Analysis of seed quality indicators based on neural network

N N Barysheva*, M V Guner1, D D Baryshev1 and S P Pronin2

1Department of Informatic Systems in Economics, Polzunov Altai State Technical University, 46 Lenin ave., Barnaul, 656038, Russia
2 Department of Informatic Technologies, Polzunov Altai State Technical University, 46 Lenin ave., Barnaul, 656038, Russia
*Corresponding author’s e-mail: barysheva.n.nik@gmail.com

Abstract. The paper presents a neural network study of the data of wheat seed quality. It is established that the analysis of bioelectrical signals of wheat seeds based on a neural network can be used in practice for the solution of two problems - diagnostics of seed material quality and the evaluation of cleaning line quality (separation into fractions). The paper presents the results of initial data preparation, formation of a neural network, analysis of training data for two practical problems of classification. It was established using a neural network that there is a nonlinear dependence of the membrane potential maximum value and the signal rise time on the seed yield. The model makes it possible to predict yield in terms of the seed material quality. A nonlinear dependence of the maximum membrane potential, the signal rise time of wheat seeds and the seeds variety to one or another faction (speed of separation into the fractions in this example) was also established in this paper. Studies have shown that the seeds variety is an important informative feature for solving the problem of classifying seeds by fractions. Therefore, it is necessary to conduct additional studies with other wheat seeds varieties to apply this method in practice.

1. Introduction
A reliable determination of wheat seed quality in the shortest possible time is an actual problem for agricultural enterprises. The seed quality is one of the main factors for determine future productivity. The seeds quality consists of three components. There are yield properties, a sowing quality and a varietal factor.

Sowing quality is determined by such indicators as laboratory germination, vigour etc. [1, 2, 3]. There are methods and tools for evaluating these indicators [2 - 6]. It is important for choose a method, that the indicators characterize not only the sowing quality, but also the yield properties of the seeds as a seed factor.

One of these methods is the method of diagnostic seed quality by bioelectric signals of wheat seeds [7]. The authors conducted an experimental study of seeds with different sowing and yield qualities. A high correlation of bioelectric signals with seed quality indicators has been established [7].

The traditional method of increasing yield properties is the seeds separation into fractions [8]. Analysis of the bioelectric signals of the seeds, divided into fractions, showed that the signals reflect the quality of each fraction.

It is possible to solve the inverse problems such as the analysis of wheat seeds quality and the cleaning line quality using bioelectric signals.

Recently, machine-learning methods, such as neural networks, decision trees and support vector machines are gaining popularity for solving classification and forecasting problems in agriculture [9, 10].
In this regard, the purpose of this study is to analyze the ability of neural networks to determine according the bioelectric signals the quality class of wheat seeds, divided into fractions.

The results of this study provide useful information for solving two problems in practice - diagnostic the quality of seed material and assessing the quality of the cleaning line (separation into fractions).

2. Materials and methods
We used for the study the bioelectric signals data of four varieties different quality wheat seeds. Seeds are represented by three different farms in the Altai Territory (LLC Wirth, LLC Rassiya, Chistyunsky farm). All wheat varieties grown in 2019 represent a fresh crop.

Research objectives:
- make a model of a neural network for diagnostic the quality of seed material based on the results of studies of seed's bioelectric signals;
- identify and analyse the significance of each factor of the model;
- create mechanisms for integrating the constructed model with applied agricultural systems for its wider testing in practice.

2.1. Method of selection and separation of seeds into fractions for experiment
Sampling for the study was performed in accordance with GOST 12036-85. It was selected a sample of each seeds variety for further separation into fractions.

It was chosen the method of separation by aerodynamic characteristics using the K-93 laboratory sailing classifier.

Air speed separation ranged from 8 m/s to 11 m/s in 1 m/s increments.

So, sixteen seed samples were obtained for research (four samples with different aerodynamic properties for each variety).

One hundred seeds were selected in each sample.

2.2. Method of bioelectric signals diagnostics
The method presents three main steps - seed preparation for research, signal measurement and data processing. Preparation for research represents seeds germination in a thermal chamber in distilled water at 20 °C for 14 hours.

Measurement is the following procedure. Grain is placed in an electrode holder; the grain shell is pierced by a needle electrode, with the help of ADC the signal is recorded for 5 seconds. The system of bioelectric signals diagnostics is presented in detail in paper [7]. As a result of measurements an array of data is obtained.

2.3. Artificial neural network
Artificial neural networks have proven to be a reliable tool for extracting knowledge from data sets. They are successfully used for solving complex, weakly formalized problems and allow working with noisy data and finding nonlinear dependencies in them.

It was used multilayer perceptrons in the present study. In these neural networks, there are as many neurons on the input layer as there are inputs in the task, and on the output layer there are as many outputs. In multilayer perceptrons there are usually one or several hidden layers, the number of neurons on which is a setting parameter.

All data was randomly divided into two subsets - training and test examples. The ratio between the number of examples in the first and second subsets was 85 by 15 percents. Accordingly, the neural networks were trained on the training subset; their testing was conducted on the test subset. Based on the results, conclusions are drawn and adequacies of the entire model are evaluated.

The test set was not used in the process of the network training, determining the stop training moment and determining additional parameters. Therefore, the quality of the results shown by the network on the examination set is a good assessment for new unknown data.

It is established that the analysis of the provided indicators allows solving several practical problems.
2.3.1. **Practice problem No. 1 - Classification by quality indicators (yield properties).** It was used the neural network with one hidden layer of five neurons to solve this problem. The neural network was trained using the backpropagation method.

2.3.2. **Practice problem No. 2 - Classification by fractions to assess the quality of the cleaning line.** Computational experiments were carried out by neural networks with one and two hidden layers to solve this problem. The total number of neurons ranged from 2 to 20. The method of training networks is the same - backpropagation.

3. **Results and discussions**

Data research by neural network provides for the following steps - preparation of initial data, formation of a neural network, preliminary training of the network, and analysis of training results. Consider each of the steps to solve the classification problems.

3.1. **Preparation of training data and selection of informative features**

Bioelectric signals represent a change in voltage over time, the so-called membrane potential. All signals represent an additive signal-to-noise mixture after measurements. Interference reduces the signal-to-noise ratio and reduces the usefulness of the data. It is possible to increase the reliability of information by applying pre-processing methods [13]. Currently, there are various methods of pre-processing [14]. Analysis of the resulting mixture showed that quantization noise and high frequency harmonic interference are superimposed on the useful signal. Noise is superimposed to a lesser degree, and noise is more imposed. Therefore, it is advisable to apply the low-pass filtering method (moving average method) [15].

Figure 1(a) shows typical graphs of changes in membrane potential after treatment. Since the graphs of changes in the membrane potential have time offsets, it was decided to allocate informative parameters for further neural network modeling. Informative parameters are presented in figure 1b. This is the maximum value of the membrane potential (MaxMP) and the rise time of the membrane potential from the value of 0.9·MaxMP to the value of MaxMP.

![Figure 1](image.png)

**Figure 1.** Graphs of the change in the signal of wheat seeds varieties "Thassos" (a) and informative parameters (b).

Figure 2 shows the distribution of informative indicators on the plane. The data are presented for four wheat seed samples with different yields. Seeds with a yield of 39 centner/ha belong to the variety "Thassos", with a yield of 22 centner/ha - "Grunny". It is important to note that wheat seeds with yields of 16 and 35 are of the same "Altai-75" variety. Seeds were grown in different fields.

The text of the paper should be formatted as follows:
Figure 2. Distribution of informative indicators on the plane.

The visualization showed that the data is divided into four yield groups. There is a nonlinear dependence of informative features. There are emissions in the data. Therefore, pre-processing of data is required before forming a training sample.

Processing represents the removal of gaps and anomalies (so-called emissions) in the data.

3.1.1. Practice problem No. 1 - Classification by quality indicators (yield properties)

Informative features for solving the classification problem by yield properties were formulated. They represent the input parameters:

- The maximum value of seeds membrane potential from the fraction of 8 m/s.
- The maximum value of seeds membrane potential from the fraction of 9 m/s.
- The maximum value of seeds membrane potential from the fraction of 10 m/s.
- The maximum value of seeds membrane potential from the fraction of 11 m/s.
- The rise time of seeds membrane potential from the fraction of 8 m/s.
- The rise time of seeds membrane potential from the fraction of 9 m/s.
- The rise time of seeds membrane potential from the fraction of 10 m/s.

The output parameter of the model is the yield level (16, 22, 35, or 39).

3.1.2. Practice problem No. 2 - Classification by fractions to assess the quality of the cleaning line.

The number of input nodes is reduced for solving the problem of classification by fractions:

- The maximum value of seeds membrane potential.
- The rise time of seeds membrane potential.
- Wheat seed variety.

The separation speed values were selected as the output parameters - 8, 9, 10, 11 m/s.
Table 1. Fragment of the training set with original data for classification by quality indicators (yield properties).

| MaxMP (8 m/s), mV | MaxMP (9 m/s), mV | MaxMP (10 m/s), mV | MaxMP (11 m/s), mV | Time rise (8 m/s), N | Time rise (9 m/s), N | Time rise (10 m/s), N | Time rise (11 m/s), N | Seed Yield |
|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|-----------------|-----------------|------------|
| 254.00            | 215.00            | 181.00            | 133.00            | 25              | 91              | 23              | 19              | 22         |
| 258.00            | 245.00            | 175.00            | 131.00            | 37              | 92              | 24              | 50              | 22         |
| 194.00            | 143.00            | 202.00            | 167.00            | 74              | 155             | 53              | 39              | 39         |
| 195.00            | 157.00            | 203.00            | 171.00            | 66              | 159             | 70              | 15              | 39         |
| 191.00            | 150.00            | 216.00            | 60.00             | 68              | 136             | 19              | 13              | 35         |
| 193.00            | 155.00            | 224.00            | 80.00             | 48              | 122             | 16              | 18              | 35         |
| 263.00            | 167.00            | 226.00            | 213.00            | 3               | 50              | 9               | 12              | 16         |
| 238.00            | 168.00            | 243.00            | 199.00            | 5               | 52              | 12              | 6               | 16         |

Table 2. Fragment of the training set with original data for classification by fractions.

| MaxMP, mV | Time rise, N | Wheat seed variety | Seed Separation into fractions |
|-----------|--------------|--------------------|-------------------------------|
| 235.00    | 88           | Grunny             | 8                             |
| 206.00    | 57           | Thassos            | 8                             |
| 158.00    | 149          | Thassos            | 9                             |
| 162.00    | 133          | Altai-75           | 9                             |
| 234.00    | 6            | Altai-75           | 10                            |
| 133.00    | 19           | Grunny             | 10                            |

Neural networks can only work with numbers. All quality parameters had to be pre-encoded. In our case, the sample set contains one quality parameter; this is the wheat seed variety.

To avoid false ordering of the characteristic values, the 1-of-N method was applied. Each of the N values of the quality parameter in the sample is replaced by a tuple of length N, in which all positions except one have zeros and one position is the unit, the location of the unit in the tuple is unique for each value of the quality parameter. After coding, instead of one column in the sample corresponding to a quality parameter having N possible values, N columns appear.

The coding of wheat seed varieties according to the 1-of-N method is presented in table 3.

Table 3. Coding of sample quality indicators.

| Wheat seed variety | Variety1 | Variety2 | Variety3 |
|--------------------|----------|----------|----------|
| Grunny             | 1        | 0        | 0        |
| Thassos            | 0        | 1        | 0        |
| Altai-75           | 0        | 0        | 1        |

3.2. Neural network formation

The formation of a neural network involves determining the type of network, the number of hidden layers, the number of neurons on hidden layers and the choice of algorithm. The number of neurons in the input and output layer is determined by the conditions of the problem as the number of input and output parameters in the model. In our case, the number of neurons in the output layer was always equal to 1.
The type of neural network is a multilayer perception. Networks were trained through the back propagation algorithm.

3.2.1. Determining the number of layers and the number of neurons
The number of hidden layers and the number of neurons were selected experimentally. The results of computational experiments to determine the best parameters of the structure of a neural network for the practical task of classification by fractions to assess the quality of the cleaning line are presented in table 4.

Table 4. The results of computational experiments for the problem of classification by fractions.

| Network structure | Training Error, % | Generalization Error, % |
|-------------------|-------------------|-------------------------|
| (2)               | 18,89             | 24,06                   |
| (2,2)             | 25,15             | 26,56                   |
| (5)               | 13,51             | 17,19                   |
| (5,5)             | 12,42             | 10,00                   |
| (10)              | 11,64             | 10,94                   |
| (10,10)           | 10,47             | 10,62                   |
| (20)              | 11,33             | 14,06                   |
| (20,20)           | 9,45              | 11,88                   |
| (30)              | 11,25             | 12,81                   |
| (30,30)           | 10,63             | 12,50                   |

The minimum generalization errors (on test data) are achieved with two hidden layers of 5 and 10 neurons, as well as with one hidden layer with 10 neurons. A further increase in the structure of the neural network entails a decrease in training errors, but at the same time, there is an increase in generalization errors, which indicates a retraining of the network. Neural networks with a small structure are not able to solve successfully the problem, so the optimal value has been experimentally established.
A similar study was conducted to solve the problem of classification by quality indicators (yield properties). It has been established that for adequate results, a neural network with one hidden layer of five neurons is optimal.

3.2.2. Setting neural network parameters
The network learning process begins with the initialization, randomly generating the initial matrix of neuron coupling coefficients within the network, the so-called synaptic weights, and consists in sequentially changing these weights so that the network error decreases. When the entire supply of training examples has been exhausted, it is believed that one epoch has ended. The average error per epoch is calculated and the examples from the training set are again used to train the same slightly trained network - this is the second epoch. The results of computational experiments with a different number of epochs are presented for solving the problem of classification by fractions to assess the quality of the cleaning line (figure 4).

![Training Error](image)

**Figure 4.** Changing network training errors depending on setting the maximum number of epochs for problem of classifications by fractions.

The results are recorded after each k-epoch of training, and instead of the training sample, examples from the testing sample are fed to the network input. They determine the network error, but they do not use for training the network. When the examples from the test set are exhausted, the average test error is determined.

In this paper, for the problem of classification by fractions, the criteria for stopping neural network training were the error level of 10% or less and the maximum number of epochs equal to 1000. To solve the problem of classification by quality indicators (yield properties), it was found that 1000 epochs allows the network to recognize all examples of the training data set with 96% accuracy.

3.3. Pre-training of the network and analysis of training results
The use of a neural network has reduced the number of training examples to 30. This means that for reliable classification it is necessary to study 30 wheat seeds. The results are shown in figure 5.
It is recommended to use all 400 examples for solving the problem of classification by fractions. Assessing the significance of the input parameters is one of the important points for increasing the reliability of the network. Therefore, to solve the problem of classifying seeds by fractions, it was established that the wheat variety is an important input parameter (figure 6).

Before starting the experiments, we found correlation coefficients that reflect the dependence of some parameters on others. Correlation coefficients can take values from -1 to +1, positive values indicate the presence of direct dependencies, negative values indicate the presence of inverse dependencies.
Table 5. Correlation coefficients between model parameters for solving the problem of classifying seeds by fractions (problem No. 2).

| Model Parameters | MaxMP | Rise Time | Variety1 | Variety2 | Variety3 | Seed Separation |
|------------------|-------|-----------|----------|----------|----------|-----------------|
| MaxMP             | 1.00000 | -0.16406 | -0.00024 | 0.14921  | -0.14942 | -0.53035        |
| Rise Time         | -0.16406 | 1.00000  | -0.29726 | -0.04886 | 0.39226  | -0.35250        |
| Variety1          | -0.00024 | -0.29726 | 1.00000  | -0.57830 | -0.57447 | -0.00000        |
| Variety2          | 0.14921  | -0.04886 | -0.57830 | 1.00000  | -0.33556 | -0.00000        |
| Variety3          | -0.14942 | 0.39226  | -0.57447 | -0.33556 | 1.00000  | 0.00000         |
| Seed Separation   | -0.53035 | -0.35250 | -0.00000 | 0.00000  | 0.00000  | 1.00000         |

The correlation coefficients between the parameters responsible for the wheat seed variety and the output parameter are almost equal to 0, which indicates the absence of a relationship between them (Table 5). However, the performed computational experiments, the results of which are shown in Figure 6, indicate the opposite. The proportion of correctly recognized examples in the test sample is significantly higher when the model is responsible for the parameters responsible for the variety. There can be only one reason for this. There is no linear dependence between the parameters (correlation coefficients determine either the presence or absence of only linear dependencies), there is a nonlinear dependence that was found and proved through the use of neural networks.

4. Conclusions

This paper presents the results of analysis of seed quality indicators based on neural network. It is important to use another informative feature - the variety of wheat seeds for the problem of classifying seeds by fractions to assess the quality of the cleaning line. That allows increasing the accuracy of the method up to 90%.

To solve the classification problem by yield properties, an artificial neural network can reduce the number of training examples to 30, while the accuracy remains 96%.

Using a neural network, a nonlinear dependence of such parameters as the maximum value of the membrane potential and the signal rise time from the yield properties of seeds was established. The model allows predicting the yield as a seed quality.

A nonlinear dependence of the maximum value of the membrane potential, the rise time of the wheat seed signal and seeds variety from the belonging of seeds to a particular separation fraction (the speed separation into fractions in this example) was established. Studies have shown that the variety is an important informative feature. Therefore, to apply this method in practice, it is important to replenish the knowledge base and conduct additional research with other varieties of wheat seeds.

The model adequately showed itself when using seeds of one variety of "Altai-75", but different yield properties and different quality for solving problems of classification by yield properties and classification by fractions.

Thus, the results of this study provide useful information in practice for solving the problem of diagnosing the quality of seed material and for assessing the quality of the cleaning line (separation into fractions), subject to additional studies with other varieties of wheat seeds.

References

[1] Hampton J G and TeKrony D M 1995 Handbook of vigour test methods (Zurich: ISTA) p 117
[2] Bewley J D and Black M 1982 Physiology and biochemistry of seeds in relation to germination (Berlin: Springer) vol 2 pp 3–9
[3] Zhang T 2018 Sensors 18(3) 813
[4] Yang L and Wen B 2017 Encyclopedia of Applied Plant Sciences 1 553–563.
[5] Colville L 2017 Encyclopedia of Applied Plant Sciences 1 335–339
[6] Anisur R and Byoung-Kwan Ch 2016 Seed Science Research 26(4) 285–305
[7] Barysheva N N and Pronin S P 2019 Engineering Technologies and Systems 29(3) 443–455
[8] Pasynkova E N, Zavalin A A and Pasynkov A V 2018 Russ. Agricult. Sci. 8 409 - 413
[9] Skvortsov E A, Nabokov V I, Nekrasov K V, Skvortsova E G and Krotov M I 2019 Agrarian Bulletin of the Urals 8(187)
[10] Caley P and Kuhnert P M 2006 Austral Ecology 31 647–655
[11] Wu W, Li Ai-Di, He Xin-Hua, Ma R, Liu Hong-Bin and Lv Jia-Ke 2018 Computers and Electronics in Agriculture (Elsevier) 144 86-93
[12] Guo P, Wu W and Sheng Q 2013 Nutr. Cycl. Agroecosyst. 95 333-344
[13] Kamruzzaman M, ElMasry G, Sun D-W and Allen P 2013 Food Chemistry 141(1) 389-396
[14] Liu D, Sun D-W and Zeng X A 2014 Food and Bioprocess Technology 7(2) 307-323
[15] Chen P and Sun Z 1991 Journal of Agricultural Engineering Research 49 85-98