Prediction of the Kind of Sprouts of *Cruciferae* Family Based on Artificial Neural Network Analysis

Adam BUCIŃSKI[1,2], Henryk ZIELIŃSKI[1] and Halina KOZŁOWSKA[1]

[1]Division of Food Science, Institute of Animal Reproduction and Food Research of Polish Academy of Sciences, Tuwima, Olsztyn, Poland; [2]Department of Biopharmacy, Faculty of Pharmacy, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland

Abstract

Buciński A., Zieliński H., Kozłowska H. (2007): **Prediction of the kind of sprouts of *Cruciferae* family based on artificial neural network analysis.** Czech J. Food Sci., **25**: 189–194.

The aim of this work was to show that artificial neural networks (ANNs) are a convenient tool for predicting the kind of sprouts originated from *Cruciferae* family. For this purpose, the known contents of bioactive compounds of small radish, radish, white mustard, and rapeseed seeds and sprouts were used for the prediction of the kind of a specific sprout. The input data reflected the contents of the following compounds in cruciferous seeds in the course of germination: soluble proteins (SP), ascorbic acid (AH_2), total glucosinolates (GLS), reduced glutathione (GSH), and tocopherols (α-T, β-T, γ-T, δ-T), expressed in their biological activity. The ANN used was trained on the learning set. The ability of the utilised ANN to generalise the gained knowledge based on the learning set was verified by the validating and testing sets. The trained and validated ANN was able to classify, with complete accuracy, the kind of sprouts out of the four kinds used. It can be concluded that ANN can be used as a useful tool for determining the identity of cruciferous seeds and sprouts based on the determined levels of their bioactive components.

Keywords: artificial neural network; sensitivity analysis; cruciferous sprout classification; bioactive compounds

The artificial neural network (ANN) analysis is a method of data analysis, which is to emulate the brain's way of work. ANN is a group of connected elements, ordered in layers, which are able to process information (Zupan & Gasteiger 1993). There are three kinds of layers: the input layer, one or more hidden layers, and the output layer (Bishop 1995). The elementary units of the network, which are called artificial neurons, are connected with one another with a different strength, and the information is encoded in the strength of the network "synaptic" connections (Rumelhart *et al*. 1986). These connections are called synaptic weights. All the information about a network is encoded in weights, because those weights are in fact numbers which determine the strength of the stimuli coming to the neurons. The most important feature of the artificial neural networks is their learning process. Learning of networks is realised by changing the values for all synaptic weights with the use of a specific algorithm. The most widely known and used learning algorithm is the so-called back-propagation training algorithm. The network learning process making use of that algorithm takes an error between the current and the desirable output of the network to improve the values of synaptic weights. Therefore, artificial neural networks differ from classical computer programs in that they "learn" from a set of examples rather than being programmed to get the right answer. For this reason, ANNs provide a simple means for predicting the outcomes that depend upon complex, possibly nonlinear, relationships between many variables (Kaliszan 1997). In some cases, sensitivity analysis is also included to give an important insight into the usefulness of individual variables (Bączek *et al.* 2004).

Vegetable materials contain many bioactive compounds which are of interest to the food industry (Hagerman *et al.* 1998). Biologically active compounds are also of interest to consumers as protective agents against diseases caused by reactive oxygen species (ROS), since a strong inverse relationship between their contents and the intake of fruit and vegetables was shown (Temple 2000). Several studies analysed bioactive compounds of a wide variety of vegetables and fruits, beverages, green and black teas, different parts of the plant (leaves, roots and bark), seeds and hulls, as well as in industrial residues and by-products other than those of the plant origin (Moure *et al.* 2001).

It is well known that the consumption of a plant-based diet, mainly vegetables, fruits, and whole grains, is recommended as one of the ways of lowering the risk of human diseases, and responds to the consumers' demand for healthy and functional foods (Roberfroid 2000). While searching for new sources of functional food, special attention has been paid to sprouts originated from *Cruciferae* family that have been increasingly used in human diets. The sprouts may thus become a potential source of nutritious food or a food ingredient.

Previously, we determined the contents of biologically active compounds and trolox equivalent antioxidant capacity (TEAC) of cruciferous seeds in the course of germination. These data were used to obtain a model of ANN for the prediction of the trolox equivalent antioxidant capacity of a range of germinated cruciferous seeds on the basis of the contents of biologically active compounds and the experimentally determined antioxidant capacity of germinated cruciferous seeds (Buciński *et al.* 2004).

The purpose of this study was to show that artificial neural networks (ANNs) are a convenient tool for predicting the kind of sprouts originated from *Cruciferae* family on the basis of the contents of soluble proteins (SP), ascorbic acid (AH_2), total glucosinolates (GLS), reduced glutathione (GSH), and tocopherols (α-T, β-T, γ-T, δ-T). Moreover, the sensitivity analysis was designed in order to discriminate which of the bioactive compounds tested are significant and important for the quality of the prediction of the kind of sprouts originated from *Cruciferae* family.

MATERIAL AND METHODS

Material. Single cruciferous seed samples were obtained from a local plant breeding station in north-eastern Poland. The samples included rapeseed (*Brassica napus* var. *oleifera*), white mustard (*Sinapis alba* L.), radish (*Raphanus sativus* L.), and small radish (*Raphanus sativus* var. *sativus*).

Seed germination. Germination was carried out in an incubator (Cliambic cabinet, model Economic Delux EC00-065, Snijders Scientific b.v., the Netherlands). The seeds were germinated in light and dark conditions, at 25°C, for up to 7 days. The seeds were layered over a moist filter paper (qualitative medium-speed filter paper) to one-third of the depth of the paper. Sprouts from each species were removed from the incubator every 24 h, frozen in liquid nitrogen and lyophilised. The germination was carried out in triplicates.

Analytical methods. Tocopherols (α-T, β-T, γ-T, δ-T) were assayed by high-performance liquid chromatography (HPLC) according to the method described by Peterson and Qureshi (1993), and soluble proteins (SP) by the dye-binding method of Bradford (1976). The content of total GLS (sum of aliphatic, indolic and aryl GLS) was evaluated by HPLC analysis according to the method of Heaney *et al.* (1986), whereas ascorbic acid (AH_2) was determined using the HPLC method as described by Oruna-Concha *et al.* (1998). The contents of total GLS and AH_2 were expressed in μmoles/g d.m., whereas SP as μmoles of albumin/g d.m., and tocopherols (μmoles/g d.m.) as d-α-tocopherol equivalents (α-TE) after calculations using current biological activities of 1.0 for α-T, 0.5 for β-T, 0.1 for γ-T, and 0.03 for δ-T (Czajka-Marins 1996; Eitenmiller *et al.* 1998).

ANN analysis. Artificial neural networks (ANNs) were run on a personal computer using Statistica Neural Networks software, Ver. 6 (StatSoft, Tulsa, OK, USA). The ANNs used were based on a multilayer perceptron. The input data reflected the contents of the following bioactive compounds in cruciferous seeds in the course of germination: soluble proteins (SP), ascorbic acid (AH_2), total glucosinolates (GLS), reduced glutathione (GSH),

Table 1. Sample fragment of the input data considered in ANN analysis

Case	GLS	AH_2	GSH	α-T	β-T	γ-T	δ-T
RW-L/I/1	175.59	0.521	0.176	0.0030	0.0088	0.0415	0.0009
RW-D/III/5	150.63	10.241	0.259	0.2040	0.0260	0.0633	0.0013
R-L/II/3	74.71	20.839	0.337	0.0363	0.0027	0.0133	0.0001
R-D/I/2	90.70	6.720	0.813	0.0486	0.0113	0.0573	0.0005
G-L/III/6	217.41	14.690	0.062	0.1714	0.0274	0.0087	0.0000
G-D/III/4	232.63	3.950	0.123	0.0211	0.0171	0.0038	0.0001
RZ-L/III/5	5.86	30.269	0.128	0.2312	0.0140	0.0158	0.0004
RZ-D/III/1	5.92	0.682	1.844	0.1278	0.0176	0.0252	0.0001

Description of cases: RW – radish seeds, R – small radish seeds, G – white mustard seeds, RZ – rapeseeds; Arabic numeral – day of germination; Roman numeral – number of repeated germination; L – light conditions; D – dark conditions

and tocopherols (α-T, β-T, γ-T, δ-T) expressed in their biological activity, and listed in Table 1. For this reason, the input layer contained 7 neurons. On the base of the input data, the ANN was responsible for the classification of the individual object (described by the content of bioactive compounds) as one of the four kinds of sprouts (rapeseed – *Brassica napus* var. *oleifera*, white mustard – *Sinapis alba* L., radish – *Raphanus sativus* L., and small radish – *Raphanus sativus* var. *sativus*). Therefore, the number of neurons in the output layer was exactly 4 neurons. Then, the next step in this study was to find out the neuron number in the hidden layer. Then, the ANNs with different neuron numbers in the hidden layer, containing 2, 3, and 4 neurons, were tested. The optimal ANN was that with 4 neurons in the hidden layer.

The architecture of the model utilised is depicted in Figure 1. A supervised method of learning with a back-propagation strategy was used. Learning of the ANN was realised during 5000 epochs. Then, the learning was continued with the use of the conjugated gradient (CG) descent algorithm as long as the root-mean-squared (RMS) error reached the smallest value. The data for analysis, reflecting the contents of biologically active compounds of cruciferous seeds in the course of germination, were randomly divided into three sets: the learning set with 86 objects, the validating set with 43 objects, and the testing set with 43 objects. A sample fragment of the input data considered in the ANN analysis is shown in Table 1. The variables considered in this study were converted with the use of the Minimax method. It made it possible to scale those data into the 0–1 range. The learning of the ANN was executed with the learning coefficient equal to 0.01 and momentum equal to 0.3. The data from the learning set were presented in a randomised manner during the learning process. Changes in RMS error were also recorded for the training and validating sets during the learning process. Following ANN classification analysis, a sensitivity analysis for the input data was carried out.

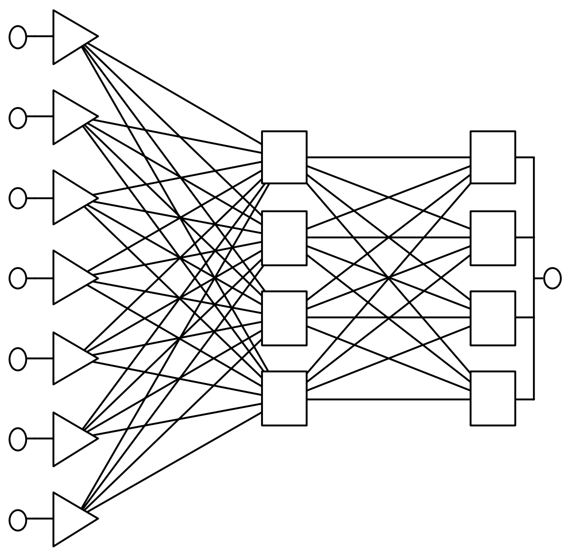

Figure 1. The architecture of artificial neural networks used for predictions of the kind of sprouts originated from *Cruciferae* family

RESULTS AND DISCUSSION

The purpose of this study was to show that artificial neural networks (ANN) are a convenient tool for predicting the kind of sprouts originated from *Cruciferae* family based on the contents of bioactive compounds. Previous studies showed the benefits of using artificial neural networks in the food industry and food processing, for example to control bread making, extrusion and fermentation processes, quality differences in soy sauce, quality of potatoes in the chip industry, prediction of milk shelf-life (Batchelor 1993; Latrille *et al.* 1993; Eerikanen & Linko 1995; Vallejo-Cordoba *et al.* 1995; Iizuka & Aishima 1997; Marique *et al.* 2003). However, no attempt has been made till now to use the artificial neural networks to predict the kind of sprouts.

In this study, the approach has been tested on the material available in different time and light conditions for cruciferous seeds during the course of germination. The reason for taking interest in cruciferous sprouts was the available evidence demonstrating that cruciferous vegetables such as broccoli, cabbage, cauliflower, and others contain among their bioactive compounds such that are able to induce the synthesis of detoxifying enzymes and may thereby be anticarcinogenic (Nestle 1998). Then, the prediction of the kind of sprouts originated from *Cruciferae* family may be helpful for the quality control.

The ANN used, which was trained on the learning set, generalised the prediction ability obtained with regard to the data contained in the validating and testing sets. The detailed data were published previously by Buciński *et al.* (2004). The learning process was completed when the artificial neural network reported the smallest RMS error with regard to the validating set of data. In the case of this network, the learning was completed in 5000 epochs by the Back Propagation (BP) method and 4 epochs by the Conjugate Gradient Descent (CGD) method (Figure 2).

The results of the study performed showed that the trained and validated model of the neural network was able to classify the investigated material without any error. The classification of the investigated sprouts with the use of the designed ANN for the learning, validating, and testing sets of data is presented in Table 2. It was also observed that the time and light conditions of germination did not influence the predictive ability of the constructed ANN model. The results of this study confirm the benefits of learned ANN as regards the prediction of the kind of sprouts originated from *Cruciferae* family.

In the next part of this study, the variables used were estimated by using the sensitivity analysis. The sensitivity analysis rates the importance of a variable with respect to the ANN model used. On the basis of the sensitivity analysis, it is possible to show which bioactive compound or group of bioactive compounds are most significant (with the "ratio" value above 1) and least significant (when the "ratio" value is below 1) for the proposed ANN model. The parameter simply describes

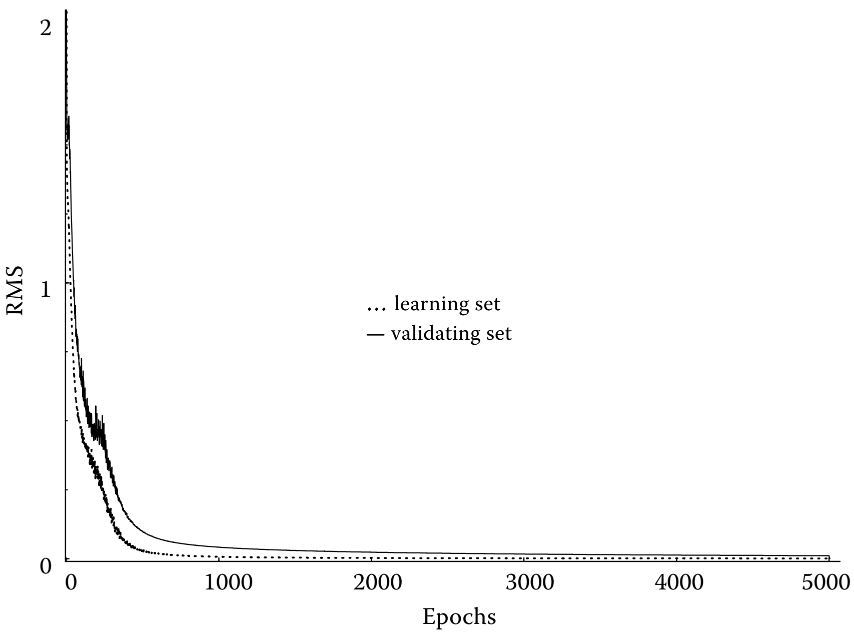

Figure 2. Error training graph

Table 2. Classification of the kind of sprouts with the use of the designed ANN for learning, validating and testing sets of data

Set	Kind of sprouts	Number of cases	Number of correctly classified cases	Percent of correct classification
Learning	radish sprouts	22	22	100
	small radish sprouts	28	28	100
	white mustard sprouts	22	22	100
	rapeseed sprouts	14	14	100
Validating	radish sprouts	12	12	100
	small radish sprouts	7	7	100
	white mustard sprouts	9	9	100
	rapeseed sprouts	15	15	100
Testing	radish sprouts	9	9	100
	small radish sprouts	8	8	100
	white mustard sprouts	12	12	100
	rapeseed sprouts	14	14	100

the ratio between an error and the baseline error (i.e. the error of the network when all these variables are available). If the "ratio" is less than one, then removing a variable from the proposed ANN model either has no effect on the performance of the network or can even enhance it. The higher the parameter ratio, the greater will be neural network prediction error in the case when the indicated variable is removed from the input data in comparison with the complete input data. In summary, as error is increasing, the neural network is more sensitive to the lack of the variable in question. The result of the sensitivity analysis is summarised in Table 3. All data reflecting the contents of bioactive compounds in sprouts were important for the ANN used. The importance of variables is ranked according to the following order: GLS > δ-T > GSH > α-T > β-T > γ-T > AH_2. The ratio parameter tells about the behaviour of the neural network when a variable is removed from the input data.

Table 3. Results of the sensitivity analysis for the variables considered in ANN analysis

Variables (µmol/g d.m.)	Ratio	Rank
GLS	14 375.6	1
δ-T	5041.1	2
GSH	2006.8	3
α-T	1339.9	4
β-T	894.3	5
γ-T	127.8	6
AH_2	3.4	7

CONCLUSION

The results obtained in this study indicate that the artificial neural network (ANN), a convenient and cheap tool, can be a promising method for predicting the kind of sprouts originated from *Cruciferae* family. Therefore, the ANN analysing the contents of bioactive compounds in sprouts may find its application in the quality control. It may be suggested that the decision system based on an artificial neural network may be a convenient tool for the automatic recognition of the plant material used in the food industry, supporting decisions being made by an analyst or even replacing him.

References

Bączek T., Buciński A., Ivanov A.R., Kaliszan R. (2004): Artificial neural network analysis for evaluation of peptide MS/MS spectra in proteomics. Analytical Chemistry, **76**: 1726–1732.

Batchelor B.G. (1993): Automated inspection of bread and loaves. In: Batchelor B.G., Solomon S.S., Waltz F.M. (eds): Proceeding SPIE – Volume 2064. Machine Vision Applications, Architectures, and Systems Integration II. August 1993. The International Society for Optical Engineering: 124–134.

Bishop C. (1995): Neural Networks for Pattern Recognition. University Press, Oxford.

Bradford M.M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, **72**: 248–254.

Buciński A., Zieliński H., Kozłowska H. (2004): The use of artificial neural networks for prediction of the total antioxidant capacity of cruciferous sprouts. Trends in Food Science and Technology, **15**: 161–169.

Czajka-Marins D.M. (1996): Vitamins. In: Mahan L.K., Escott-Stump S. (eds): Food, Nutrition and Diet Therapy. 9th Ed. W.B. Saunders Company, Philadelphia, London, Toronto, Montreal, Sydney, Tokyo: 77–122.

Eerikanen T., Linko P. (1995): Neural network based food extrusion cooker control. In: Proceedings of the International Conference EANN'95. August 1995 . Otaniemi, near Helsinki: 473–476.

Eitenmiller R.R., Landen W.O., Augustin J. Jr. (1998): Vitamin Analysis. In: Nielsen S.S. (ed.): Food Analysis, 6th Ed. Aspen Publishers Inc., Gaithersburg, Maryland: 181–191.

Hagerman A.E., Riedl K.M., Jones G.A., Sovik K.N., Ritchard N.T., Hartzfeld P.W., Riechel T.L. (1998): High molecular weight plant polyphenolics (tannins) as biological antioxidants. Journal of Agricultural and Food Chemistry, **46**: 1887–1892.

Heaney R.K., Spinks E.A., Hanley A.B., Fenwick G.R. (1986): Technical Bulletin: Analysis of Glucosinoltes in Rapeseed. AFRC, Food Research Institute, Norwich.

Iizuka K., Aishima T. (1997): Soy sauce classification by geographic region based on NIR spectra and chemometrics pattern recognition. Journal of Food Science, **62**: 101–104.

Kaliszan R. (1997): Structure and Retention in Chromatography. A Chemometric Approach. Harwood Academic Publishers, Amsterdam.

Latrille E., Corrieu G., Thibault J. (1993): pH prediction and final fermentation time determination in lactic acid batch fermentations. Computers & Chemical Engineering, **17** (Supplement 1): S423–S428.

Marique T., Kharoubi A., Bauffe P., Ducattillon C. (2003): Modeling of fried potato chips color classification using image analysis and artificial neural network. Journal of Food Science, **68**: 2263–2266.

Moure A., Cruz J.S., Franco D., Dominguez J.M., Sineiro J., Dominguez H., Nunez M.J., Parajo J.C. (2001): Natural antioxidants from residual sources. Food Chemistry, **72**: 145–171.

Nestle M. (1998): Broccoli sprouts in cancer prevention. Nutrition Review, **56**: 127–130.

Oruna-Concha M.J., Gonzalez-Castro M.J., Lopez-Hernandez J., Simal-Lozano J. (1998): Monitoring of the vitamin C content of frozen green beans and padron peppers by HPLC. Journal of the Science of Food and Agriculture, **76**: 477–480.

Peterson D.M., Qureshi A.A. (1993): Genotype and environment effects on barley and oats. Cereal Chemistry, **70**: 157–162.

Roberfroid M.B. (2000): Defining functional foods. In: Gibson G.R., Williams C.M. (eds): Functional Foods: Concept to Product. Woodhead Publishing Limited, Cambridge.

Rumelhart D.E., Hinton G.E., Williams R.J. (1986): Learning internal representations by error propagation. In: Rumelhart D.E., McClelland J.L. (eds): Parallel Distributed Processing, Vol 1. MA: MIT Press, Cambridge.

Temple N.J. (2000): Antioxidants and disease: more questions than answers. Nutrition Research, **20**: 449–459.

Vallejo-Cordoba B., Arteaga G.E., Nakai S. (1995): Predicting milk shelf-life based on artificial neural networks and headspace gas chromatographic data. Journal of Food Science, **60**: 885–888.

Zupan J., Gasteiger J. (1993): Neural Networks for Chemists. An Introduction. Wiley-VCH, Weinheim.

Received for publication July 12, 2006
Accepted after corrections April 4, 2007

Corresponding author:

Prof. Dr. Adam Buciński, Nicolaus Copernicus University, Collegium Medicum, Faculty of Pharmacy, Department of Biopharmacy, Jagiellońska 13, 85-067 Bydgoszcz, Poland
tel.: + 52 585 39 09, fax: + 52 585 38 04, e-mail: kizbiofarmacji@cm.umk.pl