NEURAL MODELLING FOR THE ANALYSIS OF CHANGES IN SELECTED FEATURES OF PLANT PRODUCTS

Jędrzej TRAJER, Ewa GOLISZ, Arkadiusz RATAJSKI

Faculty of Production Engineering, Warsaw University of Life Sciences – SGGW, POLAND
E-mail of corresponding author: jedrzej_trajer@sggw.pl

Keywords: neural modelling, classification, quality assessment, sustainable agriculture

ABSTRACT

The work investigates possibilities of plant products quality assessment by means of neural networks. A quick method of plant products assessment was proposed based on the correlations occurring between selected features of plant products and neural modelling. This approach facilitates sustainable agricultural production, which often requires making decisions based on approximate but quick assessment of the quality of produced or processed products. The method of quality assessment is presented using changes in the features of pumpkin being dried as an example. Changes in selected features of chemical composition and colour were analysed, including correlations between them. Initial analysis involved cluster analysis, which allowed for grouping data into cases characterized by similar quality. Based on the analysis, a neural model was developed, which, based on easily obtainable features, allowed for classification of products according to their quality features. This approach was positively verified based on the results of chemical composition and quality assessment performed using statistical analysis of data.

INTRODUCTION

Methods used for the assessment of plant materials include organoleptic and laboratory tests. The first method involves assessing a given object by using one's own senses: sight, smell, taste, feel or hearing, while the second mainly involves assessing an object by means of appropriate equipment and analysis of physicochemical and microbiological features. Both methods are usually time-consuming and expensive. Therefore, attempts are made to improve the assessment process. In order to achieve this, other, easily obtainable features of the product are used, e.g. image features such as geometry, colour and texture, which may be correlated with other features of plant products, or dependencies between ultrasonic wave propagation and selected features of the product, Ratajski, et al (2014). This forms the base for a neural model, which allows for the assessment of product quality based on these easily measurable features.

MATERIALS AND METHODS

A database containing 39 cases of research results for three different varieties of dried pumpkin: *Ambar, Amazonka* and *Justynka* was used for the analysis, Sojak et al (2016), Król (2017). The pumpkins were dried by three methods: convection, tunnel and hybrid method. Input data used for the analysis of changes in pumpkin features being studied were chemical composition (dry mass, total and reducing sugars, lutein, lycopene and beta carotene) as well as colour discriminant in the CIE system *L, a, b*, Hunter (1948).

The dataset was analysed using cluster analysis in order to find and classify similar cases, homogeneous in terms of features. Objects belonging to the same group should be as similar as possible to one another and as different as possible from objects belonging to other groups. The classification was based on *k-means* algorithm, Hartigan (1975). Chemical composition and colour discriminants parameters were used as classification variables.

Subsequent stage of the analysis involved developing a neural model to identify similar groups of cases (of similar quality) using only three different colour discriminants in the CIE_ *Lab* system. A multilayer perceptron networks (MLP type) with the following architecture 3:k:1(4) was used, in which the number of neurons k in the hidden layer was optimized. In the approximation problem, this number, according to Heht-Nielsen (1991), should not exceed 2n+1, where *n* is the number of neurons in the input layer. The network was trained using error backpropagation algorithm using a *bootstrap* technique Efron (1982) for the development of neural models. In comparison with an ordinary random sampling (Monte Carlo method), bootstrap avoids problems of generalization resulting from finiteness of the set of cases.

RESULTS AND DISCUSSION

The results of cluster analysis performed in *Statistica Pl version 12.5* (www.statsoft.pl), indicated four clusters, fig. 1.
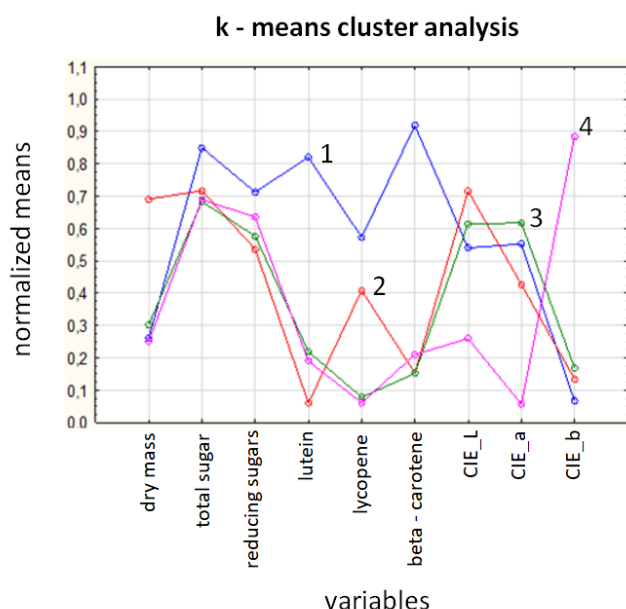

Fig.1. The results of cluster analysis for 39 cases of dried pumpkin

Cluster 1 consists of 3 cases and is characterized by the best quality features – the highest content of the examined elements (except dry mass). The second cluster consists of 9 cases and groups cases with the highest content of dry mass and lycopene. The third, and the biggest cluster, consists of 21 cases and is characterized by higher content of lutein and lycopene than the fourth cluster (which consists of 6 cases), which groups cases of the lowest quality in terms of examined features. The latter two clusters may be treated as one since the differences in chemical composition are minimal.

Table 1 shows the results of the analysis of variance *ANOVA (ANOVA* – testing significance of differences between means Ahrens (1970)) for input data describing chemical composition and colour discriminants.

Table 1. Analysis of *ANOVA* variance in the pumpkin classification task: number of clusters 4, total number of training cases: 39

Feature	Intergroup. SS	df	Intragroup SS	df	F	*p -value*

Feature	Intergroup. SS	df	Intragroup SS	df	F	p -value
dry mass	70.7	3	50.50	35	16.3410	0.000001
total sugar	374.3	3	8035.23	35	0.5435	0.655729
reducing sugars	10.3	3	272.52	35	0.4428	0.723855
lutein	178233.8	3	65736.80	35	31.6321	0.000000
lycopene	2509.6	3	1340.47	35	21.8421	0.000000
beta-carotene	17406.2	3	6106.26	35	33.2563	0.000000
CIE_L	414.6	3	529.11	35	9.1423	0.000132
CIE_a	51.8	3	36.92	35	16.3633	0.000001
CIE_b	6689.8	3	519.28	35	150.2997	0.000000

Table 1 shows that variables "total sugars" and "reducing sugars", for which *p-value* is equal more than 0.05, are of the least significance.

The performed tests suggest that the most optimal structure of the network in the classification task, Trajer et al (2012) is *MLP 3:4:1(4),* fig. 2.
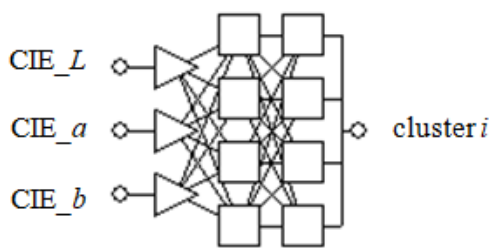

Fig.2. Neural model of the assessment of pumpkin quality

The description of the obtained network is presented in table 2.

Table 2. Parameters of the neural network for pumpkin quality assessment

Network name		Quality (training)	Quality (testing)	Quality (validation)	Training algorithm	Error function	Activation (hidden)	Activation (output)
MLP:3-4-4		95.0	100.0	100.0	BFGS 25	Entropy	Tanh	Softmax

Sensitivity analysis allowed for defining the degree of importance of each colour discriminant. Colour discriminant CIE_b, which describes share of yellow or blue (shades of yellow are positive values, and shades of blue are negative values) had the greatest impact on the results of classification, table 3.

Table 3. Results of sensitivity analysis of a neural model for the assessment of dried pumpkin quality.

MLP 3-4-4	CIE_b	CIE_a	CIE_L
Value of error	235.9761	41.84554	27.23396

The developed model for the assessment of dried pumpkin quality allows for fast monitoring of quality changes. The gains chart for the four analysed quality classes, fig. 3, graphically indicate the correctness of classification. Gain is defined as: *Gain = (Expected Response Using Predictive Model)/(Expected Response From Random Mailing)*, Brandenburger and Furth (2009). Gains chart is a graphical presentation of the usefulness of the model for predicting the value of dependent categorical variable

assuming two values. Gains charts may be used for most methods used for predicting the object assignment to a given class (in case of two or more classes). When a modelled categorical variable assumes more values, we may create gains charts separately, for each category.
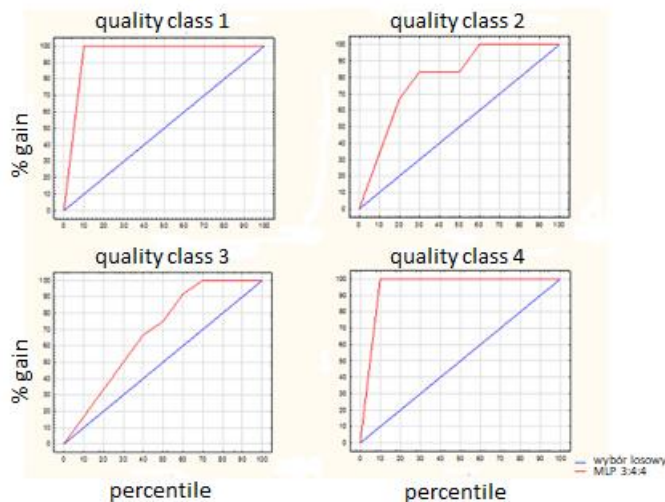

Fig. 3. Gains graph for the identified four quality classes of pumpkin

The correctness of classification of pumpkin was also confirmed by analogical results obtained using statistical data analysis: three classes in which similar cases were identified, Sojak *et al* (2016).

CONCLUSIONS

The developed neural model, based on easy to determine colour determinants in the CIE_*Lab* system allows for obtaining quick assessment of the product being examined. Training quality is equal 95%, i.e. the correctness of classification performed by the network is good. The performed sensitivity analysis showed that the most significant variable in the model was CIE_*b* – discriminant of yellow colour. Cluster analysis of 39 cases of dried pumpkin showed that the data may be divided into 4 sets (clusters), which identify four groups of pumpkin cases of similar quality. The developed MLP *3:4:1(4)* model allows for efficient classification of pumpkin into different groups of quality based on colour discriminants only, which is also indicated by positive verification of the developed neural model. This approach facilitates sustainable agricultural production, enabling quick assessment of the quality of manufactured or processed products.

REFERENCES

Ahrens, H. (1970). Analiza Wariancji. *PWN*, Warszawa.

Brandenburger, T., Furth, A. (2009). Cumulative Gains Model Quality. *Journal of Applied Mathematics and Decision Sciences.* Vol. 2009.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *Pa. Society for Industrial and Applied Mathematics.* Philadelphia.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

Heht-Nielsen, R. (1991). Neurocomputing. *Addison Wesley*, Amsterdam.

Hunter, R. S. (1948). Photoelectric Color-Difference Meter". *JOSA*. *38 (7): 661.* (Proceedings of the Winter Meeting of the Optical of America)

Król, M. (2017). Wykorzystanie eksploracyjnej analizy danych w ocenie efektów suszenia dyni. *Praca magisterska,* WIP SGGW, Warszawa.

Ratajski, A. Mikš-Krajnik M., Białobrzewski I. (2014). The use of ultrasonic measurements for the determination of particle size distribution in diluted tomato paste, *International Journal of Food Science and Technology*, 49, pp. 288–293

Sojak, M., *at all* (2016). The effect of drying and long-term storage on colour and carotenoids content of giant pumpkin (*Cucurbita Maxima*). *Technical Sciences*, Olsztyn,19(4), pp. 295-312

StatSoft Polska: www.statsoft.pl

Trajer, J., Paszek A., Iwan S. (2012). Zarządzanie wiedzą*, PWE,* Warszawa.