Estimating bulk density in leguminous grains with different traits using color parameters from digital images combined with artificial neural networks

Bruna Gava Floriam [iD], Fabíola Manhas Verbi Pereira [iD], Érica Regina Filletti+ [iD]

São Paulo State University (Unesp), Institute of Chemistry, 55 Prof. Francisco Degni St, Araraquara, São Paulo, Brazil

+Corresponding author: Érica Regina Filletti, Phone: +55 16 3301-9839 email address: erica.filletti@unesp.br

ABSTRACT: Dry grains from leguminous species, such as soybeans (*Glycine max* L.), common beans (*Phaseolus vulgaris* L.), chickpeas (*Cicer arietinum* L.) and corn (*Zea mays* L.), are regularly consumed for human nutrition. This paper showed the possibility of estimating bulk density as quality parameter of 4 different dry grains (soybeans, common beans, chickpeas and corn) in a same model using the average values of color descriptors from digital images combined with an artificial neural network, with low computational costs. These food products are good sources of carbohydrates, protein and dietary fiber, and they possess significant amounts of vitamins and minerals and a high energetic value. Estimation of the physicochemical properties of grains is challenging due to variations in shape, texture, and size and because the grain colors appear similar to the naked eye. In this work, an analytical method was developed based on digital images converted into ten color scale descriptors combined with a neural model to provide an accurate parameter for grain quality control with a low computational cost. The bulk densities of four type of grains, i.e., soybeans, beans, chickpeas and corn, were predicted using numerical data represented by the average values of color histograms of a ten color scale (red - R, green - G, blue - B, hue - H, saturation - S, value - V, relative RGB and luminosity - L) from digital images combined with artificial neural networks (ANNs). The reference bulk densities were empirically measured. A very good correlation between the reference values and values predicted by the ANN was achieved, and with a single ANN developed for the four grains, a correlation coefficient of 0.98 was observed for the test set. Moreover, the relative errors were between 0.01 and 5.6% for the test set.
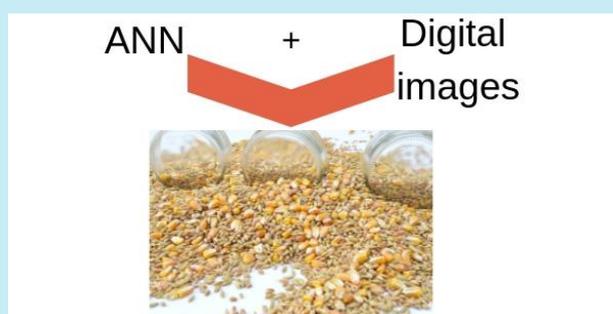
1. Introduction

Soybeans (*Glycine max* L.), common beans (*Phaseolus vulgaris* L.), chickpeas (*Cicer arietinum* L.) and corn (*Zea mays* L.) are a part of most human diets, regardless of culture. These grains are leguminous species consumed as dry grains. They are remarkable sources of carbohydrates, protein and dietary fiber, and they possess significant amounts of vitamins and minerals and a high energetic value[1].

The quality of food grains is dependent on several physicochemical parameters, including the bulk density (or density in each mass). In a grain, the bulk density is more closely related to its shape than its size[2]. Stored food materials can suffer from variations in bulk density according to the bin depth[3]. Thus, a model that can monitor this parameter would be helpful to avoid losses in agri-food supply chains. For instance, in the study by Bart-Plange and Baryeh[4] several laborious physicochemical methods were applied to evaluate cocoa beans, as the raw material for manufacturing chocolate and other food products, including the determination of their bulk density.

Using the method presented here, the quality of four types of grains can be estimated using an accurate approach based on a relevant

physicochemical parameter, i.e., the bulk density, which is related to the storage system, type of container and characteristics of the grains.

Color data from digital images are considered reliable sources of analytical information for many purposes, independent of the type of device, for example, scanners, cell phones or cameras[5]. The combination of digital images and an artificial neural network (ANN) can be adapted for applications such as the shape analysis of grains[6] and for variety identification[7]. However, in both cited studies, many complex steps were necessary to develop a predictive response model with good accuracy.

The advantage of this study is that additional information from a color histogram of a ten color scale[8,9] can be determined using a simple computational routine with fast calculations. In addition, by using the average color values instead of the entire color histogram, which includes 2560 colors[10,11], the speed of the ANN calculations is improved, and accurate results are achieved. Therefore, ANNs are useful tools for this research because they require less computational effort than other numerical techniques.

ANNs are computational models consisting of simple processing units called neurons, which are inspired by the central nervous system of intelligent organisms that acquire knowledge through experimentation; ANNs can perform machine learning to predict parameters and recognize patterns[12,13].

Initially, an ANN undergoes a learning phase in which some examples are presented to it during training, and it automatically extracts the necessary characteristics to represent the information learned by adjusting the synaptic weights of the neurons through an adequate learning algorithm. Then, these characteristics are used to generate answers to the problem studied. In other words, by providing input data to an ANN and reporting the desired output (response), the ANN can provide coherent results for new input data that are different from those used in training.

There are several advantages to using ANNs; for instance, ANNs are easy to use and update, they have data error tolerance because of the ability to respond in an acceptable way, even if partially damaged, they have great freedom in the adjustment of synaptic weights of neurons due to the presence of bias, which is a special processing unit that allows better adaptation on the part of the

neural network to the knowledge provided to it, and they provide a precise response at high speeds[14-16]. The main advantage of ANNs is their ability to generalize or learn from examples[17]; that is, ANNs can generalize learned information to provide satisfactory results for cases not seen in training. Therefore, ANNs have been used in many fields, such as chemistry[18], geology[19], medicine[20], neurocomputations[21] and biomedical engineering[22], among others.

In this sense, after an ANN has been trained and tested, it can predict the output (desired response) of new input data in the domain covered by the training examples. For food analysis, data from digital images acquired using both a camera and a desktop scanner have been applied to predict the fermentation index of cocoa beans by ANN modeling[23]. An ANN was also combined with digital images and showed excellent potential for wheat varietal identification[24] using the morphometric characteristics of these grains for the classification of different varieties with 88% accuracy and individual varieties with 84% and 94% accuracy.

Our study shows a feasible analytical method based on digital images converted into ten color scale descriptors combined with an ANN to estimate the bulk density of leguminous grains such as soybeans, beans, chickpeas and corn.

2. Materials and methods

2.1. Procurement of samples and instruments

Grains of soybeans, beans, chickpeas and corn were purchased locally. The grains were sorted to eliminate external material and damaged grains. For the tests, the grains were separated in small packets made of transparent plastic bags (10×6 cm), as shown in Fig. 1. According to the size of the grains, each plastic bag was filled with 100 soybeans, 100 beans, 50 chickpeas, or 100 kernels of corn, resulting in 56 packets of soybeans, 35 packets of beans, 30 packets of chickpeas and 50 packets of corn.

The content of each packet was weighed on an analytical balance with ±0.0001 g precision (FA-2104N, EQUIPAR, Curitiba, PR, Brazil), and the apparent volume was measured with graduated cylinders with a volume of (50.0±0.5) mL for soybeans, corn and beans and with a volume of (100±1) mL for chickpeas. Using the measured masses (m) and apparent volumes (V), the bulk

densities of the samples were estimated using the ratio $\rho = \dfrac{m}{V}$, and these values were considered the reference values in this study.
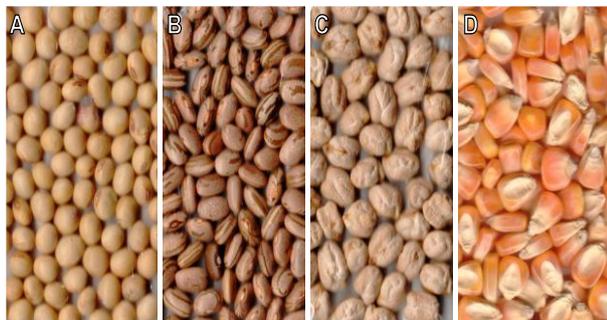

Figure 1. Examples of the scanned images of soybean (A), bean (B), chickpea (C) and corn (D) grains.

Then, each packet was digitalized using a conventional scanner (HP, LaserJet Pro 200 Color MFP M276nw, Brazil). The final size of the

images was 550×1000 pixels with a resolution of 96 dots per inch (dpi).

2.2. Image treatment

The digital images were processed and converted into the average values of ten color scale descriptors, red (R), green (G), blue (B), hue (H), saturation (S), value (V), relative red (Rr), relative green (Rg), relative blue (Rb) and luminosity (L), using the Matlab R2015b (The MathWorks, Natick, MA, USA) code available in the supplementary material of Camargo, Santos and Pereira[8]. Figure 2 shows an image of the data. Using this code, the average value of each histogram color, which includes 256 colors, is obtained; thus, each color is represented by one value per color histogram.
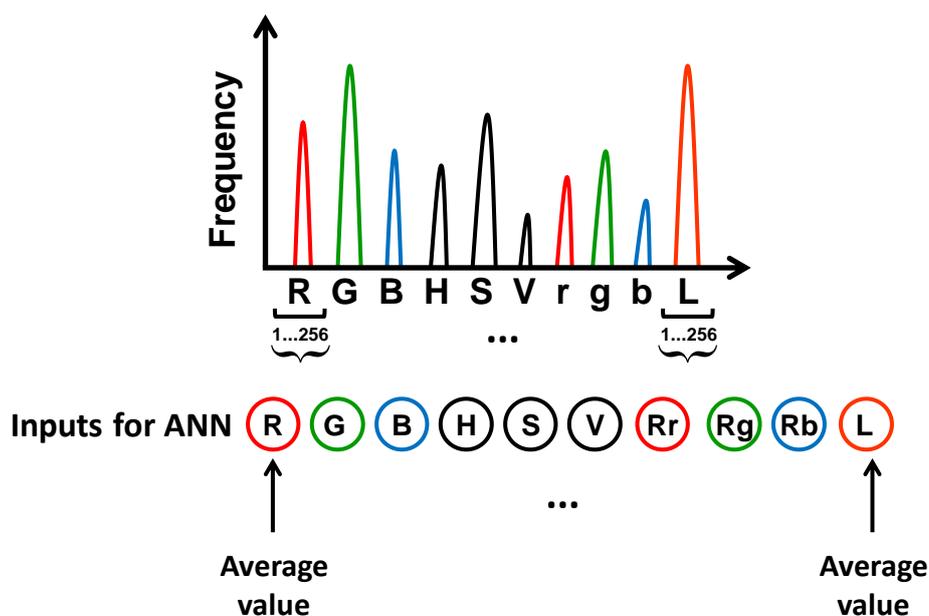
Figure 2. Diagram of the input data for the ANN calculations

2.3. ANN description

A feedforward neural network[25] was implemented in Matlab R2015b. Two learning algorithms were tested in the development of the ANN: error backpropagation and the Levenberg-Marquardt algorithm. The Levenberg-Marquardt algorithm presented correlation coefficients in the mean of 0.88 and RMSE (Root Mean Square Error) equal to 0.031 for the test set, while the backpropagation algorithm provided results with

correlation coefficient equal 0.98 and RMSE 0.014. Therefore, the backpropagation algorithm provided the best results and is described briefly below[12,13]:

1) Set the initial parameters of the network $w_{i,j}$ and $b_{i,j}$ (weights and bias) as random numbers.

2) From a training data set with pre-assigned input/output pairs, take the k-th (p_i^k, a_i^k) pair, calculate the outputs of the network with the same input, and form the new pair (p_i^k, t_i^k).

3) Calculate the error between the desired a_i^k and the obtained t_i^k output values through the formula

$$e = \sqrt{\Sigma_i \left(a_i^k - t_i^k\right)^2} \qquad (1)$$

4) Calculate the partial derivatives of error e with respect to the weights and bias.

5) Change the weights and bias according to the steepest descent strategy and a specified learning rate α:

$$b_{i,j} \leftarrow b_{i,j} - \alpha \frac{\partial e}{\partial b_{i,j}} \qquad (2)$$

and

$$w_{i,j} \leftarrow w_{i,j} - \alpha \frac{\partial e}{\partial w_{i,j}} \qquad (3)$$

6) Iterate steps 2 to 5 by successively modifying $b_{i,j}$ and $w_{i,j}$ until a defined number of learning cycles or a stopping criterion is reached.

The parameter learning rate, number of neurons per layer, number of layers, activation functions and number of epochs were varied by trial and error to obtain the best result, and early stopping of the training was performed to avoid overtraining.

3. Results and discussion

The ranges of bulk density measured using the reference method were in g cm^{-3} for soybean between 0.61 and 0.65, bean (0.75-0.78), chickpea (0.63-0.68) and corn (0.68-0.85). The variation of values among the colors converted from the images were: (i) soybean grains: R (174–192), G (128–135), B (83–91), H (0.079–0.081), S (0.52–0.57), V (0.68–071), r (0.46–0.48), g (0.32–0.33), b (0.19–0.21), and L (386–407); (ii) bean grains: R (140–147), G (91–99), B (63–72), H (0.066–0.072), S (0.56–0.61), V (0.55–0.58), r (0.51–0.53), g (0.29–0.30), b (0.18–0.20), and L (294–317); chickpea grains: R (167–174), G (131–137), B (103–111), H (0.070–0.074), S (0.39– 0.42), V (0.66–0.68), r (0.43–0.44), g (0.318–0.321), b (0.24–0.25), and L (401–421) and (iv) corn grains: R (196–217), G (122–145), B (75–100), H (0.06–0.07), S (0.55–0.64), V (0.77–0.85), r (0.49–0.53), g (0.30–0.31), b (0.17–0.20), and L (397–461). In all of cases, differences among the samples were revealed that justify an ANN model using the 10 colors as input layer. Tests were also performed without variables H and L, for example, showed worse than all input variables were used. For this

case, the correlación coefficients were 0.94, 0.84 and 0.92 for training, validation and test sets, respectively. Therefore, the input variables of the ANN training set were the R, G, B, H, S, V, Rr, Rg, Rb and L color descriptors, for a total of 10 input variables for each packet of grain sample; that is, the developed ANN had 10 neurons in the input layer (Fig. 2). The strategy of averaged value of each histogram color was applied because the entire histogram for the 10 colors corresponds to 2560 variables. For the ANN, 2560 values imply 2560 neuron in input layer, then the calculations would be slower than that computed using the averaged color values, mainly considering that the information would not be improved. The information from the hue (H) and saturation (S) color scales was also important as revealed by our previous evaluation[8].

The ANN outputs were the reference bulk density values. The samples were randomly divided into three sets as follows: a total of 70% of the samples were used in the ANN training set, 15% was used for the validation of ANN, and the remaining 15% was used to test the generalizability of the ANN. The training set consisted of 119 samples, i.e., 38 soybean, 19 bean, 24 chickpea and 38 corn samples, the validation set had 26 samples, i.e., 8 soybean, 10 bean, 2 chickpea and 6 corn samples, and the test set had 26 samples, i.e., 10 soybean, 6 bean, 4 chickpea and 6 corn samples. The number of samples were related to the size of the grains and the size of the packet (plastic bag) was the same. The main goal was to analyze the entire packet then the variation in size of the grain had determined the number of packets, which did not affect ANN results, since it was not necessary for ANN to know which grain sample was being analyzed to estimate grain density in each packet.

A single ANN was developed for all four types of grains (soybeans, beans, chickpeas and corn). Several ANN architectures were tested with one intermediate layer (varying the number of neurons from 10 to 20) and with two intermediate layers (with 8 and 4 neurons in each layer, respectively). The ANN that obtained the best result had only one intermediate layer with 15 neurons, and its performance was below 10^{-3} for the mean square error obtained after 1500 epochs, as shown in Fig. 3. The training was interrupted in 1500 epochs to avoid excessive adjustment by ANN, which could lead to an overfitting. The activation functions were a hyperbolic tangent function in

the intermediate layer and a linear function in the output layer.
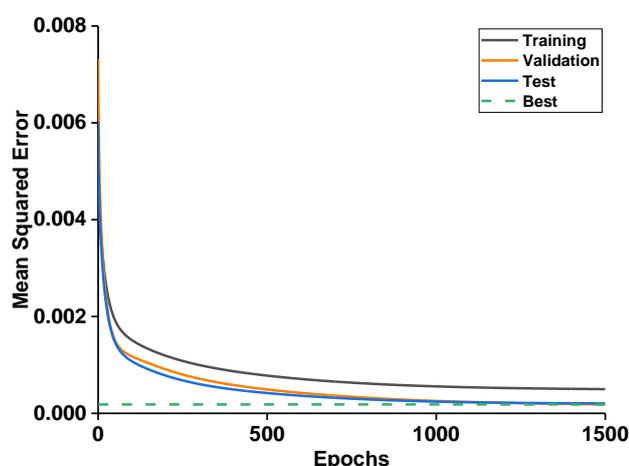
Figure 3. ANN performance results for the training, validation and test sets. The green dotted line denoted as "best" shows the lowest value of the mean square error reached in the ANN validation in 1500 epochs.

The highest efficiency for the four types of grains together had a learning rate of 0.8. Training was also performed with higher and lower learning rates, but the network became unstable with learning rates below 0.3, meaning that it failed to provide good results for the test samples. At a learning rate of 0.8, the ANN was stable, and the training was very fast (close to 1 min). The training error was small, approximately 10^{-3}, and it was stable throughout the learning process.

Among the four types of grains used in this work, the corn grains are the least uniform, that is, they present the most variation in density (between 0.68 e 0.85 g cm^{-3}), so the estimated values were more distributed along the linear adjustment, as shown in the Fig. 4. And the values of bean grains density (0.75-0.78 g cm^{-3}) are within the range of corn grains density, which agrees with the results obtained for these two grains. On the other hand, soybean and chickpea grains are more uniform with little variation in density.

The plots in Fig. 4 show a high correlation of 0.94 between the reference bulk density values and the values predicted by the ANN model for both the training sets (Fig. 4A) and a correlation coefficient of 0.98 for the data from both validation and test sets (Fig. 4B and 4C). The best fit equations for the training, validation and test data are

$y = 0.86x + 0.10$, $y = 0.92x + 0.057$, and

$y = 0.90x + 0.069$, respectively.

The mean relative error was 2.1, 1.5 and 1.4% for the training, validation and test data, respectively, with the lowest and highest relative errors of the ANN responses for the test set samples between 0.01 and 5.6%, which is a noteworthy result.
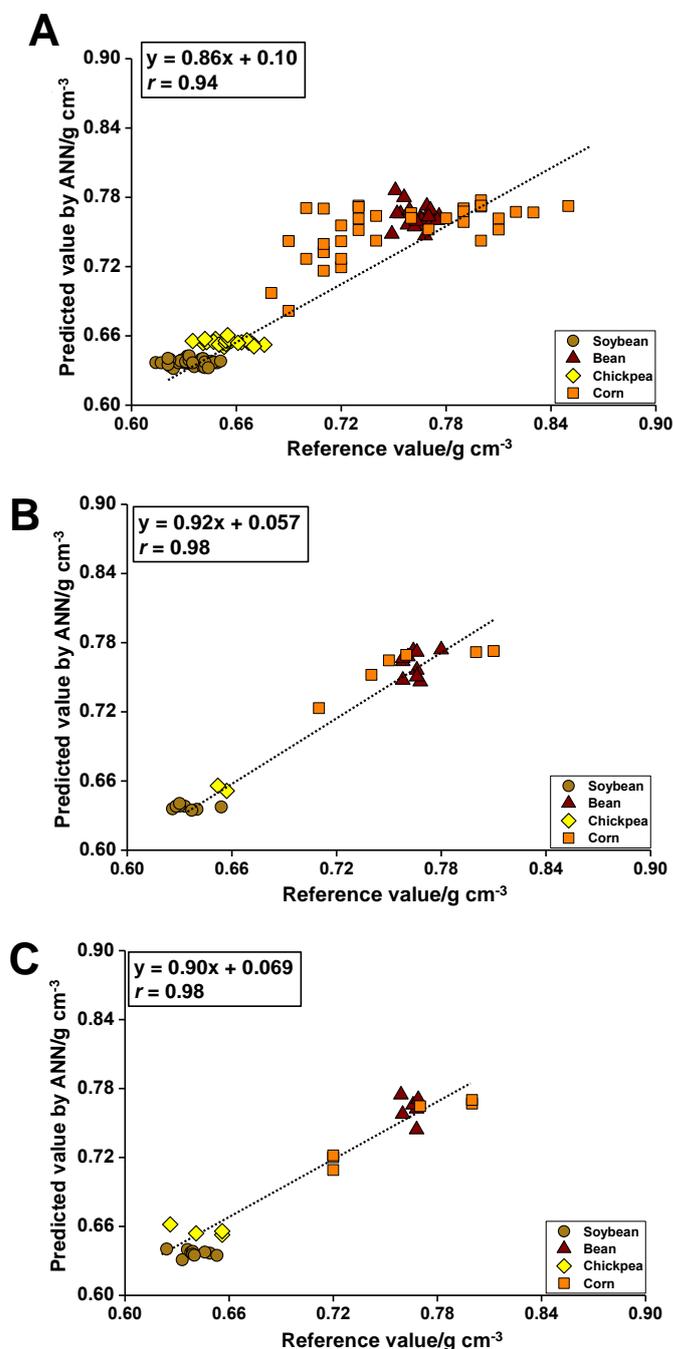
Figure 4. Predicted bulk densities (g cm^{-3}) obtained by the ANN model using image color data of the grains (soybean, bean, chickpea and corn) for the training (A), validation (B) and test (C) sets.

The accuracy of the results is shown by the plots in Fig. 5, giving the equations $y = 0.90x + 0.10$ and $y = 0.90x + 0.036$, which envelop the results obtained by the ANN calculations. These equations were based on calculations involving the linear regression of the test set (Fig. 4C), replacing the term x by each reference value and summing the linear coefficient to 3 times the standard deviation obtained for the absolute errors (difference between the reference values and those predicted by the ANN) of the 26 samples. Thus, for a given bulk density ρ, the response of the neural network will be in the interval $[\rho - 0.03, \rho + 0.03]$, where $[\rho - 3\sigma_r, \rho + 3\sigma_r]$ and $\sigma_r = 0.01$; a range with a very small amplitude indicates high reliability in the ANN response.
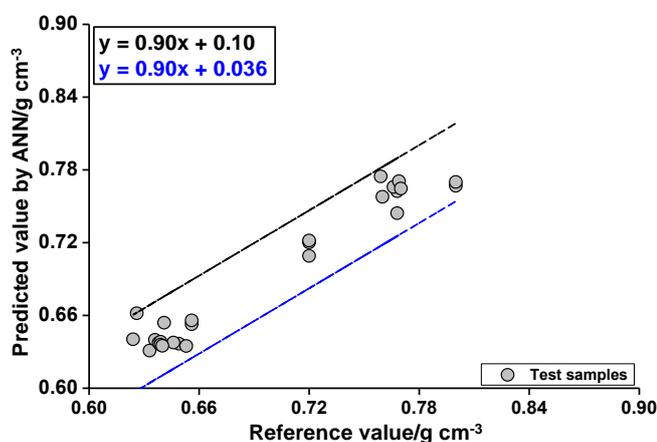
Figure 5. Bulk densities (g cm^{-3}) of soybean, bean, chickpea and corn provided by the ANN for the test set samples.

4. Conclusions

The bulk densities of four different dry grains were accurately estimated using the average values of color descriptors from digital images combined with an ANN. This method is promising, considering that these grain foods have similar colors and different shapes and textures. The reference values and the values predicted by the ANN model were highly correlated ($r = 0.98$) for the test set, with low relative errors between 0.01 and 5.6%. Thus, our study showed the possibility of using an ANN to provide accurate grain quality control parameters with low computational costs.

5. Acknowledgments

6. References

[1] Costa, G. E. A., Queiroz-Monici, K. S., Reis, S. M. P. M., Oliveira, A. C., Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes, Food Chemistry 94 (3) (2006) 327-330. https://doi.org/10.1016/j.foodchem.2004.11.020.

[2] Bhattacharya, K. R., Rice quality, A guide to rice properties and analysis. Woodhead Publishing Series Limited, Cambridge, 2013.

[3] Haque, E., Estimating bulk density of compacted grains in storage bins and modifications of Janssen's load equations as affected by bulk density, Food Science & Nutrition 1 (2) (2013) 150-156. https://doi.org/10.1002/fsn3.23.

[4] Bart-Plange, A., Baryeh, E. A., The physical properties of Category B cocoa beans, Journal of Food Engeneering 60 (3) (2003) 219-227. https://doi.org/10.1016/S0260-8774(02)00452-1.

[5] Wong, J. X. H., Liu, F. S. F., Yu, H. Z., Mobile app-based quantitative scanometric analysis, Analytical Chemistry 86 (24) (2014) 11966-11971. https://doi.org/10.1021/ac5035727.

[6] Shouche, S. P., Rastogi, R., Bhagwat, S. G., Sainis, J. K., Shape analysis of grains of Indian wheat varieties, Computers and Electronics in Agriculture 33 (1) (2001) 55-76. https://doi.org/10.1016/S0168-1699(01)00174-0.

[7] Chen, X., Xun, Y., Li, W., Zhang, J., Combining discriminant analysis and neural networks for corn variety identification, Computers and Electronics in Agriculture 71 (S1) (2010) S48-S53. https://doi.org/10.1016/j.compag.2009.09.003.

[8] Camargo, V. R., Santos, L. J., Pereira, F. M. V., A Proof of Concept Study for the Parameters of Corn Grains Using Digital Images and a Multivariate Regression Model, Food Analytical Methods 11 (7) (2018) 1852-1856. https://doi.org/10.1007/s12161-017-1028-6.

[9] Santos, P. M., Wentzell, P. D., Pereira-Filho, E. R., Scanner digital images combined with color parameters: a case study to detect adulterations in liquid cow's milk, Food Analytical Methods 5 (1) (2012) 89-95. https://doi.org/10.1007/s12161-011-9216-2.

[10] Borin, A., Ferrão, M. F., Mello, C., Cordi, L., Pataca, L. C. M., Durán, N., Poppi, R. J., Quantification of *Lactobacillus* in fermented milk by multivariate image analysis with least-squares support-vector machines, Analytical & Bioanalytical Chemistry 387 (2007) (3) 1105-1112. https://doi.org/10.1007/s00216-006-0971-7.

[11] Pereira, F. M. V., Milori, D. M. B. P., Pereira-Filho, E. R., Venâncio, A. L., Russo, M. S. T., Martins, P. K., Freitas-Astúa, J., Fluorescence images combined to statistic test for fingerprinting of citrus plants after bacterial infection, Analytical Methods 3 (2011) 552-556. https://doi.org/10.1039/C0AY00538J.

[12] Hagan, M. T., Demuth H. B, Beale M., Neural network design. PWS Publishing Company, Boston, 1996.

[13] Haykin, S., Neural Networks: a comprehensive foundation. Prentice Hall, New Jersey, 1999.

[14] Li, G., Shi, J., On comparing three artificial neural networks for wind speed forecasting, Applied Energy 87 (7) (2010) 2313-2320. https://doi.org/10.1016/j.apenergy.2009.12.013.

[15] Nafey, A. S., Neural network based correlation for critical heat flux in steam-water flows, International Journal of Thermal Sciences 48 (12) (2009) 2264-2270. https://doi.org/10.1016/j.ijthermalsci.2009.04.010.

[16] Niemi, H., Bulsari, A., Palosaari, S., Simulation of membrane separation by neural networks, Journal of Membrane Science 102 (1995) 185-191. https://doi.org/10.1016/0376-7388(94)00314-O.

[17] Zhang, Z., Friedrich, K., Artificial neural networks applied to polymer composites: a review, Composites Science and Technology 63 (14) (2003) 2029-2044. https://doi.org/10.1016/S0266-3538(03)00106-4.

[18] Tanzifi, M., Yaraki, M. T., Kiadehi, A. D., Hosseini, S. H., Olazar, M., Bharti, A. K., Agarwal, S., Gupta, V. K., Kazemi, A., Adsorption of Amido Black 10B from aqueous solution using polyaniline/ SiO_2 nanocomposite: Experimental investigation and artificial neural network modelling, Journal of Colloid and Interface Science 510 (2018) 246-261. https://doi.org/10.1016/j.jcis.2017.09.055.

[19] Ramil, A., López, A. J., Rivas, T., A computer vision system for identification of granite-forming minerals based on RGB data and artificial neural networks, Measurement 117 (2018) 90-95. https://doi.org/10.1016/j.measurement.2017.12.006.

[20] Walczak, S., Velanovich, V., Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks, Decision Support Systems 106 (2018) 110-118. https://doi.org/10.1016/j.dss.2017.12.007.

[21] Ortega-Zamorano, F., Jerez, J. M., Juarez, G. E., Franco, L., FPGA Implementation of Neurocomputational Models: Comparison Between Standard Back-Propagation and C-Mantec Constructive Algorithm, Neural Processing Letters 46 (3) (2017) 899-914. https://doi.org/10.1007/s11063-017-9655-x.

[22] Filletti, É. R., Roque, W. L., Estimating the mechanical competence parameter of the trabecular bone: a neural network approach, Research in Biomedical Engineering 32 (2) (2016) 137-143. https://doi.org/10.1590/2446-4740.05615.

[23] León-Roque, N., Abderrahim, M., Nuñez-Alejos, L., Arribas, S. M., Condezo-Hoyos, L., Prediction of fermentation index of cocoa beans (*Theobroma cacao* L.) based on colour measurement and artificial neural networks, Talanta 161 (1) (2016) 31-39. https://doi.org/10.1016/j.talanta.2016.08.022.

[24] Dubey, B. P., Bhagwat, S. G., Shouche, S. P., Sainis, J. K., Potential of Artificial Neural Networks in Varietal Identification using Morphometry of Wheat Grains, Biosystems Engineering 95 (1) (2006) 61-67. https://doi.org/10.1016/j.biosystemseng.2006.06.001.

[25] Demuth, H., Beale, M., Hagan, M., Neural Network toolbox™ 6: User's Guide. The Math Works, Natick, 2010.