Scientia Horticulturae 218 (2017) 222–229

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Application of neural image analysis in evaluating the quality of greenhouse tomatoes

Maciej Zaborowicz*, Piotr Boniecki, Krzysztof Koszela, Andrzej Przybylak, Jacek Przybył

Poznań University of Life Sciences, Institute of Biosystems Engineering, Wojska Polskiego 50, 60-627 Poznań, Poland

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 13 January 2017
Accepted 1 February 2017
Available online 27 February 2017

Keywords:
Artificial neural networks
Classification of tomatoes
Tomatoes quality
Image analysis

ABSTRACT

The paper presents the research on the use of methods of computer image analysis and artificial neural modeling in the process of assessing the quality of greenhouse tomatoes variety Cappricia. The subject of the study was tomatoes of the sizes from 40 mm to 67 mm and the colours: 1–6, include intermediate colours. The process of image acquisition and obtaining empirical data was conducted throughout the entire growing season in the period from the first harvest in the middle of May to the last harvest at the beginning of November. Satisfactory quality characteristics were obtained in the case of the *RBF 37:37-39-1:1* and *RBF 22:22-20-2:2* models. *RBF 37:37-39-1:1* network, whose output variable was the colour of the tomato, the training quality was 0.930827, the validation quality was 0.911982, and the test quality was 0.979390. The *RMSE* rate of network training for the training set was 0.075986, for the validation set it was 0.072194, and for the test set 0.061714. For the *RBF 22:22-20-2:2* network, the variables were colour and hardness. This network is characterised by a training quality of 0.985038, a validation quality of 0.990694 and a test quality of 0.985130. The *RMSE* rate of the network training is 0.065667, of the network validation it is 0.066187 and of the network test 0.073868.

The research showed that performing a correct classification requires taking two digital images of the examined tomatoes, one of the stem and one of the front of the tomato, and generating training sets that contain average values for the extracted characteristics.

1. Introduction

The Common Agricultural Policy of the Member States of European Union puts special emphasis on the high efficiency of agricultural production and the satisfactory quality of products on the market. There is a growing demand among both legislators and consumers for the sales of certified agricultural produce, which forces producers to observe a number of standards and restrictions (Butz et al., 2005).

The gradual taking over of the market for the distribution of fresh fruit and vegetables by retail chains has been observed in recent years. In fighting for customers, these chains treat the high quality of the provided product as their priority. Accordingly, retail chains impose on their suppliers quality-control obligations related to food products (Clement et al., 2012).

Tomatoes are among the many products sold by retail chains and food corporations. Due to their high supply and availability, and significant nutritional value, tomatoes have a considerable share of total vegetable consumption. Consumers notice the quality and taste of tomatoes on sale and also have certain preferences as to their appearance. Each EU Member State has its own standards regulating the quality of fruits and vegetables. However, the European Union has issued its own regulations regarding market standards for edible tomatoes. The currently binding standard describing the quality of tomatoes is included in Commission Regulation (EC) No 771/2009 of 25 August, 2009. This standard is applicable in all Member States of the European Community and is identical for all growers. In addition, retail chains sometimes set their own quality criteria to cater for their customer's preferences.

Quality evaluation is conducted by qualified graders on the basis of their specification, i.e. the product catalogue. Despite the introduction of rigorous standards, this process remains highly subjective. Each grader might perceive colours differently, focusing on different characteristics. Particular attention should be paid to the conditions in which evaluation is conducted (Goel and Sehgal 2015). Occasionally, graders assign different quality grades to the same produce. From time to time the produce is declared inconsistent with the order and rejected by the recipient.

Food producers, retail chains and scientists are making efforts to increase the objectivity of fruit and vegetable quality evaluation,

* Corresponding author.
E-mail address: maciej.zaborowicz@up.poznan.pl (M. Zaborowicz).

http://dx.doi.org/10.1016/j.scienta.2017.02.001
0304-4238/

to facilitate quick, objective and the clear evaluation of the produce put on the market (Du and Sun, 2006; Zhang et al., 2014a). No method has yet been developed that demonstrates higher-quality evaluation than qualified graders and assists them in their work (Lu et al., 2009; Iraji and Tosinia, 2011; Pabico et al., 2012).

It appears sensible, then, to use computer image analysis methods and neural network modelling that perform the evaluation of the marketable quality of greenhouse tomatoes similarly to humans (Boniecki et al., 2012a, 2012b, 2015).

Recent years have seen a considerable increase in the use of neural network-aided image-analysis methods to tackle agriculture-related issues, including those of agricultural engineering (Patil and Kumar, 2011; Gomes and Leta, 2012; Boniecki et al., 2012a; Vazquez-Cruz et al., 2013; Pentos et al., 2014). These methods increasingly often assist decision-making processes in various aspects of plant and animal production and in the process of quality evaluation and control (Gastélum-Barrios et al., 2011; Soares et al., 2013; Kujawa et al., 2014; Zhang et al., 2014b; Mohammad Reza Naroui Rad et al., 2015; Pentos, 2016; Przybylak et al., 2016).

The Yudong Zhang team (Zhang et al., 2016) presented how can be classify 18 types of fruits using in this process the 1653 processed images by using networks feed-forward neural network. The team obtained the result on the level of accuracy on 89%. It should be noted that there were used the images of relatively low resolution which also makes use of this method very interesting. Wang with the research team (Wang et al., 2015) presented a classification of 18 types of fruit based on digital images. They obtained the result of accuracy between 85% and 89%, depending on the methods of processing and analyzing data. Wu with the research team (Wu et al., 2016) developed a method to identify the tea category. It is based on pattern recognition with using tools wavelet entropy algorithm and k-Nearest Neighbors. The research team received a result at an accuracy level of 95% with a very short working time at the developed system. The application of these methods shows that the use of different techniques of image processing and analysis and the use of methods of neural modeling is appropriate to resolve this type of problems.

The purpose of the research described here is to prepare a new method for the quality evaluation of agricultural produce, including greenhouse tomatoes, using neural network and image analysis. The work began from series of trial researches (Zaborowicz et al., 2013) where was used authorship program *PiAO* (*pol. Przetwarzanie i Analiza Obrazu*, *eng. Image Processing and Analysis*) to analyze 144 variables from 2163 photos of tomatoes. It had been shown the usefulness of the produced software in the process of data acquisition and neural modelling. The trial researches obtained the following results: neural network *RBF 18:18-8-1:1* quality of learning set 0.825, 0.804 validation set and test set 0.763, the error of learning set 0.744, 0.751 validation set and test 0.704. These results formed the basis of further researches. The next step was to reduce the number of variables extracted from the image and to separate in the *PiAO* software tested objects from the background (Zaborowicz et al., 2014). It reduced the training variables and acquire data of higher accuracy due to achieve the objective present work.

2. Materials and methods

The research method consisted of six stages. In the first stage the appropriate tomato variety was selected. Next, an appropriate test station for digital-image acquisition was constructed. A digital camera was used to capture photographs of the studied tomatoes. The resulting photographs were then processed and IT systems were generated to facilitate the extraction of the characteristics of the studied objects. The collected data was converted into training sets to be used in generating neural network models and in the training process. Finally, the generated neural network model was verified.

2.1. Research material

The research material was round, “Extra” class, tomatoes of the *Cappricia* variety, from the greenhouse cultivation. The subject of the study was tomatoes of the following sizes: 40–46 mm; 47–51 mm; 52–56 mm; 57–62 mm; & 57–67 mm; and the following colours: 1; 2; 3; 4; 5; & 6, and the intermediate colours between 2 and 3, 3–4 & 4–5.

The process of image acquisition and obtaining empirical data was conducted throughout the entire growing season that is typical for greenhouse tomatoes, i.e. in the period from the first harvest in the middle of May to the last harvest at the beginning of November – every two weeks.

The research process included selected tomatoes, which left the warehouse of the grower and were intended to be transported and sold to external customers. It was the material that after the growing season ended was gathered and ended up on the automatic sorting table, where it was sorted with regard to ripeness (skin colour) and size. The tomatoes sorted and selected in this way were directed to the warehouse (storage facility) where they awaited transport. Prior to being transported these vegetables were manually evaluated by qualified experts, according to the quality requirements specified in the standards and according to the criteria provided by the clients. In the tomato production-and-supply chain it is this stage in the quality evaluation and inspection that gives rise to most difficulties for the growers.

2.2. Research methodology

The research was performed with the use of the test equipment presented below and a station prepared specifically for the purpose of image acquisition (Fig. 1).

The test kit consisted of the following elements: hardness tester, electronic scale, slide caliper. The test station consisted of: NIKON D7000 digital camera with NIKKOR AF-S DX 18–105 mm f/3.5–5.6 VR ED lens, tripod, 75 × 75 × 75-cm light tent allowing the uniform dispersion of light, black background, two 30 W lamps with a colour temperature of 5500 K.

In order to create the neural network classification model it was necessary to gather the appropriate input data-training cases for the neural network model. This is when empirical data for the studied tomatoes was acquired, such as their mass, hardness, height (h), length (a) and width (b) (Fig. 2). The image-acquisition process was performed at the same time. Each of the studied tomatoes was photographed twice: on the stem side and on the front side (Fig. 2). A total of 4326 photographs of the studied objects were taken.

The exposure parameters of the photographs were determined based on a grey card. The image-acquisition process was conducted with the use of identical parameters of the camera, i.e. a constant focal length of 35 mm, exposure: diaphragm f/8 and exposure time 1/5s (Fig. 3). The diaphragm was selected first and then the appropriate exposure time was selected with the use of the light meter built into the camera. The exposure metering mode in the camera was set to centre-weighted. All the photographs were taken with the use of ISO 100 sensitivity, without the use of white balance adjustment or exposure adjustment, and without the use of a flash lamp. Lens stabilisation was disabled. White balance was set to the default value.

The photographs were saved, without any loss, into the *RAW* format *NEF* (*Nikon Electronic File*) files with 14-bit colour depth. Recording in the *sRGB* (*standarised Red Green Blue*) colour space was used. Each image had a size of 4928 × 3264 pixels.

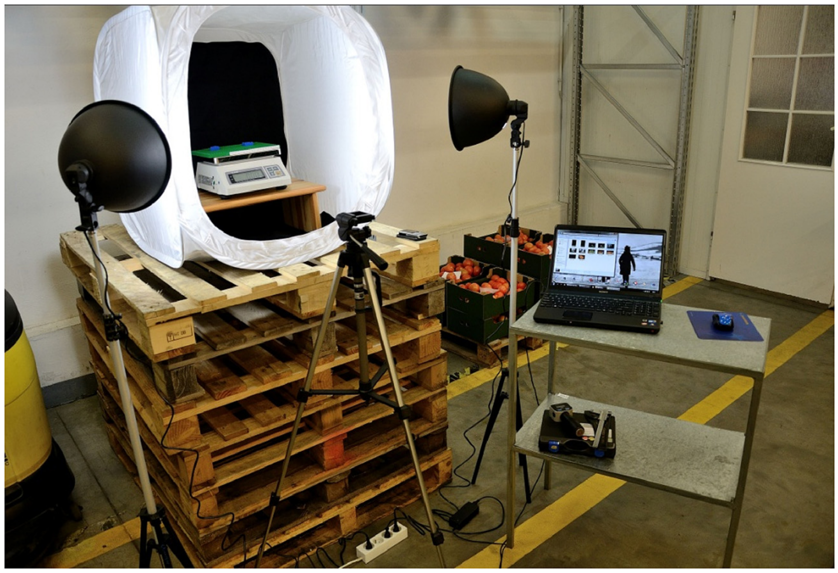

Fig. 1. Image acquisition test station.

(source: own materials)

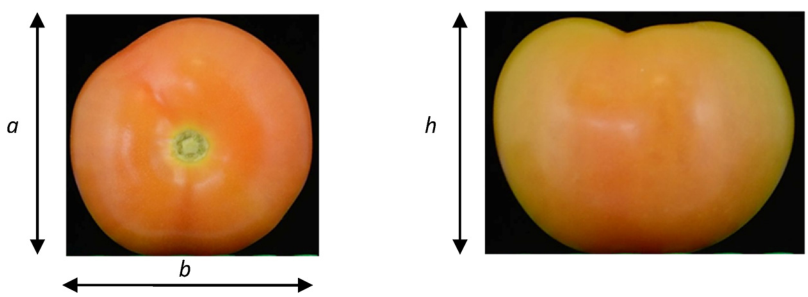

Fig. 2. Views of a tomato on the stem side (on the left) and on the front side (on the right).

(source: own materials)

Fig. 3. The original form of the analysed image of a tomato.

(source: own materials).

The gathered research material was converted to the *JPEG* (*Joint Photographic Experts Group*) format in the *ViewNX 2* program from NIKON. The default parameters of the program were used in the processing of the images, apart from *Picture Control Recorded Value [SD] STANDARD* and *Sharpening with the value of 3*. The obtained images were characterised by the same resolution as the *NEF* files.

Next, the digital images of the tomatoes in the *JPEG* format, together with the description of the physical parameters, were subjected to the process of image processing and analysis in the

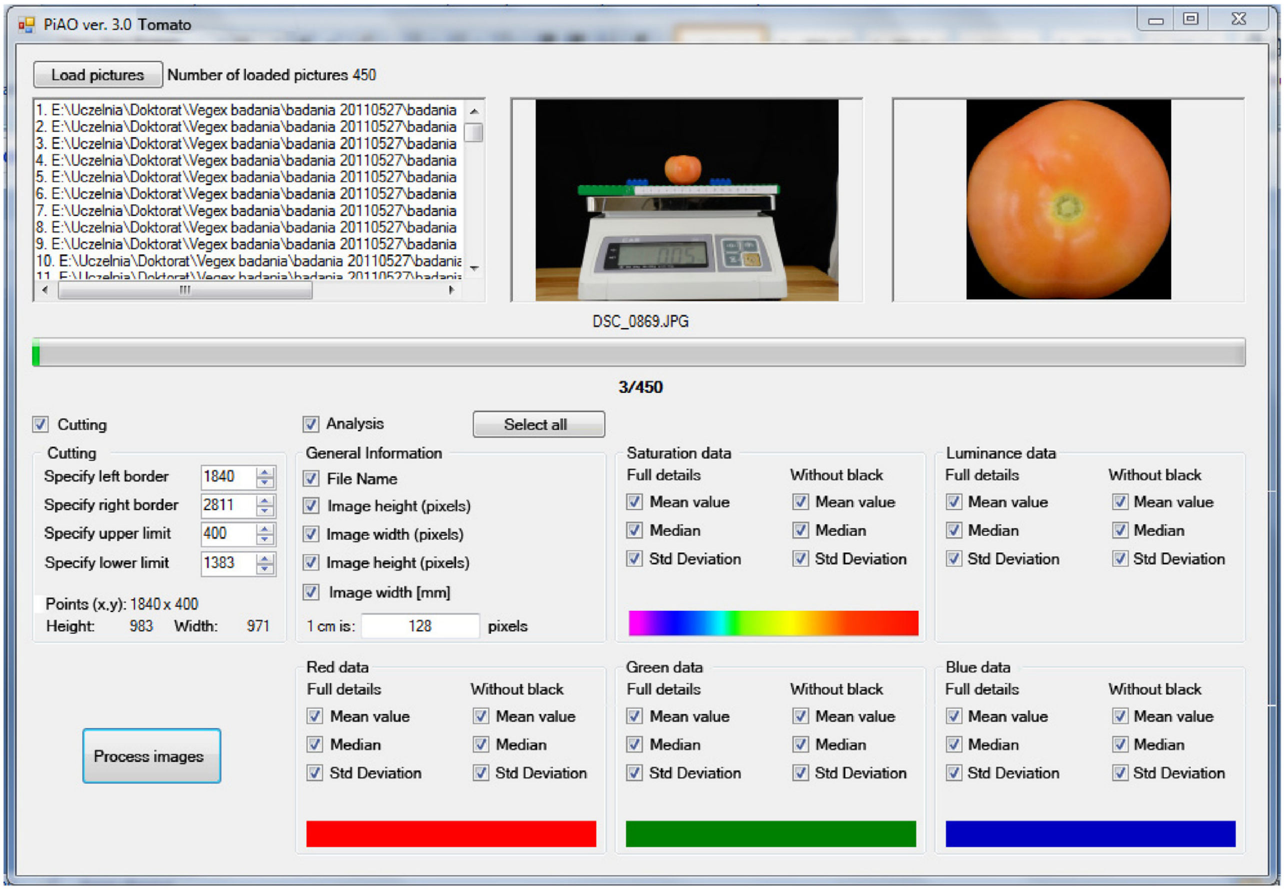

Fig. 4. IT system *PiAO ver. 3.0 Tomato*.

(source: own materials).

author's IT system named *PiAO ver. 3.0 Tomato (Image Processing and Analysis)*] (Fig. 4).

The presented software *PiAO* was prepared in the *C#* programming language on the *Microsoft .Net Framework 4.0* platform with the use of the *AForge.NET* library and the *Microsoft Visual Studio 2010* programming suite. The task of the created software was the transformation of the input and the extraction of the features characteristic of the obtained digital images to the form of input variables of training sets for the *ANN* (*Artificial Neural Network*) simulators.

The IT system processes images and performs calculations based on the standard *RGB* colour palette. Its basic functions include the "*Cutting*" of an object from an image. This function requires the left, right, top and bottom cutting borders to be specified. The next function is "*Analysis*", which facilitates the obtaining of a range of data from an image, such as "*General information*", i.e. "*File name*", "*Image height [pixels]*", "*Image width [pixels]*", "*Image height [mm]*", and "*Image width [mm]*". Determining and saving information on the size of the analysed object in millimetres is possible once the scale is specified, i.e. the number of pixels per 10 mm of length of the real image. The analysed image parameters also include "*Saturation data*", "*Lumination data*", "*Red-colour data*", "*Green-colour data*", and "*Blue-colour data*". It is also possible to obtain information on the value of the average, the median and the standard deviation for these parameters. In addition, the IT system allows the measurement and calculation of values in two modes: "*Full data*" – obtaining properties in the full range of colours of the standard *RGB* palette, and *"Without black"* – an analysis of the same image with the exclusion of the black colour (*RGB*: 0 0 0). The first action to perform after the launch of *PiAO ver. 3.0 Tomato* is to load images that are intended to be processed and/or analysed. Next, the user needs to select which of the functions in the *"Cutting" and "Analysis"* group should be performed. Once the operating parameters of the program are set, the user needs to click the "*Process images*" button. The user will be asked to specify the folder into which to save the processed images and the procedure of the processing and analysis of the loaded photographs will be initiated. Once it is completed, the user is asked to name the file and the directory into which the generated file containing data is to be saved. The file is saved in the *CSV* format, which can be imported directly into the *ANN* simulators.

In the course of the research process it turned out that it was necessary to create an additional IT system which would allow the dividing of the obtained data into two independent sets – stems and fronts. Data generated with the use of *PiAO ver. 3.0 Tomato* was uniform and continuous, and one tested tomato constituted another two rows in a training set. A rule was adopted for taking photographs of the objects requiring that the first photograph always presented the front and the other presented the side of the tomato stem. *System Data processing – Stem and face* allowed the dividing of the originally generated training set into two subsets, representing the stem and the face, and to save them into a *CSV* file.

2.3. Generating neural network model training sets

PiAO ver. 3.0 Tomato made it possible to automatise the process of the generation and conversion of graphic data into a form accepted by the *ANN* simulator. Next, the *Data processing – Stem and face* IT system facilitated dividing the obtained training variables into two groups, stem and face, which became the output sets in the further stages of the generation of training sets.

Directly after the generation, each of the sets describing the stem and the face contained 41 characteristic variables. The first eight parameters in the training set were physical variables obtained in the process of empirical studies. The others were obtained in the process of computer image analysis with the use of *PiAO ver. 3.0 Tomato*.

Eight training sets were created in the course of tests under the following names: *Front*, *Front clean*, *Stem*, *Stem clean*, *Stem and Front average*, *Stem and Front average clean*, *Stem and Front rows*, *Stem and Front rows clean*.

Subsequent analyses determining that reading the size of a tomato directly from a digital image is more accurate than a measurement with the use of a caliper. Therefore, the following variables were removed from the sets: *a* – tomato length, *b* – tomato width, *h* – tomato height.

The seasonal nature of the occurrence of particular sizes and colours of tomatoes was taken into account in the created training sets. In the course of the study there appeared the so-called intermediate colours, e.g. the colours between numbers 2 and 3, 3 and 4 and 4 and 5. For the above-mentioned reasons, the decision was made to formulate their two types when preparing training sets: the first containing the full spectrum of colours including intermediate colours, and the second sets without the intermediate colours (the adjective clean was added to the names of these sets). Each of the sets contained 36 independent input variables.

The *Front* set included 2144 training cases characterising the front of the studied tomatoes. The set contained intermediate colours.

The *Front clean* set contained 1479 training cases of the view of the front of the tomatoes, without intermediate colours.

Stem is a training set containing 1964 training cases with characteristics describing the tomato stem. The training cases include intermediate colours.

Stem clean is a training set of 1419 training cases without intermediate colours describing the tomato stem.

Stem and Front average – stem and front training sets combined into a single joint set. The values of specific variables were averaged and recorded as a single variable, e.g. a single case of a variable of the median of front lumination and the median of stem lumination were averaged and recorded as the lumination median. The created set contains 38 variables and 2144 training cases, including intermediate colours. The set includes the following variables: the height of the tomato on the front side[mm], the width of the tomato on the front side[mm], the height of the tomato on the stem side[mm], the width of the tomato on the stem side[mm].

Stem and Face average clean is a set generated on the basis of the *Stem and Face average*. Like the original set, it contains 38 of the same training variables, and the intermediate colours of tomatoes were eliminated from the data set. A total of 1494 training cases remained.

Stem and Face rows are sets of front and stem combined in rows. This set contains 68 variables and 2144 training cases. The set contains intermediate colours.

Stem and Face rows clean is a set that is similar to *Stem and Face rows*, but training cases containing intermediate colours were eliminated from it. Similarly as in the previous case, the set included 68 variables, but it contained 1494 training cases.

The created training sets assigned to generate neural models were used in the *STATISTICA* software. Each of the prepared sets was divided into three subsets (according to the standard division implemented in the statistical package 2:1:1)

- training subset (U) used for network training,
- validation subset (W), which facilitates the controlling of the results of the operation of the training algorithm in the course of the learning process,
- test subset (T), which allows the performing of the evaluation of the generated neural network.

2.4. The description of neural model training

The process of neural network modelling was conducted in two stages, using an *ANN* simulator implemented in the *STATISTICA* suite. In the first stage, during the preliminary generation of *ANN* models, the *Automatic Network Designer* function was used, and in the second stage *User Network Designer* was employed.

Automatic Network Designer is a heuristic algorithm that experimentally determines the optimal network structure for the current training set. A user cannot interfere in the network structure, except for determining its boundary conditions and general network parameters. The generation of models proceeds automatically.

User Network Designer allows full control over the process of neural network modelling. Apart from determining the type of *ANN*, it is possible to define particular inputs, layers and the number of neurons in the layers, as well as the method of network training and the criteria for the selection of the optimal model.

Automatic Network Designer made it possible to generate neural models which were used as the basis for further research and analyses. It was preliminarily assumed that the simulator should test 20 networks of each type and retain 10 that produce the best results. The criterion for the network retention was the balance between error and network diversity. The aim of this criterion was to obtain a wide spectrum of generated models in order to determine the optimal topology. Additionally, a linear-activation function was established for *MLP* networks.

The simulator tested the following networks

1. Linear,
2. *PNN* (*Probabilistic Neural Network*),
3. *GRNN* (*Generalised Regression Neural Network*),
4. *RBF* (*Radial-Basis Function*),
5. Three-layer perceptron – *MLP* (*Multi-Layer Perceptron*),
6. Four-layer perceptron – *MLP* (*Multi-Layer Perceptron*).

All eight prepared training sets and all variables obtained from the empirical research and the computer image analysis were used in the process of neural network modelling. In the first stage of the modelling, the output variable was colour, determined by the colour of the tomato skin: 2; 2–3; 3; 3–4; 4; 4–5; 5–6. This stage of the modelling resulted in the generation of eight sets of *ANN's* containing 10 neural models each. The generated *ANN's* were analysed and those with the highest quality and the lowest error rate were chosen for conducting another modelling process.

The process of *ANN* training consists of adjusting the neural model parameters to the available training data; thus the quality of *ANN* is understood as the number of correct answers given by the network set against the total number of all answers. A network error is understood as *RMSE* (*Root Mean-Square Error*), i.e., as the square root of the mean-square error. This error is calculated by aggregating the differences between the squares of the errors of particular cases and dividing the sum by the number of considered cases, then extracting the square root of the obtained quotient.

The best results were produced by *RBF* networks, which were consequently chosen for the next stage of the modelling. Similarly to in the first stage, for every training set a new network was generated with one output variable – colour. In this stage, the advanced *User Network Designer* function was used, and also additional models were generated for the sets *Stem and Front average* and *Stem*

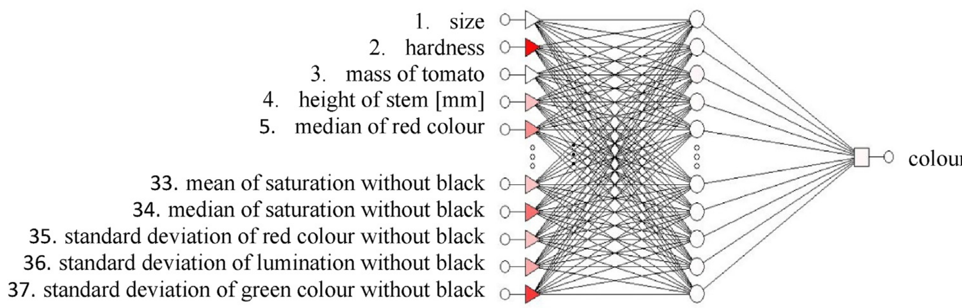

Fig. 5. A screenshot of the generated *RBF 37:37-39-1:1* model.

and Front average clean with two output variables – colour and hardness.

As a result of the conducted research, a set of neural models of the *RBF* type was obtained. The networks were characterised by a different number of input variables: from 22 to 35, and one or two output variables. The generated models contained one hidden layer consisting of 1–78 neurons, depending on the analysed case. The networks were trained with the following algorithms: *EX* (*Explicit* – determining radial deviation), *IS* (Isotropic – determining radial deviation), *KM* (K – determining radial deviation), *KN* (K-nearest neighbour – determining radial deviation), *PI* (Pseudoinverse) and *SS* (*SubSample*).

The first stage of the neural network modelling made it possible to identify those automatically generated *ANN*'s that were the most effective in the process of evaluating the quality of tomatoes. Since the quality of these models was low, i.e., between 50% and 80%, further research was conducted with *User Network Designer*.

3. Results

In the second stage of the research, *User Network Designer* was employed to generate seven neural networks of the *RBF* type with one or two outputs that satisfied the appropriate quality criteria. The best models proved to be two from all the networks that were generated. These were *RBF 37:37 39 1:1* and *RBF 22:22-20 2:2*. They were generated on the basis of the *Stem and Front average* set. The *RBF* models were characterised by high quality and a low *RMSE* rate.

For the *RBF 37:37-39-1:1* network, whose output variable was the colour of the tomato, the training quality was 0.930827, the validation quality was 0.911982, and the test quality was 0.979390. The *RMSE* rate of network training for the training set was 0.075986, for the validation set it was 0.072194, and for the test set 0.061714. The network was trained with the *SS*, *EX* and *PI* algorithms. The discussed result was obtained in the 8th iteration of the process of neural model training. Decreasing the number of variables, as well as performing the 9th iteration of the training process, resulted in the lowering of the quality of this model.

For the *RBF 22:22-20-2:2* network, the variables were colour and hardness. This network is characterised by a training quality of 0.985038, a validation quality of 0.990694 and a test quality of 0.985130. The *RMSE* rate of the network training is 0.065667, of the network validation it is 0.066187 and of the network test 0.073868. In the process of network training, the *SS*, *IS* and *PI* algorithms were used. The presented result was obtained after conducting the model training process three times. As in the previous case, the reduction of the training variables, as well as performing the process for the fourth time, lowered the quality of the network.

In the process of generating *ANN* models, a sensitivity analysis was carried out, which was one of the stages in the neural network modelling. The aim of the analysis is to establish the importance of particular input variables. This process makes it possible to determine which variables are crucial for the training process and the functioning of a neural network model and which are less important. The criterion of the sensitivity analysis was the quotient of the error obtained after running a model without a given variable and the error obtained with all training variables.

The *RBF 37:37-39-1:1* network (Fig. 5) is characterised by 37 input variables and one output variable – colour. The five variables that are the most crucial for the functioning of the neural model are *the size, hardness* and *mass of the tomato*, *the height of the stem [mm]* and *the median of the red colour*. The five least-important variables are *the mean of saturation without black*, *the median of saturation without black*, *the standard deviation of the red colour without black*, *the standard deviation of lumination without black* and *the standard deviation of the green colour without black*.

The *RBF 22:22-20-2:2* model (Fig. 6) was generated on the basis of 22 input variables and has two output variables – colour and hardness. The variable which was the most crucial for the functioning of the *ANN* model was *the mass of the tomato* and the variables specifying the red colour: *median*, *mean without black*, *median without black* and *mean*. The following variables proved to be the least important: *the standard deviation of the blue colour without black*, *the standard deviation of the blue colour*, *the standard deviation of the*

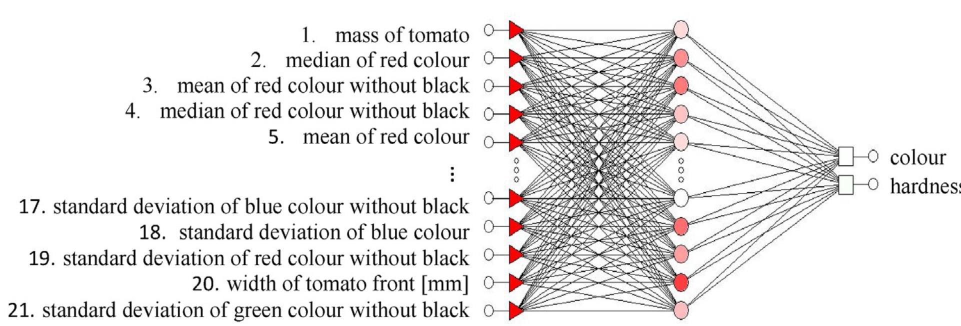

Fig. 6. A screenshot of the generated *RBF 22:22-20-2:2* model.

red colour without black, *the width of the tomato front [mm]*, *standard deviation of the green colour without black*.

RBF neural models are unidirectional networks. They are characterised by three-layer architecture. They have an input layer, a hidden layer and an output layer. The activation function of input and output neurons is linear, while the neurons of the hidden layer are non-linear, i.e., radial. *RBF* networks learn relatively fast in comparison to other models. Since their major advantage is the lack of global data extrapolation, these networks can be considered secure. It needs to be noted that these networks can model any non-linear function using one hidden layer. Linear transformation, carried out in the output layer, is fully optimised using neural modelling techniques, which allows the elimination of local minima found in the process of the training of other models. *RBF* networks make it possible to solve complex non-linear classification problems.

4. Summary

Performing the computer image analysis made it possible to identify the representative features necessary to build training sets for *ANN's*. They were the basis for the generation of neural models used for the independent and effective quality evaluation of greenhouse tomatoes.

The process of neural network modelling was conducted using the *STATISTICA ANN* simulator. *User Network Designer* was used in the process of generating the *ANN's*.

Out of all the generated neural models, the best proved to be the *RBF* network. As a result of the conducted research, we tested and chose the best model serving a classification function. The best network was chosen on the basis of the quality criteria for *ANN's* and *RMSE*.

Satisfactory quality characteristics were obtained in the case of the *RBF 37:37-39 -1:1* and *RBF 22:22-20 -2:2* models. The research showed that performing a correct classification requires taking two digital images of the examined tomatoes, one of the stem and one of the front of the tomato, and generating training sets that contain average values for the extracted characteristics.

The best generated model, i.e., *RBF 22:22-20-2:2*, has an error rate of 1.49%, which means that for every 100 tomatoes subjected to quality evaluation, 98 will be correctly classified. The *RBF* neural model is used in the quality evaluation of one uniform batch of tomatoes intended for sale. Identifying subtle differences in such a batch, which would classify given tomatoes as belonging to another quality category, is very difficult and often impossible to do even by an experienced grader.

A significant advantage of the presented neural model over a qualified grader is the repeatability of the identification process and the objectivity of the evaluation. Obtaining an answer from an *ANN* model requires performing the process of image acquisition and analysis and using the appropriate neural model, which, compared to the work done by a grader, slightly prolongs the process of quality evaluation. However, it is worth noting that the objectivity of the presented method and its ability to eliminate the inaccuracies related to quality evaluation override the inconveniences caused by the prolongation of the process. When compared to a human, a neural model can work without any restrictions, it does not get tired, nor is it prone to external factors. In order for the presented model to work properly, the appropriate and fixed conditions of the image acquisition are required.

The use of the generated neural model allows the non-invasive evaluation and classification of tomatoes, which eliminates the need for using a hardness tester or a caliper that almost always cause damage to the structure of the product.

5. Conclusions

The following conclusions have been drawn from the conducted research

1. The implementation of the methods of computer image analysis and neural network modelling to identify the quality characteristics of greenhouse tomatoes made it possible to create an original method for evaluating and classifying these vegetables.
2. The best classification model that was generated proved to be the *RBF 22:22-20-2:2* type of *ANN*, with 22 input neurons, 20 neurons in the hidden layer and 2 outputs representing the colour and the hardness of the tomato. The generated *RBF* topology is indicative of the non-linear nature of the examined issue.
3. Performing a sensitivity analysis made it possible to determine the input parameters that have a crucial effect on the quality of the functioning of the generated model. These variables are *mass*, *the median of the red colour*, *the mean of the red colour without black*, *the median of the red colour*, *the mean of the red colour*.
4. The elimination of cases representing the intermediate colours of tomatoes (2–3, 3–4, 4–5) from the training sets resulted in the lowering of the quality of the generated neural models.
5. The generated classification neural model *RBF 22:22-20-2:2* allows the fast and effective quality evaluation of greenhouse tomatoes.
6. Generated neural model can be useful as a kernel of a computer expert system to support the work of classifiers in evaluation greenhouse tomatoes.

Acknowledgement

The study was supported by Grant no. PBS3/B8/26/2015 from the National Centre for Research and Development.

References

Boniecki, P., Dach, J., Pilarski, K., Piekarska-Boniecka, H., 2012a. Artificial neural networks in modeling process of emission of ammonia during composting sewage. Atmos. Environ. 57, 49–54.

Boniecki, P., Nowakowski, K., Tomczak, R., Kujawa, S., Piekarska-Boniecka, H., 2012b. The application of the Kohonen neural network in the non-parametric quality-based classification of tomatoes. In: 4th International Conference on Digital Image Processing, Kuala Lumpur (Malaysia), http://dx.doi.org/10.1117/12.954158, proceedings of SPIE, Vol. 8334, Article Number: 833427.

Boniecki, P., Koszela, K., Piekarska-Boniecka, H., Weres, J., Zaborowicz, M., Kujawa, S., Majewski, A., Raba, B., 2015. Neural identification of selected apple pests. Comput. Electron. Agric. 110 (2015), 9–16.

Butz, P., Hofmann, C., Tauscher, B., 2005. Recent developments in noninvasive techniques for fresh fruit and vegetable internal quality analysis. J. Food Sci. 70 (9), 131–141, 2005.

Clement, J., Novas, N., Gazquez, J., Manzano-Agugliaro, F., 2012. High speed intelligent classifier of tomatoes by colour, size and weight. Span. J. Agric. Res. 10 (2), 314–325, 2012.

Du, C., Sun, D., 2006. Learning techniques used in computer vision for food quality evaluation: a review. J. Food Eng. 72, 39–55.

Gastélum-Barrios, A., Bórquez-López, R., Rico-García, E., Toledano-Ayala, M., Soto-Zarazúa, G., 2011. Tomato quality evaluation with image processing: a review. Afr. J. Agric. Res. 6 (14), 3333–3339.

Goel, N., Sehgal, P., 2015. Fuzzy classification of pre-harvest tomatoes for ripeness estimation – an approach based on automatic rule learning using decision tree. Appl. Soft Comput. 36, 45–56, http://dx.doi.org/10.1016/j.asoc.2015.07.009.

Gomes, J., Leta, F., 2012. Applications of computer vision techniques in the agriculture and food industry: a review. Eur. Food Res. Technol. 235, 989–1000, 2012.

Iraji, M., Tosinia, A., 2011. Classification tomatoes on machine vision with fuzzy the Mamdani inference adaptive neuro fuzzy inference system based (Anfis-Sugeno). Aust. J. Basic Appl. Sci. 5 (11), 846–853.

Kujawa, S., Nowakowski, K., Tomczak, R.J., Dach, J., Boniecki, P., Weres, J., Mueller, W., Raba, B., PiechotaT. Carmona, P.C.R., 2014. Neural image analysis for maturity classification of sewage sludge composted with maize straw. Comput. Electron. Agric. 109 (2014), 302–310.

Lu, Q., Wang, J., Gomez, A., Perier, A., 2009. Evaluation of tomato quality during storage by acoustic impulse response. J. Food Process. Preserv. 33, 356–370, 2009.

Naroui Rad, Mohammad Reza, koohkan, Shirali, Fanaei, Hamid Reza, Pahlavan Rad, Mohammad Reza, 2015. Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo L.*). Sci. Hortic. 181, 108–112, http://dx.doi.org/10.1016/j.scienta.2014.10.025, ISSN 0304-4238.

Pabico, J., De Grano, A., Zarsuela, A., 2012. Neural Network Classifiers for Natural Food Products. ICS Research, Philippines.

Patil, J., Kumar, R., 2011. Advances in image processing for detection of plant diseases. J. Adv. Bioinf. Appl. Res. 2 (2), 135–141, ISSN 0976-2604.

Pentos, K., Luczycka, D., Wrobel, R., 2014. The identification of the relationship between chemical and electrical parameters of honeys using artificial neural networks. Comput. Biol. Med. 53, 244–249, http://dx.doi.org/10.1016/j.compbiomed.2014.08.008.

Pentos, K., 2016. The methods of extracting the contribution of variables in artificial neural network models – comparison of inherent instability. Comput. Electron. Agric. 127, 141–146, 2016.

Przybylak, A., Boniecki, P., Koszela, K., Ludwiczak, A., Zaborowicz, M., Lisiak, D., Stanisz, M., Ślósarz, P., 2016. Estimation of intramuscular level of marbling among Whiteheaded Mutton Sheep lambs. J. Food Eng. 168, 199–204, http://dx.doi.org/10.1016/j.jfoodeng.2015.07.035.

Soares, J.D.R., Pasqual, M., Lacerda, W.S., Silva, S.O., Donato, S.L.R., 2013. Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. Sci. Hortic. 155, 24–29, http://dx.doi.org/10.1016/j.scienta.2013.01.026, ISSN 0304-4238.

Vazquez-Cruz, M.A., Jimenez-Garcia, S.N., Luna-Rubio, R., Contreras-Medina, L.M., Vazquez-Barrios, E., Mercado-Silva, E., Torres-Pacheco, I., Guevara-Gonzalez, R.G., 2013. Application of neural networks to estimate carotenoid content during ripening in tomato fruits (Solanum lycopersicum). Sci. Hortic. 162, 165–171, http://dx.doi.org/10.1016/j.scienta.2013.08.023, ISSN 0304-4238.

Wang, S., Zhang, Y., Ji, G., Yang, J., Wu, J., Wei, L., 2015. Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-Scaled chaotic ABC and biogeography-based optimization. Entropy 17, 5711–5728, http://dx.doi.org/10.3390/e17085711.

Wu, X., Yang, J., Wang, S., 2016. Tea category identification based on optimal wavelet entropy and weighted k-nearest neighbors algorithm. Multimed. Tools Appl., 1–15, http://dx.doi.org/10.1007/s11042-016-3931-z.

Zaborowicz, M., Boniecki, P., Koszela, K., Przybył, J., Mazur, R., Kujawa, S., Pilarski, K., 2013. Use of artificial neural networks in the identification and classification of tomatoes. Proc. SPIE 8878, Fifth International Conference on Digital Image Processing (ICDIP 2013), 88782R (July 19, 2013); doi:10.1117/12.2030696.

Zaborowicz, M., Przybył, J., Koszela, K., Boniecki, P., Mueller, W., Raba, B., Lewicki, A., Przybył, K., 2014. Computer image analysis in obtaining characteristics of images greenhouse tomatoes in the process of generating learning sets of artificial neural networks. Proc SPIE. 9159, Sixth International Conference on Digital Image Processing (ICDIP 2014), 91590D (April 16, 2014) ISSN: 0277-786X doi:10.1117/12.2064066.

Zhang, D., Lee, D.J., Tippetts, B.J., Lillywhite, K.D., 2014a. Date maturity and quality evaluation using color distribution analysis and back projection. J. Food Eng. 131, 161–169, http://dx.doi.org/10.1016/j.jfoodeng.2014.02.002.

Zhang, Y.D., Wang, S.H., Ji, G.L., Phillips, P., 2014b. Fruit classification using computer vision and feedforward neural network. J. Food Eng. 143, 167–177.

Zhang, Y., Phillips, P., Wang, S., Ji, G., Yang, J., Wu, J., 2016. Fruit classification by biogeography-based optimization and feedforward neural network. Expert Syst. 33 (3), 239–253, http://dx.doi.org/10.1111/exsy.12146.