships with the target attribute, the asymmetry coefficient has a moderate positive correlation with the category field. In contrast, the fields area, compactness, perimeter, and width exhibit a moderate negative correlation with the target field. The length attribute shows a weak negative correlation with the category attribute [28-31].

Next, Kohonen Self-Organizing Maps (SOM) were applied to the data. The data was divided into three clusters, with the profiles presented in Figure 3 and the breakdown by absolute frequency shown in Figure 4.

#	Name	Total	Cluster 2	Cluster 0	Cluster 1
		100%	34%	33%	33%
1	9.0 compactness_n	74	59	58	59
2	9.0 width_n	73	67	68	62
3	9.0 perimeter_n	73	69	65	61
4	9.0 area_n	72	70	66	62
5	9.0 length_n	67	69	60	55
6	9.0 groove_length_n	66	70	47	46
7	9.0 asymmetry_coe...	44	33	43	39

Fig. 3. Clustering using Kohonen Self-Organizing Maps

The most significant attribute for the clusters was compactness, while the least significant was the asymmetry coefficient.

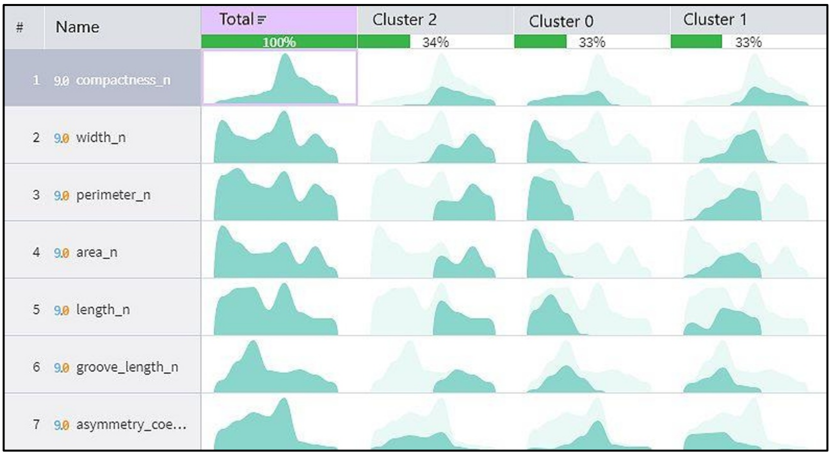

Fig. 4. Absolute frequencies of clusters

Two models were constructed using neural networks. The performance results are presented in Figures 5 and 6.

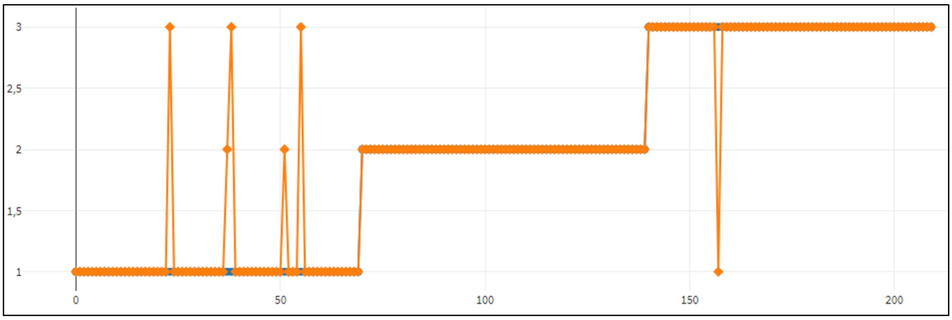

Fig. 5. The first data model

The first data model utilized all attributes from the dataset. The accuracy of this model was 0.97, which is a very high score [32].

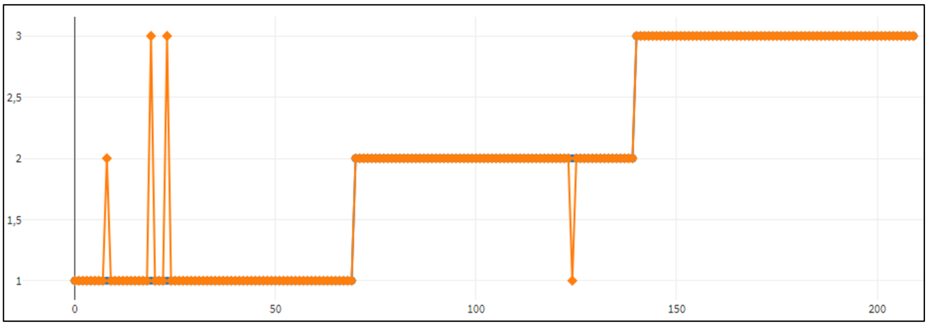

Fig. 6. The second data model

The second model used a subset of attributes excluding the asymmetry coefficient. This model achieved an accuracy of 0.98, slightly better than the first model.

5 Conclusion

Wheat grain classification is a fundamental aspect of modern agriculture, significantly impacting all its areas. It not only simplifies the selection of suitable varieties, ensuring they meet specific growing conditions and market requirements, but also aids in effective resource management and enhances overall productivity. In the context of global climate change and increasing demands for food security, accurate classification becomes a vital tool for ensuring sustainability and adaptation in agricultural production.

This system plays a key role in improving grain quality, which directly affects its use in various products and the final product reaching consumers. Classification facilitates the optimization of trading processes, simplifying export and import operations, crucial for maintaining price stability and transparency in international markets. It also helps farmers make informed decisions, leading to more efficient and profitable farming.

In a scientific context, wheat grain classification forms the foundation for research and development of new varieties capable of adapting to changing conditions and demands. This opens opportunities for implementing innovative agronomic practices and technologies that improve agricultural production.

Thus, systematic wheat grain classification ensures sustainability, efficiency, and innovative development in agriculture. It plays an indispensable role in shaping the future of the agricultural sector, equipping it to meet contemporary challenges and satisfy the growing needs of the global population.

The analysis of the wheat dataset determined the significance of various attributes for grain classification. Data normalization and correlation analysis revealed key relationships between attributes, specifically a moderate positive correlation of the asymmetry coefficient with the target field and negative correlations of other attributes. Kohonen Self-Organizing Maps showed that compactness, width, perimeter, and area are crucial for clustering, while the asymmetry coefficient was least significant. Modeling confirmed that excluding the asymmetry coefficient from the attribute set increased classification accuracy from 0.97 to 0.98. These results highlight the importance of considering significant attributes and optimizing the input data set to achieve maximum accuracy in wheat grain classification tasks.

References

1. A. Boyko, A. Kukartsev, T. Dolgova, K. Lobkov, A. Bagaeva, Simulation model of the urban population taking into account the influence of the manufacturing sector based on big data, 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE, 15 March 2023, DOI:10.1109/infoteh57020.2023.10094081.
2. K. Lobkov, D. Eremeev, A. Rubinskaya, E. Melnikova, I. Panfilov, Determination of the degree of impact of natural disasters on the level of migration of the population by simulation modelling, 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE, 15 March 2023, DOI:10.1109/infoteh57020.2023.10094109.
3. I.A. Panfilov, N.A. Shepeta, A.A. Stupina, A.A. Boyko, V.V. Kondratiev, The electrochemical technology of gold extraction from sulfide ores, MIAB. Mining Inf. Anal. Bull., 2023;(11-1):226-238. [In Russ]. DOI:10.25018/0236_1493_2023_111_0_226.

4. V. V. Gerasidi, R. G. Dubrovin, V. V. Kukartsev, T. A. Panfilova, G. V. Stas, Boost system diagnostic parameters of coherent gas piston installations of mining enterprises, Series of Geology and Technical Sciences, 15 August 2024, DOI: 10.32014/2024.2518-170x.423.
5. V. O. Gutarevich, N. V. Martyushev, R. V. Klyuev, V. A. Kukartsev, V. V. Kukartsev, L. V. Iushkova, L. N. Korpacheva, Reducing oscillations in suspension of mine monorail track, Applied Sciences, 7 April 2023, DOI: 10.3390/app13084671.
6. B. V. Malozyomov, N. V. Martyushev, V. A. Kukartsev, V. V. Kukartsev, S. V. Tynchenko, R. V. Klyuev, N. A. Zagorodnii, Y. A. Tynchenko, Study of supercapacitors built in the start-up system of the main diesel locomotive, Energies, 1-15 (2023), DOI:10.3390/en16093909.
7. K. Kravtsov, I. Panfilov, E. Filyushina, E. Vaitekunene, D. Tikhonenko, Creation of multi-link automatic parameter control systems at nuclear power plants, 2024 12th International Conference on Smart Grid (icSmartGrid), IEEE, 1-6 (2024), DOI:10.1109/icsmartgrid61824.2024.10578151.
8. I. I. Bosikov, N. V. Martyushev, R. V. Klyuev, I. A. Savchenko, V. V. Kukartsev, V. A. Kukartsev, Y. A. Tynchenko, Modeling and complex analysis of the topology parameters of ventilation networks when ensuring fire safety while developing coal and gas deposits, Fire, 1-15 (2023), DOI:10.3390/fire6030095.
9. A. Shutaleva, N. Martyushev, Z. Nikonova, I. Savchenko, V. Kukartsev, V. Tynchenko, Y. Tynchenko, Sustainability of inclusive education in schools and higher education: teachers and students with special educational needs, Sustainability, 7 February 2023, DOI:10.3390/su15043011.
10. V. Kukartsev, A. Gantimurov, K. Kravtsov, A. Borodulin, Y. Tynchenko, Application of non-parametric learning method in soil suitability assessment in present day economy, Journal of Infrastructure, Policy and Development, 1 August 2024, DOI:10.24294/jipd.v8i7.4074.
11. T. Panfilova, V. Kukartsev, V. Tynchenko, Y. Tynchenko, O. Kukartseva, I. Kleshko, X. Wu, I. Malashin, Flood susceptibility assessment in urban areas via deep neural network approach, Sustainability, 29 August 2024, DOI:10.3390/su16177489.
12. V. Kukartsev, K. Kravtsov, O. Stefanenko, N. Podanyov, A. Bezvorotnykh, Using machine learning techniques to simulate network intrusion detection, 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS), 3 May 2024, DOI: 10.1109/iscs61804.2024.10581097.
13. K. Degtyareva, D. A. Ageev, V. V. Kukartsev, Finding patterns in employee attrition rates using self-organizing Kohonen maps and decision trees, 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 14 December 2023, DOI: 10.1109/icses60034.2023.10465548.
14. A. Gladkov, V. Kukartsev, A. Kozlova, D. Grigorev, Development of requirements for AIS aimed at controlling high turnover, 2023 IEEE International Conference on Computing (ICOCO), 9 October 2023, DOI: 10.1109/icoco59262.2023.10397670.
15. Y. Tynchenko, V. Kukartsev, V. Tynchenko, O. Kukartseva, T. Panfilova, A. Gladkov, V. Nguyen, I. Malashin, Landslide assessment classification using deep neural networks based on climate and geospatial data, Sustainability, 17 August 2024, DOI:10.3390/su16167063.
16. T. Panfilova, V. Tynchenko, V. Nelyub, A. Glinscaya, A. Borodulin, A. Gantimurov, Y. Tynchenko, Introducing machine learning to analyze factors aimed at successful

development of the individual social qualities, Journal of Infrastructure, Policy and Development, 6 August 2024, DOI:10.24294/jipd.v8i8.4165.
17. M. Savchenko, V. Tynchenko, Unsupervised production machinery data labeling method based on natural language processing, 2024 International Russian Smart Industry Conference (SmartIndustryCon), IEEE, 25 March 2024, DOI:10.1109/smartindustrycon61328.2024.10515763.
18. D. Strateichuk, N. Martyushev, R. Klyuev, V. Gladkikh, V. Kukartsev, Y. Tynchenko, A. Karlina, Morphological features of polycrystalline CdS1−xSex films obtained by screen-printing method, Crystals, 1-15 (2023), DOI:10.3390/cryst13050825.
19. N. V. Martyushev, B. V. Malozyomov, I. H. Khalikov, V. A. Kukartsev, V. V. Kukartsev, V. S. Tynchenko, Y. A. Tynchenko, M. Qi, Review of methods for improving the energy efficiency of electrified ground transport by optimizing battery consumption, Energies, 1-15 (2023), DOI:10.3390/en16020729.
20. V. A. Kukartsev, V. V. Kukartsev, V. S. Tynchenko, S. O. Kurashkin, R. B. Sergienko, S. V. Tynchenko, I. A. Panfilov, S. V. Eremeeva, T. A. Panfilova, Study of the influence of the thermal capacity of the lining of acid melting furnaces on their efficiency, Metals, 1-15 (2023), DOI:10.3390/met13020337.
21. N. V. Martyushev, D. A. Bublik, V. V. Kukartsev, V. S. Tynchenko, R. V. Klyuev, Y. A. Tynchenko, Y. I. Karlina, Provision of rational parameters for the turning mode of small-sized parts made of the 29 NK alloy and beryllium bronze for subsequent thermal pulse deburring, Materials, 1-15 (2023), DOI:10.3390/ma16093490.
22. V. Brigida, V. I. Golik, E. V. Voitovich, V. V. Kukartsev, V. E. Gozbenko, V. Y. Konyukhov, T. A. Oparina, Technogenic reservoirs resources of mine methane when implementing the circular waste management concept, Resources, 17 February 2024, DOI: 10.3390/resources13020033.
23. B. V. Malozyomov, N. V. Martyushev, V. V. Kukartsev, V. Y. Konyukhov, T. A. Oparina, N. S. Sevryugina, V. E. Gozbenko, V. V. Kondratiev, Determination of the performance characteristics of a traction battery in an electric vehicle, World Electric Vehicle Journal, 12 February 2024, DOI: 10.3390/wevj15020064.
24. V. Orlov, V. Tynchenko, A. Nizameeva, D. Shalaeva, D. Ageev, Development of a multifunctional cross-platform system for automation of energy data and resource management, E3S Web of Conferences, 2023, DOI:10.1051/e3sconf/202346007002.
25. K. Kravtsov, V. Tynchenko, E. Semenova, D. Shalaeva, I. Pinchuk, Workflow automation and performance improvement based on PostgreSQL, E3S Web of Conferences, 2023, DOI:10.1051/e3sconf/202345809022.
26. N. Fedorova, I. Panfilov, E. Moiseeva, K. Degtyareva, E. Markovskaya, Analytical methods and tools for business process optimization, BIO Web of Conferences, 2024, DOI:10.1051/bioconf/202411305009.
27. T. Panfilova, V. Tynchenko, O. Kukartseva, A. Kozlova, A. Glinscaya, Modernization of electronic document management and systems analysis processes using an automated platform, E3S Web of Conferences, 2024, DOI:10.1051/e3sconf/202454909018.
28. N. V. Martyushev, B. V. Malozyomov, I. H. Khalikov, V. A. Kukartsev, V. V. Kukartsev, V. S. Tynchenko, Y. A. Tynchenko, M. Qi, Review of methods for improving the energy efficiency of electrified ground transport by optimizing battery consumption, Energies, 8 January 2023, DOI: 10.3390/en16020729.
29. V. A. Kukartsev, V. V. Kukartsev, V. S. Tynchenko, S. O. Kurashkin, R. B. Sergienko, S. V. Tynchenko, I. A. Panfilov, S. V. Eremeeva, T. A. Panfilova, Study of the

influence of the thermal capacity of the lining of acid melting furnaces on their efficiency, Metals, 8 February 2023, DOI: 10.3390/met13020337.

30. N. V. Martyushev, D. A. Bublik, V. V. Kukartsev, V. S. Tynchenko, R. V. Klyuev, Y. A. Tynchenko, Y. I. Karlina, Provision of rational parameters for the turning mode of small-sized parts made of the 29 NK alloy and beryllium bronze for subsequent thermal pulse deburring, Materials, 30 April 2023, DOI: 10.3390/ma16093490.
31. V. Kukartsev, K. Kravtsov, Y. Tynchenko, T. Panfilova, Influence of mountain factors on salt excess and soil toxicity in mountain conditions, Sustainable Development of Mountain Territories, 30 September 2023, DOI: 10.21177/1998-4502-2023-15-3-784-797.
32. K. Yelemessov, M. Isametova, N. Saydinbayeva, V. Kukartsev, Mathematical and computer modeling of gantry crane load-beam system oscillation, Sustainable Development of Mountain Territories, 30 June 2023, DOI: 10.21177/1998-4502-2023-15-2-450-461.