

Artificial neural networks and computer image analysis in the evaluation of selected quality parameters of pea seeds

Katarzyna Szwedziak¹

¹Chair of Biosystem Engineering, Opole University of Technology, 76 Prószkowska Street, Opole, Poland

Abstract. The aim of the study was to develop an innovative method of modelling the process of evaluating the quality of agricultural crops on the basis of computer image analysis and artificial neural networks (ANN). It was therefore assumed that on the basis of the prepared application for processing and analysing the acquired digital images, based on the *RGB* colour recognition model, a quick and good method of assessing the quality of products would be obtained. An experiment was conducted on the evaluation of selected parameters of pea seeds quality using computer image analysis and the obtained results were verified by artificial neural networks using the geostatic function.

1 Introduction

Examining the quality characteristics of agricultural and food products is increasingly important because of their suitability for further processing and marketing. After Poland's accession to the European Union, domestic products must meet the requirements of other European countries. The task of the agri-food industry is, apart from processing, also appropriate protection of generally not very durable plant raw materials and their processing into safe and durable food products with preservation of appropriate taste qualities. To achieve such an effect, the agricultural and food industry has a number of technologies and processes in place. Of course in some cases, and using the wrong parameters of these processes, the raw materials may change their quality characteristics, which will affect the final product. Therefore, it is important to continuously monitor the quality of raw materials and food products during storage, warehousing and heat treatment. The necessity of introducing continuous monitoring programmes for the evaluation of the quality of food and agricultural products and raw materials also results from the increasing consumer demands. Raw materials must also be properly classified and appropriate quality control methods must be selected. For example, other quality characteristics should be considered for vegetables and other for fruit or raw materials that are only suitable for further processing, e.g. cereal or legume grains. Factors connected with the quality of raw materials are primarily the product's specific characteristics, such as the physical and chemical properties of the products, colour, shape, variety, content of basic chemical

components, degree of damage or pest infestation. It therefore seems appropriate to develop an innovative method for identifying all these characteristics in an easy, accessible and rapid manner. Undoubtedly, modern techniques include neural networks, i.e. artificial intelligence and computer image analysis. [8,10]

2 Purpose and scope of work

The aim of the study was to develop an innovative method of modelling the process of evaluating the quality of agricultural crops on the basis of computer image analysis and artificial neural networks (ANN).

It was therefore assumed that on the basis of the prepared application for processing and analysing the acquired digital images, based on the *RGB* colour recognition model, a quick and good method of assessing the quality of products would be obtained. [2,3]

To this end, the works were broken down as follows:

- preparation of pea seed samples,
- preparation and development of a computer application for the assessment of the degree of grain mass contamination based on the *RGB* model,
- creation of a measuring stand, which enables to take digital photographs of appropriate quality,
- taking digital photographs of the samples of the pea seeds taken,
- analysing the photographs of pea seed samples obtained and converted to .bmp format by means of the APR computer application and further converting them to binary form,
- tabular summary of the binarisation results ("1"; "0"),
- neural analysis,
- geostatic.

To identify and solve the problems mentioned above, an innovative method of computer image analysis was developed, and artificial neural networks were used, which allowed for quick and precise evaluation. To this end, a method for taking samples for testing and taking digital photos in an appropriate way was developed to obtain the desired information.

3 Test Methodology

Several series of studies were carried out during the experiment. Pea seeds were used for the study. In the first stage of the research, an experiment was carried out in the laboratory, preparing samples of pea seeds by artificially contaminating them. In this way, samples with a known mass of contaminants were obtained. Photographs were taken with a digital camera and analysed using the "APR" computer application. The performed analysis concerned the contaminants in the mass of pea seeds. The second stage of the research was based on neural modelling of the obtained results. To determine the purity of seed mass and tests performed in the laboratory, a test stand for computer image acquisition was made. This stand is equipped with appropriate light sources. It consists of the following elements:

- webcam, digital camera,
- computer for collecting and processing data,
- illuminated table on which the test samples are placed,
- software for viewing camera images and taking pictures (Fig. 1).



Fig.1 General view of the basic module of the stand for computer image acquisition. K. Szwedziak

A very important element in the process of image acquisition was the selection of appropriate lighting and the position of the acquisition equipment. [4,5]. This stand ensures a smooth, in terms of intensity, inflow of incident light, both from the side of the camera and from the side opposite to the position of the camera, in relation to the material tested. The illuminated table was an important element in achieving high contrast between the object and its backgrounds. Strong, omnidirectional lighting from the camera side ensures that any shadows are removed. It is important that the lighting intensity is selected taking into account the sensitivity range of the image acquisition equipment sensors. In particular, it was necessary to examine the histogram of pixel brightness levels for maximum use of the range (Fig. 2).[9]

In order to maintain the repeatability of the measurements, the side walls were made of light-scattering material. The main source of lighting was internal lamps. It is important that the outdoor lighting is not too intense and does not fall on the side walls of the stand locally.

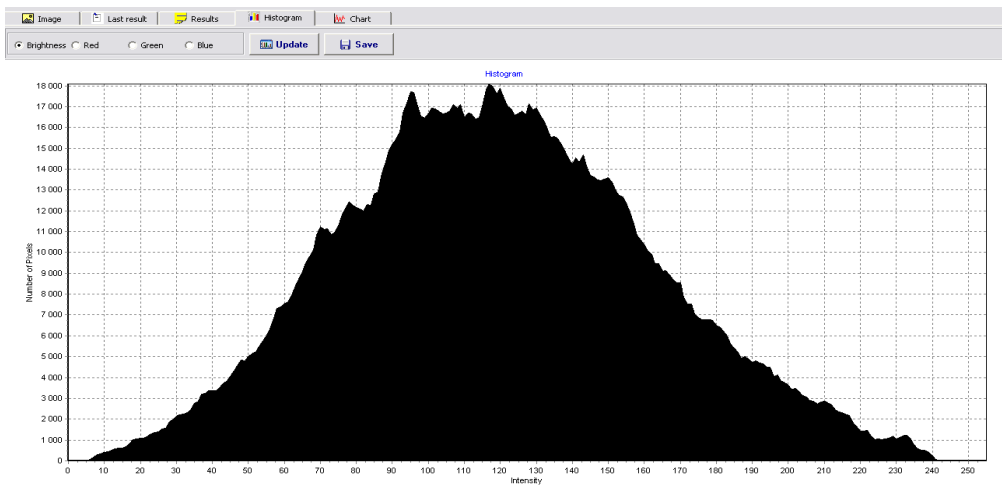


Fig. 2 Example histogram of pixel brightness levels obtained on the basis of a digital photo analysed by the "APR" application. K. Szwedziak

APR (Analyses Processing Recognition) is an application for analysing, processing and recognising images. Its basic feature is the ability to build image processing scripts. For this purpose, a scripting language has been built in to allow a number of graphical operations. In

addition, it is possible to enter commands directly in the command line. Part of the operations is available from a panel containing an appropriate communication interface with the user. [9]

51 trials each were prepared. Seed samples were contaminated by each time increasing the amount of contamination for each trial by 1 g. Samples weighing 50 g were prepared for testing.

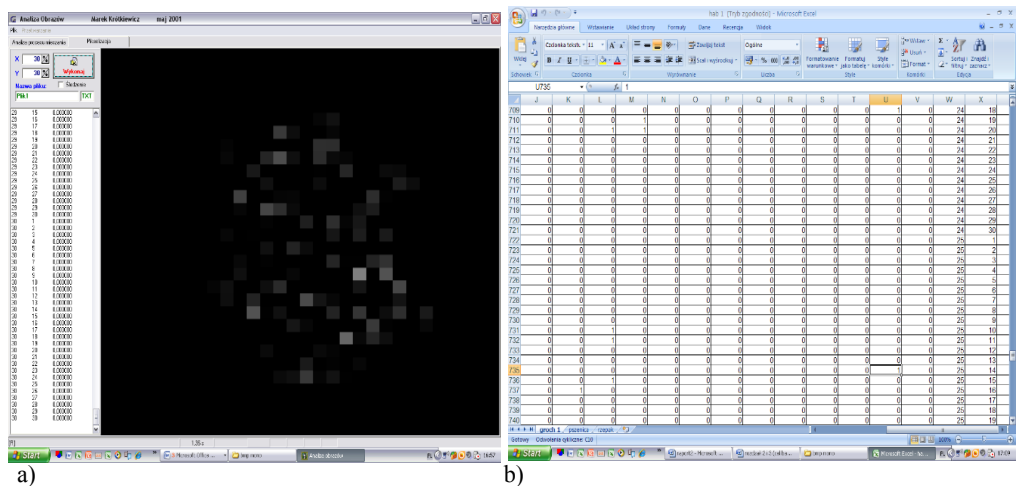
At the same time, the first trial contains 100% pea seeds and the last 51th sample contains 100% contaminants. These two cases were treated as control for the examined variant.

Seed samples prepared in such a way were mixed in a static mixer to obtain the mass of mixed pea seeds with contaminants. Rapeseed was used to contaminate pea seeds. The next stage of the experiment was the preparation of samples and image acquisition stand, so as to obtain images of appropriate quality with relevant information visible to the computer application. A digital camera was used to take photographs; the photographs were then framed, i.e. the area subject to analysis outlined, and saved in .bmp format. They were converted into black and white images using a computer graphics software (Fig. 3).



a) b)
Fig. 3 Example of pea sample a) full-colour; b) black and white. K. Szwedziak

Then they were digitized for an arbitrarily defined resolution of the research window (30x30=900 pixels). (Fig. 4a)



a) b)
Fig. 4(a). Picture after pixilation: 30x30 cells b). Summary of results obtained after pixilation. K. Szwedziak

On the basis of the obtained grayscale degree of individual pixels of the window, they were binarised into contaminants and "background" (in the binary record they were marked as "1" and "0" respectively). The results of binarisation of individual photographic images are summarised in tables for the next stage of calculation using neural network (Fig. 4b)

Neural analysis was performed in a package [Neal R. (2000). Flexible Bayesian Models on Neural Networks, Gaussian Processes, and Mixtures v. 2000-08-13. University of Toronto]. The adopted regression model assumes a properly selected neural network architecture. In order to determine the most frequently "represented" locations of the tested contaminants in particular groups, the coordinates of points (dependent variables - *targets*) were contrasted with binary variables (independent - *inputs*). The number of independent variables in each group for each contaminant type was determined by the number of binary images accepted for this analysis - i.e. with a point fraction not exceeding 10 % of the total set of points in the test window - i.e. approximately 90 points of contaminants.[1]. The arbitrary level of acceptance of images was justified by the fear of possible blurring of contaminants with the background in case of overly darkened photographs (the most "clear" photographs were analysed). The number of hidden layers was set at a level enabling the most effective network learning, controlled by the so-called rejection rate, hyperparameters and their plots. A numerical simulation was performed for 100 iteration steps after the rejection of the first 20% of the so-called burn-in steps.

4 Analysis and discussion of results

In order to assess the conformity of a given point distribution, the non-parametric test of Kolmogorov - Smirnov was used, by means of which the observed spatial locations of points in the research window were estimated and compared with the expected values. Test statistics in the Kolmogorov - Smirnov test is as follows:

$$D_{n1,n2} = \max[F_{n1}(x) - F_{n2}(x)] \quad (1)$$

where: $F_{n1}(x)$ and $F_{n2}(x)$ mean the respective empirical distribution functions for the first and second sample, $n1$ and $n2$ mean the respective sample sizes. The method of distribution of contaminants in the mass of seeds was described using one of the geostatic functions – the K function.

The analysis of spatial distribution of observed particles concerns a set of points irregularly distributed in a limited area of the surface. An attempt was made to adapt the geostatistical function K to the analysis of the distribution of contaminants in the mass of seeds. The properties of the spatial distribution of points describing the relationships between them can be easily determined using the K function. The K function is defined as follows:

$$K(d) = \lambda - 1E [\text{number of points} \leq \text{distance } d \text{ from arbitrarily defined distance}] \quad (2)$$

where:

λ - intensity (number of points per unit area),

E - expected value (number of points \leq distance d from arbitrarily defined distance)

The estimator for the above K function was given by Ripley [5]. The K function has the advantage that its theoretical value is known for several useful spatial distribution models. An example of K function for a homogeneous process without spatial dependencies between points in space is the Πd^2 function plot. In case the points are clustering in a certain area, we will expect a predominance of events for short distances, so the function $K(d)$ will be greater than the Πd^2 ($K(d) > \Pi d^2$) determined for a homogeneous process. If the function $K(d)$ is less than Πd^2 ($K(d) < \Pi d^2$), then the points are distributed regularly. During modelling, network learning parameters were controlled to obtain "rejection rate"

values for initiations between 0.1 and 0.3 and for production runs between 0.2 and 0.8. The element is one of the network parameters indicating the correct modelling of data. These coefficients were 0.2 and 0.4 respectively. An additional element determining the properly assumed values of the network algorithm was the observation of the trend of learning hyperparameters plots [6,7]. Fig. 5-6 graphically presents geostatic data for pea seeds.

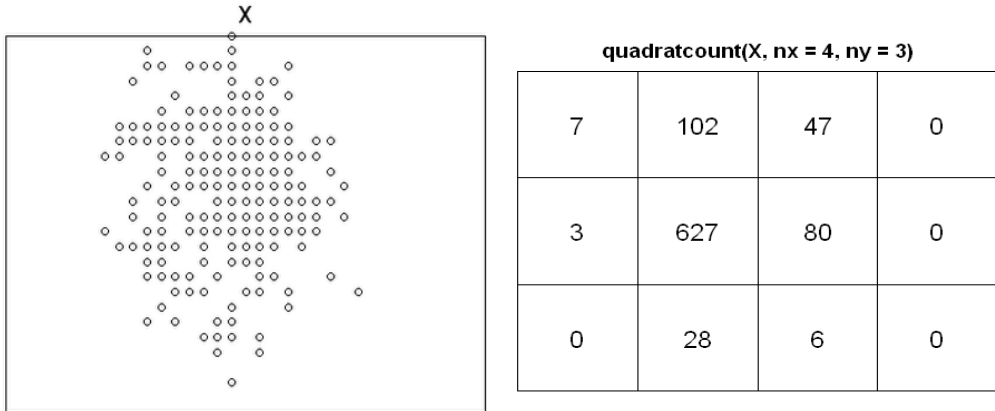


Fig. 5 Distribution and square of counting the distribution of contamination points in the mass of pea seeds. K. Szwedziak

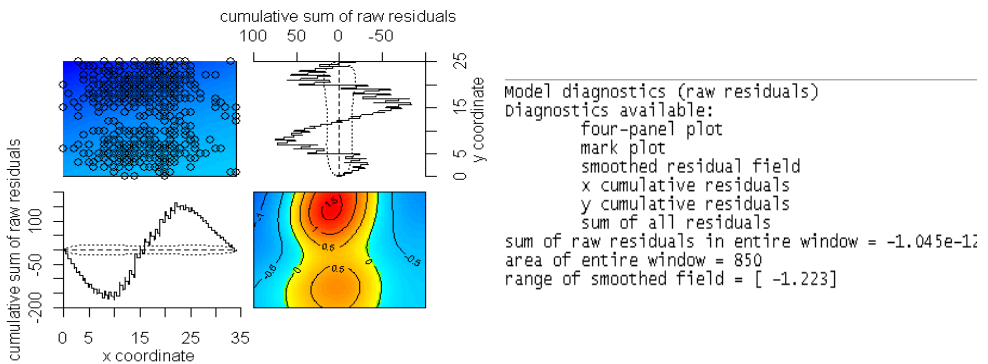


Fig. 7. Model matching diagram - cumulative sum of raw residuals K. Szwedziak

Based on the obtained graphs and geostatic parameters it can be said that the distribution of points in the research window is uneven. The value of $X\text{-square} = 46.06$ and $p < 2.2e^{-16}$ was obtained for the distribution of contaminants in the mass of pea seeds. There are deviations from the independent arrangement of points, centralization of the point process - there is a geostatic interaction. There are no grounds for rejecting H_0 , which stands for no geostatic interaction. On the basis of the K function plot, a strong tendency for clustering of points can be observed. On the basis of a plot of hyperparameters of network learning we can say that the model is a fitting model. The model fitting is also confirmed by plot 8, which shows that the range of smoothing the function (residuals plot) is at the level of $|1,223|$

5 Conclusions

Based on the analysis of the image, neural prediction and geostatistical analysis it can be said that:

1. Logically and experimentally developed and tested neural models for the evaluation of pea contamination distribution confirmed that it is advisable to use them on the basis of colour characteristics obtained from the created image analysis software.
2. The use of computer image analysis allowed for a significant acceleration of the assessment of the distribution of contaminants in the examined material compared to traditional methods.
3. The "APR" computer application based on the RGB model recognises colours occurring in the mass of seeds and on the fruit and seed coat, which was confirmed in laboratory tests and model tests for pea seeds.
4. The application of K function allowed for observation and statistical identification of regularity of contamination distribution in pea seeds mass.
5. The combination of computer image analysis, neural analysis and geostatic analysis is a cost effective and fast way to perform the analysis of the locality of contaminants in the mass of pea seeds.

Bibliography

1. B. A. Adewumi, O. C. Ademosun, A. S. Ogunlowo *Agricultural Engineering International*, **VIII**:1-12 (2006).
2. I. Białobrzewski, *Inżynieria Rolnicza* **8** 68, 7-14, (2005).
3. P. Boniecki, A. Paryś, *Jurnal of Research and Application in Agricultural Engineering* **51** 1, (2006).
4. P. M. Granitto, P. F. Verdes, H. A. Ceccatto, *Computers and Electronics in Agriculture* **33**, 91-103, (2000).
5. B. D. Ripley, *Journal of Applied Probability*, **13**, 255-266, (1976).
6. R. Tadeusiewicz, *Wprowadzenie do sieci neuronowych*. StatSoft Polska, Kraków, (2001)
7. R. Tadeusiewicz, P. Kohorda, *Komputerowa analiza i przetwarzanie obrazów*, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków, (1997).
8. J. Trajer, *Inżynieria Rolnicza* **2** 62, 173-180, (2005).
9. K. Szwedziak, M. Krótkiewicz, *Inżynieria Rolnicza*, **13** 88, 429–435, (2006).
10. K. Szwedziak, K. Wojtkiewicz, *Inżynieria Rolnicza*, **2** 154, 109-118, (2015).