Technical Sciences, 2021, 24, 135–155

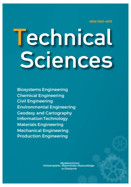
UNIWERSYTET
WARMIŃSKO-MAZURSKI
W OLSZTYNIE

Yearbook peer-reviewed scientific journal

ISSN 1505-4675
e-ISSN 2083-4527

TECHNICAL SCIENCES

Homepage: https://czasopisma.uwm.edu.pl/index.php/ts/



ISSN 1505-4675

Technical Sciences

Biosystems Engineering
Chemical Engineering
Civil Engineering
Environmental Engineering
Geodesy and Cartography
Information Technology
Materials Engineering
Mechanical Engineering
Production Engineering

Wydawnictwo
Uniwersytetu Warmińsko-Mazurskiego
w Olsztynie

DOI: https://doi.org/10.31648/ts.7062

BOOTSTRAP AGGREGATION TECHNIQUE FOR EVALUATING THE SIGNIFICANCE OF MANUFACTURING PROCESS PARAMETERS IN THE GLASS INDUSTRY

Łukasz Paśko[1], *Aneta Kuś*[2]

[1] ORCID: 0000-0001-8175-2295
[2] ORCID: 0000-0002-5028-0261
Department of Computer Science
Faculty of Mechanical Engineering and Aeronautics
Rzeszów University of Technology

Received 27 August 2021, accepted 14 September 2021, available online 22 September 2021.

Keywords: manufacturing process, glassworks, neural networks, bagging, manufacturing parameters.

Abstract

The article presents the application of the bootstrap aggregation technique to create a set of artificial neural networks (multilayer perceptron). The task of the set of neural networks is to predict the number of defective products on the basis of values of manufacturing process parameters, and to determine how the manufacturing process parameters affect the prediction result. For this purpose, four methods of determining the significance of the manufacturing process parameters have been proposed. These methods are based on the analysis of connection weights between neurons and the examination of prediction error generated by neural networks. The proposed methods take into account the fact that not a single neural network is used, but the set of networks. The article presents the research methodology as well as the results obtained for real data that come from a glassworks company and concern a production process of glass packaging. As a result of the research, it was found that it is justified to use a set of neural networks to predict the number of defective products in the glass industry, and besides, the significance of the manufacturing process parameters in the glassworks company was established using the developed set of neural networks.

Correspondence: Łukasz Paśko, Zakład Informatyki, Wydział Budowy Maszyn i Lotnictwa, Politechnika Rzeszowska im. Ignacego Łukasiewicza, al. Powstańców Warszawy 12, 35-959 Rzeszów, e-mail: lpasko@prz.edu.pl

Introduction

Nowadays, it is difficult to find an industry that does not use artificial neural networks (ANNs). ANNs are even used in agriculture, to assess the degree of maturity of apples, as described by GÓRSKI et al. (2008), to assess the wheat damage, as presented by GOLKA et al. (2020) or to determine the hardness of wheat kernels, as discussed by HEBDA and FRANCIK (2006). Neural modelling in the agriculture field was also used by NIEDBAŁA et al. (2015) to predict starch content in potatoes. The methods of using ANN also vary. Staying on the topic of the non-obvious application of ANN in agriculture, but also in mechanical engineering, FRANCIK (2006) describes the use of the method of forecasting time series with the use of ANN to characterize agricultural machines. The use of ANNs in mechanical engineering has been discussed for many years; already LEFIK (2005) described the applications of ANNs in mechanics and engineering, but it can certainly be said that this application is still showing an upward trend and we are witnessing incredible developments in the field of ANNs.

The use of ANNs is also an indispensable element of the Big Data phenomenon. Collecting an enormous amount of data, including those that may seem to be useless, in combination with machine learning allows us to detect dependencies that may affect an entire process under study, and which were previously not taken into account. This combination is particularly useful in production processes, e.g. for the assessment of production parameters and detection of the causes of undesirable phenomena. As assessed by TADEUSIEWICZ and HADUCH (2015), ANNs are used, for example, for the examination of motor gasoline and constitute an excellent prognostic tool supporting the management of the production process. ANNs are also used in production quality control processes, as described by ROJEK (2015).

There are many types of ANNs and their applications, but one of the most popular types is the Multilayer Perceptron (MLP) network. This is confirmed by the scale of application of MLP-type neural networks in classification and regression, as discussed by KURT et al. (2014). MLP networks are used where there is a large number of input and output data, it is necessary to define the relationship between them and when the problem is very complex. The MLP network consists of layers: input, hidden and output neurons and is a type of unidirectional network, i.e. a signal that flows from the input neurons through the hidden layer to the output neurons no longer returns to “earlier” neurons. Various learning algorithms are used to train MLP neural networks. One of the most popular is the error backpropagation algorithm.

One of the concepts that has emerged in machine learning is ensemble machine learning. This concept implies the use of more than one model (e.g. ANN) to derive a classification or regression result. An example of ensemble machine learning is bootstrap aggregation–otherwise also known as bagging.

This is an aggregation method that uses the same training algorithm on different subsets of data created using cases from the original training dataset. These subsets of data are called "bags of data". According to the described method, from a training dataset of size n, m subsets of a dataset D ($D1$, $D2$, ..., Dm) of size n', are created at random, using cases from the original dataset. Cases for subsets are selected randomly on the basis of replacement, which means that cases may repeat in one subset and this is not undesirable. The bagging method uses the $n' < n$ rule. It means that each of the subsets of D contains fewer cases than the original dataset. Then each of the subsets is used as a training dataset, in effect creating m models of the same type – each of them trained on a different set of data. The results of each model are then averaged, resulting in the final classification or regression result.

The study described in this article is not the first use of bagging with ANNs. OPITZ and MACLIN (1997) assessed the possibility of using bagging in this field, ZHIBIN et al. (2018) described possibility of using bagging method in ANN for prediction of thermal comfort in buildings. Also, already in 1996, BREIMAN (1996) confirmed that the described method of data aggregation solves the problem of the lack of stability of classifiers. However, above-mentioned studies are related to the classifiers topic while in this research the bagging method was used in the regression problem. Another novelty introduced by this article is the use of the described method on a real-world data set obtained from a glassworks production process while OPITZ and MACLIN (1997) conduct research on a publicly available data set, very often used in this type of research. In addition, mentioned authors used cross validation method, while in the studies described in this article, the dataset is sufficient to divide it into three subsets: train, validation and test, without using cross validation method.

Having access to an enormous number of parameters that can be input variables using ANNs, it is extremely important and difficult to determine the significance of each of them. The methods of carrying out significance analysis are described by JASIŃSKI and BOCHENEK (2016), where the authors use ANN to forecast real estate prices. Examples of methods for determining the significance of parameters are: searching for the optimal set of parameters by building many models with a different set of explanatory variables or testing the sensitivity of the output variable to the input data. The result of carrying out such analyses is to remove those input parameters that have the least impact on the output variable. The results of tests carried out by the RODZIEWICZ and PERZYK (2016) who analysed the significance of the parameters of a continuous steel casting process, confirm the expediency of determining the significance of the input variables for the detection of defects in manufactured products.

This article presents the use of the previously described bagging (bootstrap aggregation) method and MLP-type ANNs to determine the significance of the manufacturing process parameters in a glassworks, which can be used to determine the causes of defects in the glass products. The article is a continuation of previous research (PAŚKO 2020), which considers the possibility of using ANNs to determine the influence of production parameters on the number of defective products for a single ANN.

Description of the Case Study

The aim of the research was to predict the number of defective products based on the values of selected manufacturing process parameters. Additionally, an attempt was made to determine the impact of the analyzed manufacturing process parameters on the number of defective products. Thanks to the conducted experiment, it is also possible to assess the usefulness of the bootstrap aggregation technique combined with ANNs.

The problem of product defectiveness discussed in the article is one of the most important problems for manufacturing companies. Product quality assurance is a key element of the strategy of manufacturing companies, because these companies are forced to constantly meet the needs of customers by offering them the highest quality products. For this reason, research on the quality of products and the search for the causes of product defects are still up-to-date and needed.

The research was based on a case study of a glassworks that produces glass packaging. In the glassworks, several dozen types of defects in products have been defined, but only one type of defect was considered in this study. The analyzed defect consists in the presence of small size air bubbles located inside the glass, which are visible in the finished products.

The more complicated the manufacturing process, the more various types of product defects may appear. It is particularly visible on the example of the manufacturing process in a glassworks, which consists of the following stages: preparation of an appropriate mixture of raw materials, melting of the prepared mixture, molding of liquid glass, hot refining of products, annealing of products to reduce internal stresses, and cold refining to increase the aesthetic value of the product. The occurrence of abnormalities in any of these stages can deteriorate the quality of the finished products.

The manufacturing process in a glassworks requires the maintenance of appropriate parameters, such as:

– temperature – determines the effectiveness of melting raw materials and the correctness of the annealing process of semi-finished products;

– atmospheric conditions – maintaining appropriate conditions affects the cleanliness of the glass;

– gas pressure – is responsible for the homogeneity of the glass mass.

It is necessary to establish appropriate values for the above parameters in order to avoid the production of defective products.

The experiment discussed in this paper concerns the problem of regression. In this type of problem, the value of the explained variable is predicted based on the values of the explanatory variables (predictors). The research was carried out on the basis of a dataset from the glassworks. The dataset consists of 70 explanatory variables and one explained variable. The explanatory variables concerned selected parameters of the manufacturing process. Their values were recorded every second, but for this study the dataset was transformed with an interval of ten minutes. The parameters of the manufacturing process that were taken into account in the research concern:

– the temperature of the glass measured in several places of the glass furnace (each measurement location was associated with a separate explanatory variable);

– the temperature of the glass measured in several places of the forehearth, which distributes the glass to the production line (each measurement place was associated with a separate explanatory variable);

– selected operating parameters of the glass furnace and the forehearth (e.g. the level of glass in the furnace, the speed of loading raw materials into the furnace, combustion air pressure in several dozen locations of the forehearth, and each operating parameter was associated with a separate explanatory variable);

– selected cooling parameters of glass molds.

The explained variable is the number of defective products in which the analyzed defect was identified. Data on the number of defective products was recorded every ten minutes. This means that each data point contains information on how many defective products have been identified at the quality control station in the last ten minutes. The dataset included the values of the above-mentioned manufacturing process parameters recorded during 27 days of continuous production.

Table 1 presents the basic statistics on selected predictors and the dependent variable. The table reveals the clear differences between the variables. These differences are also confirmed by the graphs shown in Figures 1-2. Figure 1 presents time series plots of three selected explanatory variables and the dependent variable. These types of graphs are used to assess the basic characteristics of variables, such as long-term trends, regular periodic fluctuations, the level of values, and the homogeneity of variance. For many predictors, it has been noted that their values fluctuate around more than one constant level, and the transition between levels occurs quickly. When observing the variance of the variables (differences in values around the average level of the variable), it is noticed that the variance is not homogeneous. For some variables, their variance increases over time. Focusing on observing periodic fluctuations, it can be stated that in some cases the periodic fluctuations are clearly visible. The best example is temperature variable, which contains data about the air temperature measured in the production hall. Its periodicity is related to the daily fluctuations in air temperature.

Table 1

Basic statistics of selected variables

Variable	Min.	Mean	Median	Max.	Standard deviation
J2FFONTTPTCPV	1,271.3	1,300.2	1,299.9	1,321.2	8.9
J21BLCBPRPV	9.8	10.3	10.4	11.2	0.3
Pressure	987.9	1,015.7	1,016.7	1,028.1	7.7
No. of defects	1	36.7	30	255	28.4

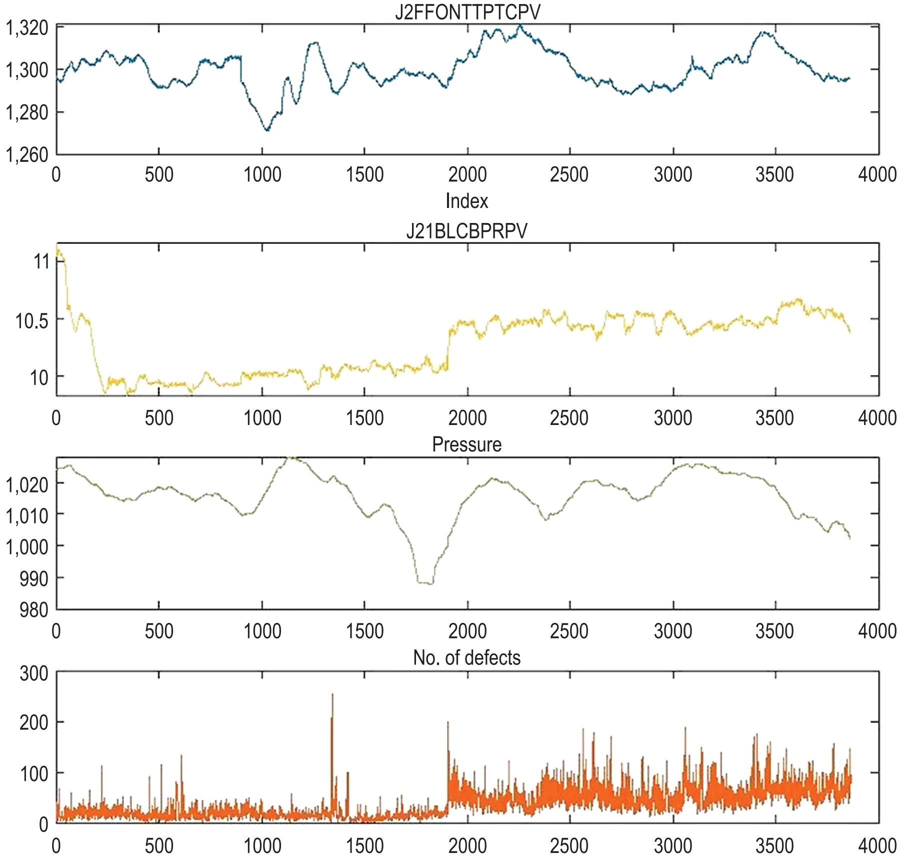


Fig. 1. Time series plots for selected variables

In turn, Figure 2 contains histograms of selected variables. The histograms show that distributions of the variables are diversified. In the case of many variables, bimodal distributions can be noticed, in which two histogram bars are clearly higher than the others. This arrangement of the histogram bars shows that most of the observations are concentrated around two levels of values.

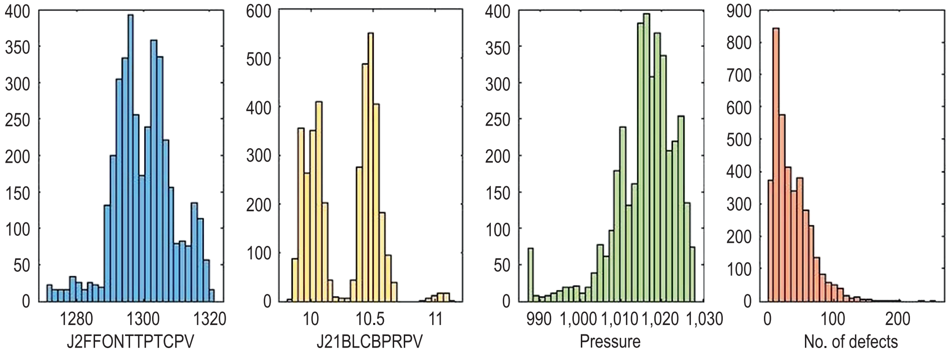


Fig. 2. Histograms for selected variables

Among the analyzed variables, there are also those that their histogram is characterized by a clear left-skewed (negative skewness), as in the case of the Pressure variable. In turn, a small part of the variables has a right-skewed distribution (positive asymmetry). Such a distribution applies to the dependent variable (no. of defects). The conducted exploratory data analysis showed that the methods of data analysis used during the research should be insensitive to heterogeneity of variance and to the distributions of the variables, as well as should be able to map the levels of the variables and their seasonal fluctuations.

Research Methodology

The research was carried out according to the framework shown in Figure 3. MATLAB software was used to carry out all the activities in this experiment. The first stage was the exploratory data analysis. One of the explanatory variables was rejected at this stage because its values did not change over the period considered, so this variable could not affect the number of defective products in the analyzed period. Therefore, in the next stages, 69 explanatory variables were taken into account. Apart from the aforementioned rejection of one variable based on the time series graph, no other feature extraction techniques were used. This was because feature extraction techniques such as principal component analysis or independent component analysis transform the original variables into a set of new variables called components. A characteristic feature of the components is that they are artificial variables, the value of which is difficult to interpret in relation to the manufacturing process parameters. One of the goals of the conducted research was to establish the significance of the manufacturing process parameters, which would be difficult after the use of techniques that transform the original parameters.

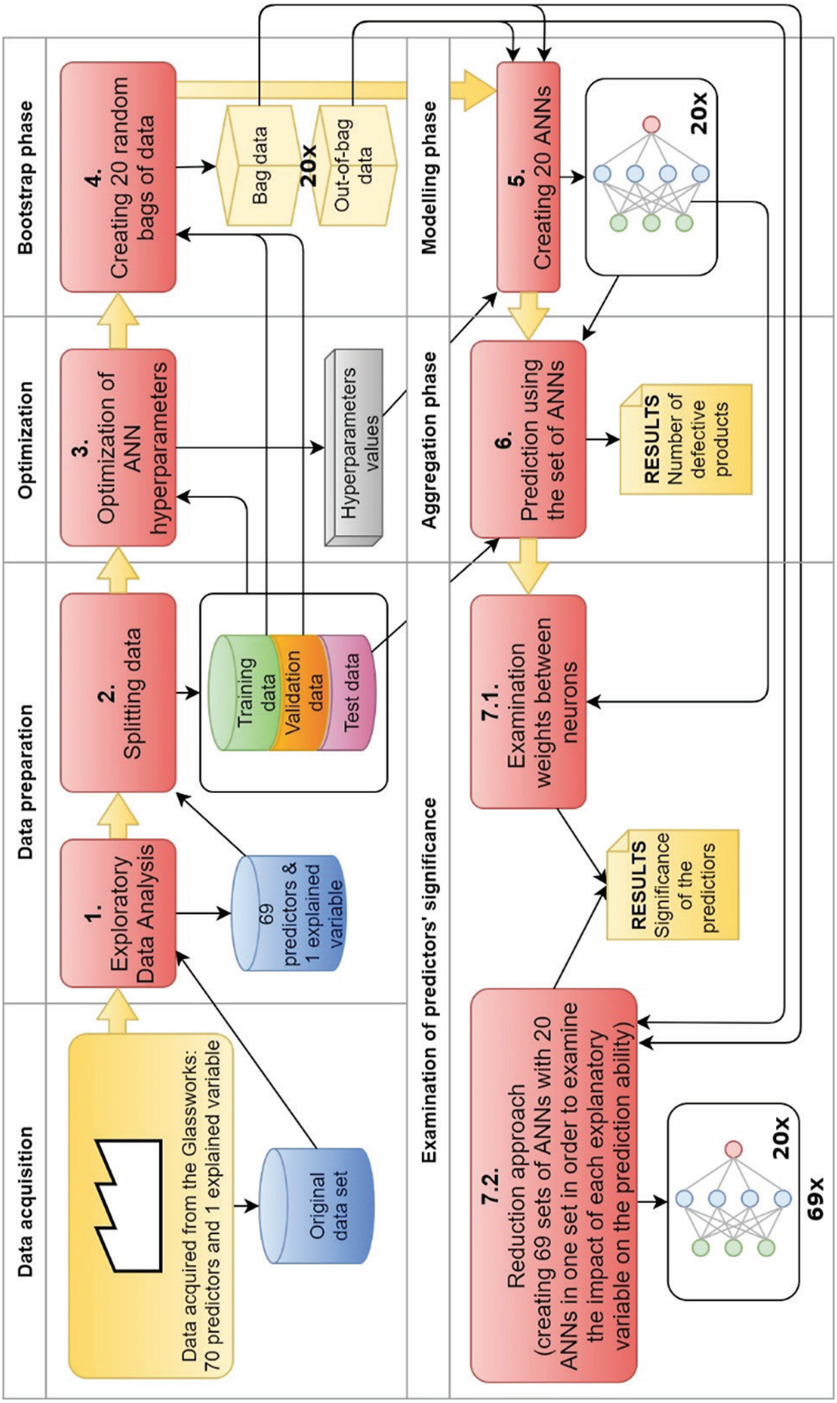

Data acquisition
Data preparation
Optimization
Bootstrap phase
Data acquired from the Glassworks: 70 predictors and 1 explained variable
Original data set
1. Exploratory Data Analysis
69 predictors & 1 explained variable
2. Splitting data
Training data
Validation data
Test data
3. Optimization of ANN hyperparameters
Hyperparameters values
4. Creating 20 random bags of data
Bag data
20x
Out-of-bag data
Examination of predictors' significance
Aggregation phase
Modelling phase
5. Creating 20 ANNs
20x
6. Prediction using the set of ANNs
RESULTS Number of defective products
7.1. Examination weights between neurons
RESULTS Significance of the predictiors
7.2. Reduction approach (creating 69 sets of ANNs with 20 ANNs in one set in order to examine the impact of each explanatory variable on the prediction ability)
20x
69x

Fig. 3. Research framework

The second stage was to divide the dataset into training, validation and test subsets. The number of observations in each variable was 3,866. 3,094 cases were randomly selected for the training subset, which constitutes about 80% of all cases from the dataset (a "case" is a vector composed of the observations of all explanatory variables and the explained variable registered at a given time *t*). The validation subset consisted of 386 randomly selected cases (approx. 10%). The remaining cases (386) were included in the test subset.

The third stage was to optimize the ANN hyperparameters. The aim of this stage was to determine the optimal values of the three MLP neural network hyperparameters: the number of neurons in the hidden layer, the activation function of hidden neurons, and regularization term strength (denoted as lambda). Table 2 shows the values of the hyperparameters that were taken into account during the optimization. 50 network sizes (the number of hidden neurons), 4 activation functions, and 51 lambda values were analyzed. Consequently, 10,200 ANNs were created. Such approach that involves generating all possible combinations of the analyzed hyperparameters would be computationally inefficient in the case of Big Data. However, the analyzed data does not have three basic features ascribed to Big Data, called 3V:

– Volume — a huge amount of data;

– Variety — various data formats and different degrees of data structuring;

– Velocity — high speed of generating new observations that need to be take into account during the analysis and the need to generate analysis results as soon as possible.

Table 2

Hyperparameter values taken into account during optimization

Hyperparameter	Considered values
Number of hidden neurons	2, 4, 6, …, 100
Activation function of hidden neurons, f	Rectified linear unit function (ReLU), which performs the following operation on each input element x: $f(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$ Hyperbolic tangent function (tanh): $f(x) = \tanh(x)$ Sigmoid function: $f(x) = \frac{1}{1 + e^{-x}}$ Identity function: $f(x) = x$
Regularization term strength (lambda)	$0.05 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, …, $2.5 \cdot 10^{-3}$

Due to the fact that the analyzed data does not have the features of Big Data, it was decided to use the approach described below to optimize the hyperparameters.

Each of the created ANNs was trained with a different combination of hyperparameters. The learning algorithm called Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) was used to train the ANNs. The training subset was used to train the ANNs. The validation subset was used to control the end of ANN training. The ANN training process ended when the mean square error (MSE) calculated on the validation subset increased in ten consecutive iterations of the training algorithm. The MSE value is calculated according to the following formula:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

where:

y_i – the observed value of the explained variable,

$\hat{y}_i$ – the predicted value of the explained variable,

n – the number of cases in the dataset.

The role of the objective function was played by the root mean square error $(\mathrm{RMSE} = \sqrt{\mathrm{MSE}})$ calculated on the test subset. Thanks to the use of RMSE, the error the ANNs is expressed in the same units as the original explained variable. The learning process of all ANNs was repeated ten times, each time assuming different initial values of the weights of connections between neurons. After ten iterations, the mean RMSE was calculated for each combination of the hyperparameters considered. Figure 4 shows three plots of averaged RMSE measure. The plots show that the RMSE value is most influenced by the number of hidden neurons – the more neurons, the smaller the RMSE (although over 50 neurons, the RMSE does not decrease significantly). There are also clear differences in the activation function. Relatively low RMSE values occur for the ReLU and sigmoid functions, while for the hyperbolic tangent the RMSE values are the highest. The lambda parameter has the least influence on the RMSE. For all tested values of the lambda parameter, the RMSE values calculated on the test subset do not differ significantly from each other.

Finally, the set of hyperparameters for which the average RMSE calculated on the test subset had the smallest value was selected. As a result of optimization, the following set of hyperparameters was established: 80 hidden neurons, ReLU activation function, lambda value of $1.1 \cdot 10^{-3}$. Such values of hyperparameters were adopted in the subsequent stages of the experiment.

The fourth stage of the research is the preparation of bags that are subsets of the original dataset. It was assumed that the number of bags $m = 20$, and the number of cases in each bag $n' = 3400$. Cases for each bag were drawn from among the cases included in the training and validation subsets. Those cases

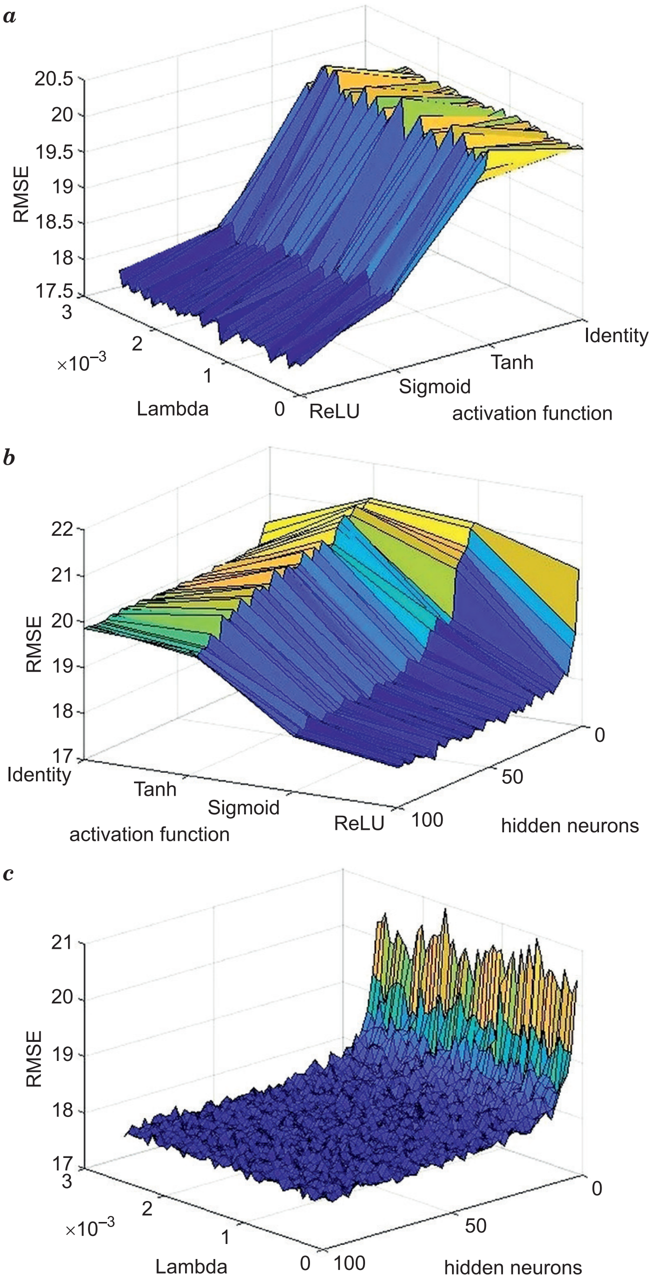


Fig. 4. Plots showing the influence of activation functions, number of hidden neurons, and lambda on RMSE obtained during optimization of hyperparameters: *a* – influence of lambda and activation function on RMSE (ANNs with 80 hidden neurons), *b* – influence of activation function and number of hidden neurons on RMSE (ANNs with lambda = 0.0011), *c* – influence of lambda and number of hidden neurons on RMSE (ANNs with ReLU activation function)

that were not selected for a given bag were placed in a set called out-of-bag. The cases included in the test subset were not used when creating bags.

In the fifth stage, a set of ANNs was created. The set consisted of $m = 20$ ANNs. All ANNs were developed using the hyperparameters established in the third stage. Each ANN was trained on one of the bags created. The out-of-bag set, which was created from cases not included in the given bag, was used for validation. When the MSE value calculated for cases from the out-of-bag set increased in ten learning iterations, the ANN training process was completed. This situation can be seen in Figure 5. The graph shows the MSE values determined during the training of one of twenty ANNs. The lowest MSE value for the validation subset was achieved in iteration no. 109. From iteration no. 110, the MSE value was slightly higher than the lowest value. This caused the learning process to end in iteration 119.

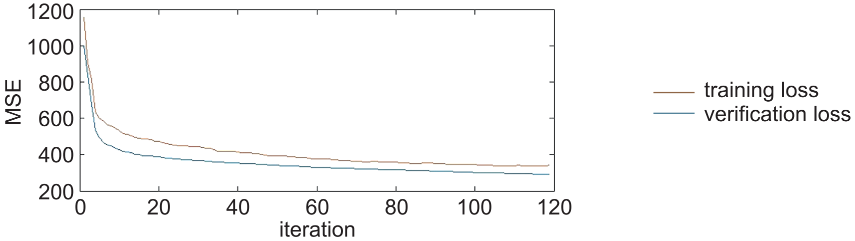


Fig. 5. MSE values calculated during training of one of the ANNs

The sixth stage is the prediction of the number of defective products using the twenty ANNs created. This stage can be called the aggregation phase in which the prediction results of each single ANN are aggregated by computing the arithmetic mean. The determined average number of defective products based on the results of twenty ANNs is treated as the final result of prediction by the set of ANNs. Prediction of the number of defective products was performed mainly for cases from the test subset. Based on the prediction results, the RMSE was calculated for each ANN separately and for the set of ANNs. For comparison, the RMSE on the training and validation subsets was also calculated.

In the last, seventh stage of the research, the significance of ANNs input variables, which are manufacturing process parameters, was examined. Two approaches were used to determine the significance of the input variables:

– Stage 7.1: examination of connections weights between input neurons and hidden neurons,

– Stage 7.2: reduction approach.

In step 7.1 it was assumed that for a given neural network v the significance of its input variable X_i is equal to the sum of the absolute values of connections weights $w_{Ii\to Hj}$ from the input neuron I_i which represents the variable X_i to each of the K hidden neurons (designation H) in this ANN:

$$\mathrm{ANN}_v sig_{X_i} = \sum_{j=1}^{K} \left| w_{Ii\to Hj} \right| \tag{2}$$

Having determined the significance of all input variables for each of twenty ANNs in the above manner, then three methods of calculating the significance of input variables were used, taking into account the entire set of ANNs:

– Method I: Significance of the variable X_i expressed as the sum of the significance of this variable for each of the ANNs:

$$Sig_{X_i}^{\mathrm{I}} = \sum_{v=1}^{m} \mathrm{ANN}_v sig_{X_i} \tag{3}$$

– Method II: Significance of the variable X_i expressed as the arithmetic mean of the significance of this variable for each of the ANNs:

$$Sig_{X_i}^{\mathrm{II}} = \frac{1}{m} \sum_{v=1}^{m} \mathrm{ANN}_v sig_{X_i} \tag{4}$$

– Method III: Significance of the variable X_i expressed as the median significance of this variable for each of the ANNs:

$$Sig_{X_i}^{\mathrm{III}} = \underset{v\in\{1,\ldots,m\}}{\mathrm{median}} \left(\mathrm{ANN}_v sig_{X_i} \right) \tag{5}$$

In turn, in step 7.2 a reduction approach was applied, thanks to which it was possible to use the fourth method of calculating the significance of input variables:

$$Sig_{X_i}^{\mathrm{IV}} = \frac{\mathrm{RMSE}_{X_i}}{\mathrm{RMSE}} \tag{6}$$

The given set of ANNs, which is used to determine the significance of the variable $Sig_{X_i}^{\mathrm{IV}}$ is trained on a subset of data that is devoid of the variable X_i. After generating the set of ANNs, the RMSE_{X_i} error for that set is determined. Finally, the significance of the variable X_i is treated as the quotient of the prediction error of the ANNs trained on the dataset devoid of the variable X_i and the prediction error of the ANNs trained on the dataset containing all input variables. The rejection of the input variable usually results in the deterioration of the predictive ability of the model (the RMSE error value increases). Therefore, it is assumed that the greater the RMSE_{X_i}, the more significant the variable X_i is. It is also possible that removing a given variable X_i from the training set will improve the predictive ability of the model. Then the measure $Sig_{X_i}^{\mathrm{IV}}$ takes values smaller than 1, and the variable X_i is treated as insignificant for the model. The reduction approach required the creation of N sets of ANNs, where N is the number of input variables. Each of the N sets was trained on a subset of data generated using the bagging technique.

Results and Discussion

The plot in Figure 6 shows the RMSE values taking into account the prediction results of each of the twenty ANNs generated in the fifth stage. The RMSE was computed for the training, validation, and test subsets. In the case of most ANNs, the following relationship can be seen: the RMSE for the training subset has the lowest value (ranging from approx. 16.6 to 20.4), while the RMSE for the validation subset has the highest value (from approx. 18.6 to 21.5). Between these values is the RMSE calculated on the test subset (from about 18.2 to 20.5). In turn, the RMSE value calculated from the prediction results of the entire set of ANNs is 18.31 for the test subset. This result can be compared with the naive method of predicting the number of defective products. The naive method may consist, first of all, in determining the arithmetic mean of the number of defective products, taking into account all cases included in the training subset (36.74). Next, the computed score can be taken as the prediction score for each case in the test set. Prediction using the naive method has the RMSE error of 27.84. This value is much higher than the RMSE determined for the set of ANNs, which proves that prediction using developed ANNs is more effective.

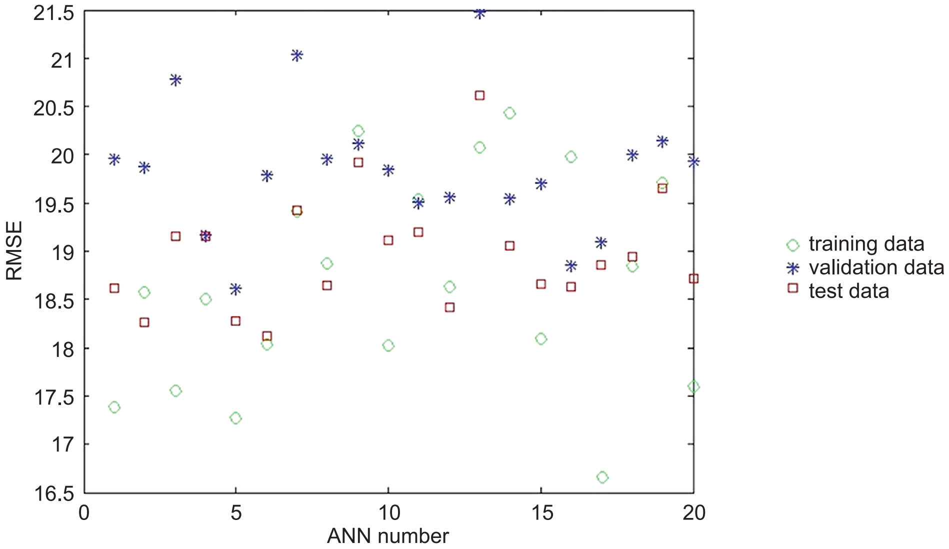


Fig. 6. RMSE values for each ANN belonging to the set of ANNs

Looking at the plot in Figure 7, it can be also concluded about the good quality of prediction using the set of ANNs. This plot shows the actual number of defective products for each case in the test subset along with the predicted number of defective products using the set of ANNs. The plot is in the form of a time series, but it should be noted that the time distances between adjacent cases may be different due to the fact that cases were assigned to the test subset randomly.

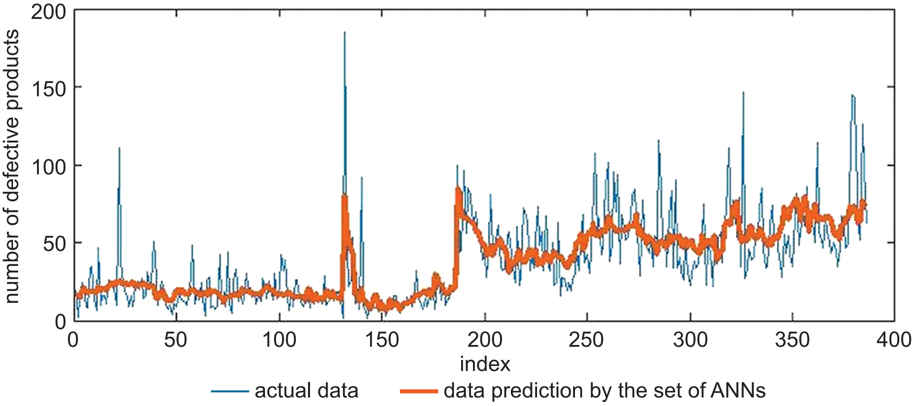


Fig. 7. Prediction results of the number of defective products compared to the actual number of defective products in the test subset

By comparing the actual number of defective products with the prediction results, it can be seen that for some cases there are large discrepancies (e.g. case no. 132: the actual value is 185 and the predicted value is 80.28). However, these single, large differences between the prediction result and the real value do not indicate a bad quality of prediction. The biggest differences are observed for the so-called outliers. The value of outliers is difficult to predict by any method. Moreover, predicting values for outliers is not the aim of this study. The purpose of the research was to create a model capable of capturing the most important properties of the explained variable for the analyzed period. The discussed plot shows that the set of ANNs is able to map the most important features of the explained variable in an appropriate way. In the left part of the plot (up to about 170 cases) it can be seen that the average number of defective products is less than the number of defects in the right part of the plot (above case no. 170). Similar dependencies can be observed for the predicted number of defective products. The set of ANNs is able to map not only the changing levels of the explained variable, but also correctly represent its variability understood as differences between the values in a given interval. For example, the variability in the number of defective products from 50 to 100 cases is less than the variability between 300 and 350 cases. This phenomenon can also be seen for prediction results generated by the set of ANNs.

Since it was found that the created set of ANNs can correctly map the most important phenomena occurring in the analyzed data, an attempt can be made to use the set of ANNs to determine the significance of the predictors. The significance of a given predictor is understood as the influence of this predictor on the prediction results, which is determined by the set of ANNs. The partial results of the predictor significance analysis are shown in Figure 8. This figure contains four column graphs corresponding to the four measures of significance described by Formulas 3, 4, 5, and 6. Each graph includes 10 most significant predictors of all 69 predictors analyzed.

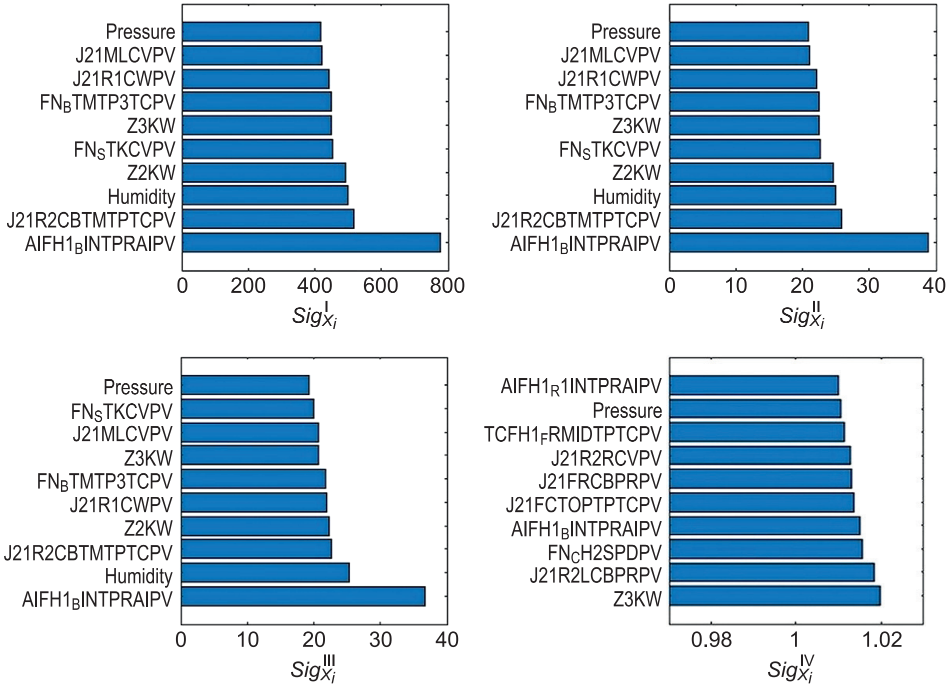


Fig. 8. Predictor significance rankings (included only 10 predictors with the highest value of a given measure of significance)

In the case of measures $Sig^{\mathrm{I}}_{X_i}$, $Sig^{\mathrm{II}}_{X_i}$ and $Sig^{\mathrm{III}}_{X_i}$ it can be seen a high similarity of the three significance rankings created. The top ten most important predictors are the same. Moreover, in the rankings created for the measures $Sig^{\mathrm{I}}_{X_i}$ and $Sig^{\mathrm{II}}_{X_i}$, the predictors with the highest significance occupy the same positions. In the case of the ranking for $Sig^{\mathrm{III}}_{X_i}$, the order of the variables is slightly different.

The first place in the three above-mentioned rankings is taken by the variable AIFH1$_{\mathrm{B}}$INTPRAIPV. The advantage of this variable over the other predictors is relatively large. This variable records the pressure values in one of zones of the forehearth that supplies liquid glass to the production line. The time course of the AIFH1$_B$INTPRAIPV variable over the entire period under study is shown in Figure 9. The course of this variable is compared with the time series presenting the number of defective products. Comparing the course of both time series, it can be seen that the variability of these time series is similar. In the periods when the value of AIFH1$_B$INTPRAIPV remains approximately constant, the explained variable also shows relatively little variability (an example may be the period between cases no. 1500 and 1900). On the other hand, in periods of high variability of AIFH1$_B$INTPRAIPV, the number of defective products also fluctuates relatively strongly, which can be seen, for example, in the period from case no. 2000 to the last case. This observation may explain why the AIFH1$_B$INTPRAIPV is the most significant variable for the set of ANNs.

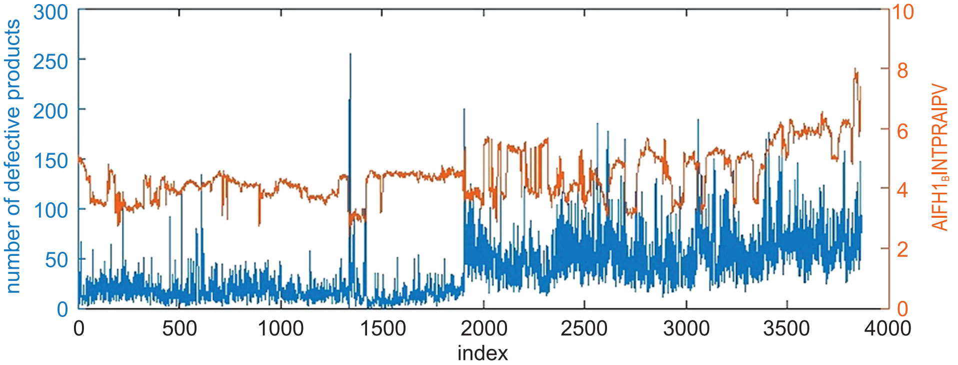


Fig. 9. Time series plots of the number of defective products and the predictor ranked highest in the three rankings

The differences in significance measures between the next variables in the rankings are much smaller. The following places are occupied by variables, which record: glass temperature in one of the forehearth zones, power of electric reheating, glass temperature in the lower part of the glass furnace, the degree of opening of valves that control cooling of some forehearth zones, humidity, and atmospheric pressure.

In the analysis of predictors significance in the stage 7.1, the values of bias of hidden neurons and weights of connections between hidden neurons and output neuron were also taken into account. However, the results obtained in that manner were similar to the results presented above.

The ranking for the measure $Sig_{X_i}^{\mathrm{IV}}$ looks slightly different. The differences between the obtained values are small and fall within the range from 0.98 to 1.02. Such small differences make it hard to draw unambiguous conclusions about the significance of the analyzed predictors. 13 out of 69 predictors have a value of $Sig_{X_i}^{\mathrm{IV}}$ less than 1. As mentioned in the previous section, this value suggests that removing a given predictor from the training subset could improve the quality of the prediction. However, no such analyzes were carried out in the conducted experiment.

Looking at the top ten of the ranking for the significance measure $Sig_{X_i}^{\mathrm{IV}}$, it can be seen that some variables also appear in the top ten of the remaining three rankings. Apart from these variables, the discussed ranking includes the following predictors: the degree of opening of valves that regulate the gas supply, the glass temperature in the upper part of one of the forehearth zones, and the speed of loading the raw materials into the furnace. The variable AIFH1$_{\mathrm{B}}$INTPRAIPV, which is the most significant in the other three rankings, occupies fourth place here.

It turns out that in the reduction approach the most important variable is Z3KW, which registers the power of electric reheating. This variable was also in the top positions in the three other rankings. Figure 10 shows the values of the variable Z3KW along with the number of defective products. Looking at the values of the Z3KW over time, one can notice clear, sudden changes in the levels of this variable. Comparing this with the time series of the explained variable, it can be seen that some changes in the levels of Z3KW are associated with relatively large increases or decreases in the number of defective products (for example, an increase in the value of Z3KW to the maximum level that begins in the case no. 1907 coexists with a sudden increase in the number of defects). Moreover, the minimum Z3KW value in the case no. 1339 coincides with the maximum number of defective products (255). These observations may explain why this predictor is important to the set of ANNs.

Of course, it is difficult to formulate an unequivocal conclusion that low or high values of a given predictor contribute to low or high values of the explained variable. The complexity of the glass packaging production process

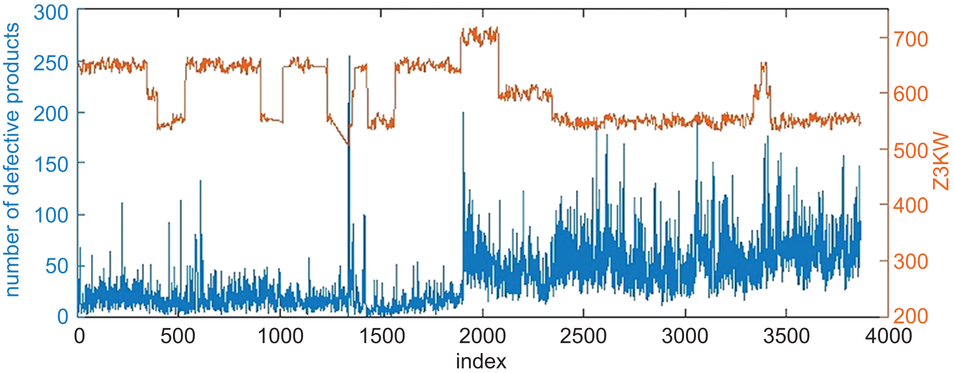


Fig. 10. Time series plots of the number of defective products and the predictor ranked highest in the fourth ranking

and the multitude of variables affecting the final quality of products make it difficult to formulate simple and unambiguous conclusions. However, the methods of predictor significance computation presented in this paper help to estimate which explanatory variables have the greatest impact on the results of prediction carried out using the set of ANNs and the bootstrap aggregation technique.

The presented research methodology has some limitations. The approach discussed in the paper takes into account only one type of products' defect. The applied techniques are aimed at predicting one type of defect, and the measures of significance of the predictors concern the influence of the predictors on the number of products with only a given type of defect. It should be borne in mind that glass products are characterized by the possibility of many types of defects, which are related to, among others, wrong shape of products, inadequate quality of a product's surface, or the occurrence of undesirable inclusions in the glass. In order to take into account more than one type of defect, the presented approach would have to be modified. There are also limitations to the applied method of ANNs hyperparameters optimization. For the considered dataset, the applied optimization method allowed for the selection of hyperparameters in a satisfactory time. However, in the case of a larger amount of data (in particular Big Data), the applied optimization method may be too time-consuming and computationally inefficient. In such a situation, other methods of selecting ANNs hyperparameters should be used, which were proposed by e.g. REN et al. (2021) and ELSKEN et al. (2019).

Conclusions

The research described in this article were aimed at determining the significance of the manufacturing process parameters of the glass products in relation to the quality of these products, as well as the prediction of the number of defective products based on the values of the manufacturing process parameters. The fulfilment of the assumed goals was achieved with the use of the MLP-type ANNs. Moreover, the bagging (bootstrap aggregation) technique was used to create the ANNs. The research framework included seven stages: data acquisition, data preparation, optimization, bootstrap phase, modelling phase, aggregation phase, examination of predictors' significance. This article describes a data analysis model capable of predicting the number of defective products for real data obtained from a glassworks, and presents four ways of determining the significance of predictors. MATLAB software was used to carry out the experiment.

Despite the very high complexity of the production process and the enormous amount of data, the assumed goals were fulfilled. The validity of the use of the bagging technique with ANNs in predicting the number of defective products and determining the significance of production parameters was confirmed. During the research, the development potential of the experiment was also identified. In subsequent iterations, it is planned to determine the optimal number of neural networks placed in the set of ANNs (20 ANNs were adopted in the described research) and to determine the optimal number of cases to be placed in one bag of data when using the bootstrap aggregation technique (a number of 3,400 cases was assumed) – the values were assumed intuitively; however, with appropriate optimization, it may be possible to obtain better prediction results. It is also planned to use other types of ANNs, e.g. networks with more than one hidden layer, which can model more complex relationships between predictors and the dependent variable. Continuing the research, other regression models can be applied, such as regression trees, random forests or support vector regression. Thanks to this, it will be possible to compare the predictive ability of these models in the task of predicting the number of defective products.

References

Breiman L. 1996. *Bias, variance and arcing classifiers. Technical Report TR 460. Dept. of Statistics*. University of California, Berkeley, CA, USA.

Elsken T., Metzen J.H., Hutter F. 2019. *Neural Architecture Search: A Survey*. Journal of Machine Learning Research, 20: 1-21.

Francik S. 2009. *Metoda prognozowania szeregów czasowych przy użyciu sztucznych sieci neuronowych*. Inżynieria Rolnicza, 13(6): 53-59.

Golka W., Arseniuk E., Golka A., Góral T. 2020. *Sztuczne sieci neuronowe i teledetekcja w ocenie porażenia pszenicy jarej fuzariozą kłosów*. Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin, 288: 67-75.

Górski M., Kaleta J., Langman J. 2008. *Zastosowanie sztucznych sieci neuronowych do oceny stopnia dojrzałości jabłek*. Inżynieria Rolnicza, 12(7): 53-56.

Hebda T., Francik S. 2006. *Model twardości ziarniaków pszenicy wykorzystujący sztuczne sieci neuronowe*. Inżynieria Rolnicza, 10(13): 139-146.

Jasiński T., Bochenek A. 2016. *Prognozowanie cen nieruchomości lokalowych za pomocą sztucznych sieci neuronowych*. Studia i Prace WNEIZ US, 45(1): 317-327.

Kurt I., Ture M., Unubol M., Katranci M., Guney E. 2014. *Comparing Performances of Logistic Regression, Classification & Regression Trees and Artificial Neural Networks for Predicting Albuminuria in Type 2 Diabetes Mellitus*. International Journal of Sciences: Basic and Applied Research (IJSBAR), 16(1): 173-187.

Lefik M. 2005. *Zastosowania sztucznych sieci neuronowych w mechanice i inżynierii*. Zeszyty Naukowe. Rozprawy Naukowe, 341: 3-258.

Niedbała G., Lenartowicz T., Kozłowski J.R., Zaborowicz M. 2015. *Modelowanie neuronowe jako metoda prognozowania zawartości skrobi w ziemniakach na potrzeby Porejestrowego Doświadczalnictwa Odmianowego i Rolniczego (PDOiR)*. Nauka Przyroda Technologie, 9(2): 1-7.

Opitz D.W., Maclin R.F. 1997. *An empirical evaluation of bagging and boosting for artificial neural networks*. Proceedings of International Conference on Neural Networks (ICNN'97), 3: 1401-1405.

Paśko Ł. 2020. *Significance of Manufacturing Process Parameters in a Glassworks*. Advances in Manufacturing Science and Technology, 44(2): 39-45.

Ren P., Xiao Y., Chang X., Huang P., Li Z., Chen X., Wang X. 2021. *A Comprehensive Survey of Neural Architecture Search: Challenges and Solutions*. ACM Computing Surveys, 54(4): 1-34.

RODZIEWICZ A., PERZYK M. 2016. *Application of significance analysis to finding root causes of product defects in continuous casting of steel*. Computer Methods in Materials Science, 16(4): 187-195.

ROJEK I. 2015. *Sieci neuronowe w kontroli jakości procesu*. Studies & Proceedings Polish Association for Knowledge Management, 74: 91-100.

TADEUSIEWICZ R., HADUCH B. 2015. *Wykorzystanie sieci neuronowych do analizy danych i pozyskiwania wiedzy w systemie ekspertowym do oceny parametrów benzyn silnikowych*. Nafta-Gaz, 71(10): 776-785.

ZHIBIN W., NIANPING L., JINQUNG P., HAIJIAO C., PENGLONG L., HONGQIANG L., XIWANG L. 2018. *Using an ensemble machine learning methodology – Bagging to predict occupants' thermal comfort in buildings*. Energy and Buildings, 173: 117-127.