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

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

**Abstract**


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 (AH₂), 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
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 (Kaliszczuk 1997). In some cases, sensitivity analysis is also included to give an important insight into the usefulness of individual variables (Baczek 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 (Bucinski 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 ($\text{AH}_2$), total glucosinolates (GLS), reduced glutathione (GSH), and tocopherols ($\alpha$-TE, $\beta$-TE, $\gamma$-TE, $\delta$-TE). 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 rape-seed (Brassica napus var. oleifera), white mustard ($\text{Sinapis alba}$ L.), radish ($\text{Raphanus sativus}$ L.), and small radish ($\text{Raphanus sativus}$ 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 $25^\circ$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 ($\alpha$-T, $\beta$-T, $\gamma$-T, $\delta$-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 ($\text{AH}_2$) was determined using the HPLC method as described by Oruna-Concha et al. (1998). The contents of total GLS and $\text{AH}_2$ were expressed in mmoles/g d.m., whereas SP as mmoles of albumin/g d.m., and tocopherols (mmoles/g d.m.) as $d$-$\alpha$-tocopherol equivalents ($\alpha$-TE) after calculations using current biological activities of 1.0 for $\alpha$-T, 0.5 for $\beta$-T, 0.1 for $\gamma$-T, and 0.03 for $\delta$-T (Czajka-Marins 1996; Eitzenmiller 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 ($\text{AH}_2$), total glucosinolates (GLS), reduced glutathione (GSH),
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.

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

<table>
<thead>
<tr>
<th>Case</th>
<th>GLS</th>
<th>AH₂</th>
<th>GSH</th>
<th>α-T</th>
<th>β-T</th>
<th>γ-T</th>
<th>δ-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW-L/I/1</td>
<td>175.59</td>
<td>0.521</td>
<td>0.176</td>
<td>0.0030</td>
<td>0.0088</td>
<td>0.0415</td>
<td>0.0009</td>
</tr>
<tr>
<td>RW-D/III/5</td>
<td>150.63</td>
<td>10.241</td>
<td>0.259</td>
<td>0.2040</td>
<td>0.0260</td>
<td>0.0633</td>
<td>0.0013</td>
</tr>
<tr>
<td>R-L/II/3</td>
<td>74.71</td>
<td>20.839</td>
<td>0.337</td>
<td>0.0363</td>
<td>0.0027</td>
<td>0.0133</td>
<td>0.0001</td>
</tr>
<tr>
<td>R-D/I/2</td>
<td>90.70</td>
<td>6.720</td>
<td>0.813</td>
<td>0.0486</td>
<td>0.0113</td>
<td>0.0573</td>
<td>0.0005</td>
</tr>
<tr>
<td>G-L/III/6</td>
<td>217.41</td>
<td>14.690</td>
<td>0.062</td>
<td>0.1714</td>
<td>0.0274</td>
<td>0.0087</td>
<td>0.0000</td>
</tr>
<tr>
<td>G-D/III/4</td>
<td>232.63</td>
<td>3.950</td>
<td>0.123</td>
<td>0.0211</td>
<td>0.0171</td>
<td>0.0038</td>
<td>0.0001</td>
</tr>
<tr>
<td>RZ-L/III/5</td>
<td>5.86</td>
<td>30.269</td>
<td>0.128</td>
<td>0.2312</td>
<td>0.0140</td>
<td>0.0158</td>
<td>0.0004</td>
</tr>
<tr>
<td>RZ-D/III/1</td>
<td>5.92</td>
<td>0.682</td>
<td>1.844</td>
<td>0.1278</td>
<td>0.0176</td>
<td>0.0252</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

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

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
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 > AH₂. The ratio parameter tells about the behaviour of the neural network when a variable is removed from the input data.

**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**


Received for publication July 12, 2006

Accepted after corrections April 4, 2007