

Research Article

Analysis of the Oil Content of Rapeseed Using Artificial Neural Networks Based on Near Infrared Spectral Data

Dazuo Yang,^{1,2} Hao Li,³ Chenchen Cao,¹ Fudi Chen,¹ Yibing Zhou,¹ and Zhilong Xiu²

¹ Key Laboratory of Marine Bio-Resources Restoration and Habitat Reparation in Liaoning Province, Dalian Ocean University, Dalian 116023, China

² College of Life Science and Technology, Dalian University of Technology, Dalian 116021, China

³ College of Chemistry, Sichuan University, Chengdu, Sichuan 610064, China

Correspondence should be addressed to Dazuo Yang; dzyang1979@hotmail.com

Received 9 May 2014; Accepted 2 June 2014; Published 23 June 2014

Academic Editor: Qingrui Zhang

Copyright © 2014 Dazuo Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The oil content of rapeseed is a crucial property in practical applications. In this paper, instead of traditional analytical approaches, an artificial neural network (ANN) method was used to analyze the oil content of 29 rapeseed samples based on near infrared spectral data with different wavelengths. Results show that multilayer feed-forward neural networks with 8 nodes (MLFN-8) are the most suitable and reasonable mathematical model to use, with an RMS error of 0.59. This study indicates that using a nonlinear method is a quick and easy approach to analyze the rapeseed oil's content based on near infrared spectral data.

1. Introduction

Infrared absorption spectroscopy is a common approach for analyzing food composition [1–3]. For a certain characteristic absorption frequency, Lambert's law provides the following equation [4, 5]:

$$\log \frac{I_0}{I} = \varepsilon \ell c, \quad (1)$$

where I_0 represents incident light intensity, I represents transmission light intensity, ε represents the attenuation coefficient, ℓ represents the distance the light travels through the material, and c represents concentration.

Equation (1) is widely used for determining food composition. However, because the wavelengths in the infrared absorption spectrum are diverse and the force of penetration is tiny, infrared absorption spectroscopy can only be used for analyzing transparent liquids. It is of great difficulty to analyze the oil content of rapeseed using infrared absorption spectroscopy. Therefore, to solve this problem, this study instead uses a nonlinear approach to analyze near-spectral data to determine the oil content of rapeseed.

2. Artificial Neural Networks

2.1. Fundamental of ANN Models. Artificial neural networks (ANN) model is composed of an interconnected group of artificial neurons. In most circumstances, an artificial neural network is an adaptive system that is equipped to be adapting continuously to new data and learning from the accumulated experience and noisy data [6, 7]. Apart from that, the system structure can be changed based on external or internal information that flows through the network during the learning phase. Meanwhile, essential information can be abstracted from data or model complex relationships between inputs and outputs [8–10].

As can be seen from Figure 1, the main structure of the artificial neural network (ANN) is made up of the input layer and the output layer. The input variables are introduced to the network by the input layer [11]. Also, the response variables with predictions, which stand for the output of the nodes in this certain layer, are provided by the network. Additionally, the hidden layer is included. The type and the complexity of the process or experimentation usually iteratively determine the optimal number of the neurons in the hidden layers [12].

TABLE 1: Data of 29 rapeseed samples.

| Sample | c (%) | Wavelength (μm) | | | | |
|--------|---------|------------------------------|------------------|------------------|------------------|------------------|
| | | $\lambda = 1.68$ | $\lambda = 1.73$ | $\lambda = 1.94$ | $\lambda = 2.10$ | $\lambda = 2.18$ |
| 1 | 39.60 | 0.70059 | 0.76184 | 0.85572 | 0.86851 | 0.87385 |
| 2 | 40.33 | 0.70218 | 0.77053 | 0.89213 | 0.87777 | 0.88360 |
| 3 | 38.71 | 0.69428 | 0.75829 | 0.88461 | 0.86710 | 0.87264 |
| 4 | 38.53 | 0.70560 | 0.77184 | 0.89991 | 0.88426 | 0.88747 |
| 5 | 37.57 | 0.69840 | 0.76610 | 0.90143 | 0.88410 | 0.88841 |
| 6 | 37.76 | 0.69448 | 0.76032 | 0.89164 | 0.87292 | 0.87551 |
| 7 | 38.22 | 0.73863 | 0.81279 | 0.93508 | 0.92134 | 0.92534 |
| 8 | 37.57 | 0.69165 | 0.76566 | 0.90251 | 0.88688 | 0.88994 |
| 9 | 39.17 | 0.68090 | 0.74233 | 0.87103 | 0.85752 | 0.85983 |
| 10 | 39.98 | 0.68634 | 0.75908 | 0.88710 | 0.87267 | 0.87799 |
| 11 | 38.93 | 0.69500 | 0.76286 | 0.88532 | 0.87407 | 0.87893 |
| 12 | 36.61 | 0.69674 | 0.76537 | 0.89669 | 0.88421 | 0.89025 |
| 13 | 39.74 | 0.69335 | 0.76262 | 0.88754 | 0.87327 | 0.87888 |
| 14 | 38.12 | 0.69504 | 0.76614 | 0.89400 | 0.88275 | 0.88853 |
| 15 | 38.89 | 0.69658 | 0.76258 | 0.88753 | 0.87480 | 0.87960 |
| 16 | 39.85 | 0.68976 | 0.75463 | 0.85723 | 0.86513 | 0.86898 |
| 17 | 39.90 | 0.68955 | 0.75582 | 0.87775 | 0.86404 | 0.86850 |
| 18 | 38.84 | 0.69057 | 0.75927 | 0.88184 | 0.87312 | 0.87900 |
| 19 | 38.32 | 0.69654 | 0.76187 | 0.88378 | 0.87097 | 0.88071 |
| 20 | 38.21 | 0.69645 | 0.76287 | 0.89703 | 0.87551 | 0.87432 |
| 21 | 38.13 | 0.70596 | 0.77703 | 0.90508 | 0.88704 | 0.89242 |
| 22 | 38.51 | 0.69456 | 0.76413 | 0.89476 | 0.87563 | 0.88302 |
| 23 | 38.31 | 0.70629 | 0.77119 | 0.89479 | 0.88420 | 0.88730 |
| 24 | 39.21 | 0.68059 | 0.74624 | 0.87088 | 0.85682 | 0.86421 |
| 25 | 38.61 | 0.68567 | 0.75328 | 0.88502 | 0.86225 | 0.87175 |
| 26 | 39.03 | 0.68050 | 0.75124 | 0.87930 | 0.86176 | 0.87014 |
| 27 | 39.31 | 0.68053 | 0.74778 | 0.87492 | 0.85486 | 0.86043 |
| 28 | 36.50 | 0.68824 | 0.75139 | 0.88715 | 0.86963 | 0.87842 |
| 29 | 37.40 | 0.69583 | 0.76104 | 0.89541 | 0.88044 | 0.88611 |

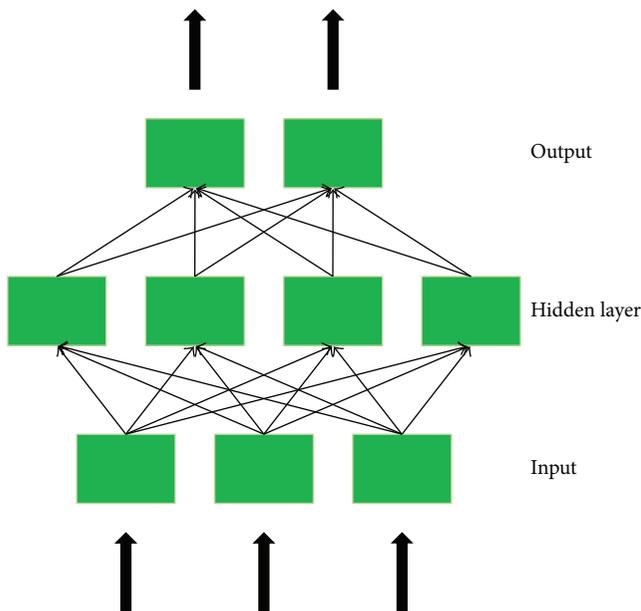


FIGURE 1: A schematic view of artificial neural network structure.

2.2. *Model Development.* Gu and Wang [12] have accomplished a series of researches from correlative precision instrument from which we could obtain data of rapeseeds' near infrared spectroscopy by analyzing absorbance under different wavelengths. We defined c (%) as the percentage composition of the oil in rapeseed. Data of 29 rapeseed samples are shown on Table 1.

In order to confirm the most suitable and robust ANN model in analyzing the oil content of rapeseed, 21 models were established including linear prediction model, general regression neural networks (GRNN) [14] and multilayer feed-forward neural networks (MLFN) [15, 16]. Into that matter, nodes of MLFN models were set to be from 2 to 20, so that the most robust MLFN model could be found. The independent variables are the absorbancies under the wavelength of $1.68 \mu\text{m}$ (reference wavelength), $1.73 \mu\text{m}$ (characteristic absorption wavelength of fat), $1.94 \mu\text{m}$ (characteristic absorption wavelength of water), $2.10 \mu\text{m}$ (characteristic absorption wavelength of starch), and $2.18 \mu\text{m}$ (characteristic absorption wavelength of protein), respectively, while the dependent variable is the percentage composition of the oil in rapeseed. Training set is consist of 24 samples while the rest

TABLE 2: Results of different models in analyzing oil content of rapeseed.

| Model | Trained samples | Tested samples | RMS error | Stopped reason |
|-------------------|-----------------|----------------|-----------|----------------|
| Linear prediction | 24 | 5 | 0.60 | Autostopped |
| GRNN | 24 | 5 | 0.74 | Autostopped |
| MLFN 2 nodes | 24 | 5 | 1.04 | Autostopped |
| MLFN 3 nodes | 24 | 5 | 1.23 | Autostopped |
| MLFN 4 nodes | 24 | 5 | 0.88 | Autostopped |
| MLFN 5 nodes | 24 | 5 | 1.29 | Autostopped |
| MLFN 6 nodes | 24 | 5 | 1.31 | Autostopped |
| MLFN 7 nodes | 24 | 5 | 2.20 | Autostopped |
| MLFN 8 nodes | 24 | 5 | 0.59 | Autostopped |
| MLFN 9 nodes | 24 | 5 | 3.39 | Autostopped |
| MLFN 10 nodes | 24 | 5 | 1.93 | Autostopped |
| MLFN 11 nodes | 24 | 5 | 0.83 | Autostopped |
| MLFN 12 nodes | 24 | 5 | 1.54 | Autostopped |
| MLFN 13 nodes | 24 | 5 | 1.10 | Autostopped |
| MLFN 14 nodes | 24 | 5 | 1.57 | Autostopped |
| MLFN 15 nodes | 24 | 5 | 3.02 | Autostopped |
| MLFN 16 nodes | 24 | 5 | 1.67 | Autostopped |
| MLFN 17 nodes | 24 | 5 | 1.07 | Autostopped |
| MLFN 18 nodes | 24 | 5 | 1.72 | Autostopped |
| MLFN 19 nodes | 24 | 5 | 0.85 | Autostopped |
| MLFN 20 nodes | 24 | 5 | 2.79 | Autostopped |

of the samples are considered to be the testing set. To ensure the accuracy of the experiments, we did the training process repeatedly. The composing of trained samples and tested samples is different in each experiment. Results of the 21 models were obtained by correlative software, which are shown in Table 2.

Results presented by Table 2 imply that the lowest RMS error of testing exists in the MLFN model with 8 nodes (MLFN-8), which is 0.59, lower than those generated by linear prediction model and GRNN model. And the accuracy rate of the testing is 100% with the permission error. Therefore, the MLFN-8 model is proved to be an accurate and robust model.

3. Results and Discussion

3.1. Training Results of MLFN-8. Training and testing results of MLFN-8 model were extracted from the experiments. For more intuitionistic, six figures described by data are used to portray the training and testing results, which are shown in Figures 2 to 7.

In training process, the comparison result between predicted values and actual values is depicted by Figure 2. The regulation between predicted values and actual values implies that the training process is precise.

Figure 3 depicts the relationship between residual values and actual values during training process, showing that the residual values are relatively concentrated.

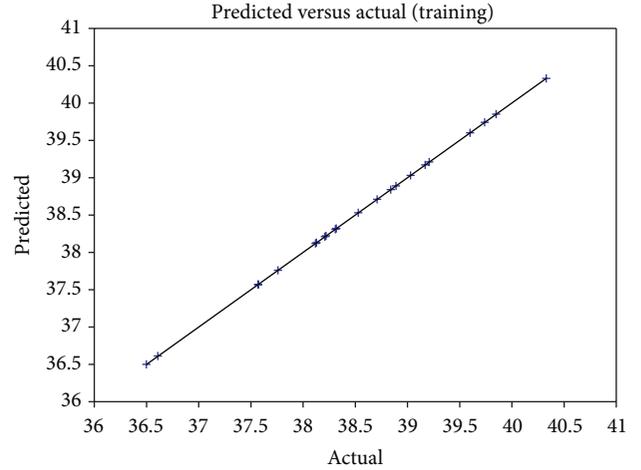


FIGURE 2: Comparison between predicted values and actual values during training process.

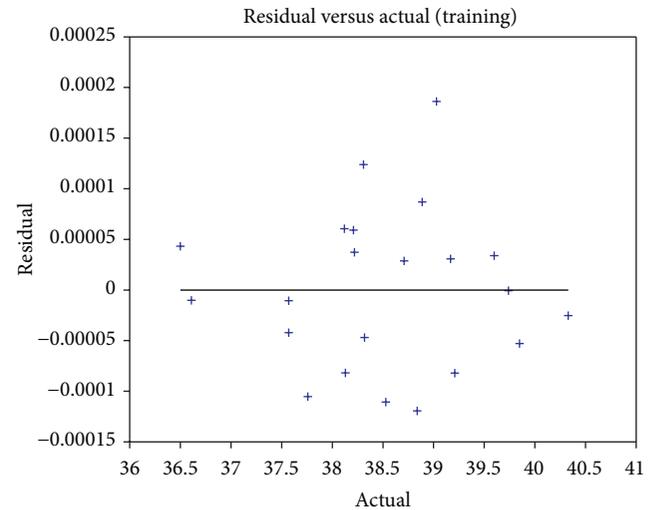


FIGURE 3: Comparison between residual values and actual values during training process.

Different from Figure 3, Figure 4 depicts the relationship between residual values and predicted values during training process. Similar to the result shown in Figure 3, the residual values present the same phenomenon as Figure 3, which indicates that the training process is precise.

In general, Figures 2, 3, and 4 depict the results of training process, showing that the values are concentrated and correspond with the normal training process of MLFN-8 model. It is worth mentioning that the residue values are generally tiny and close to zero, which implies that the training process is correct and precise.

3.2. Testing Results of MLFN-8. To analyze the testing process, three figures were used to present the average values of testing results, which are shown in Figures 5 to 7.

In testing process, as shown in Figure 5, the comparison between predicted values and actual values is also close to linear situation, which means that the MLFN-8 model is precise while predicting.

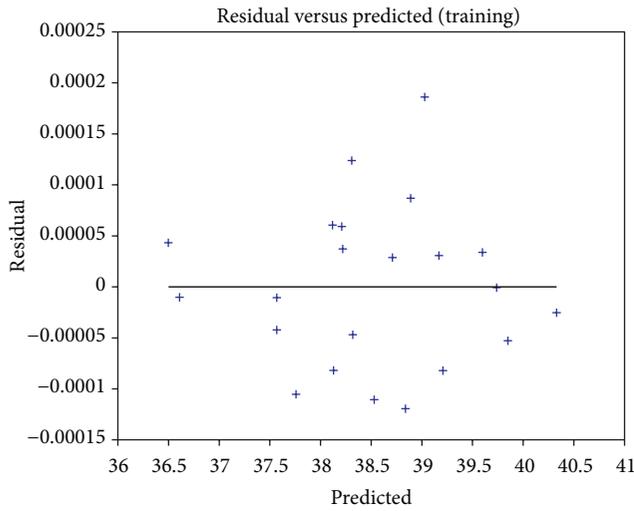


FIGURE 4: Comparison between residual values and predicted values during training process.

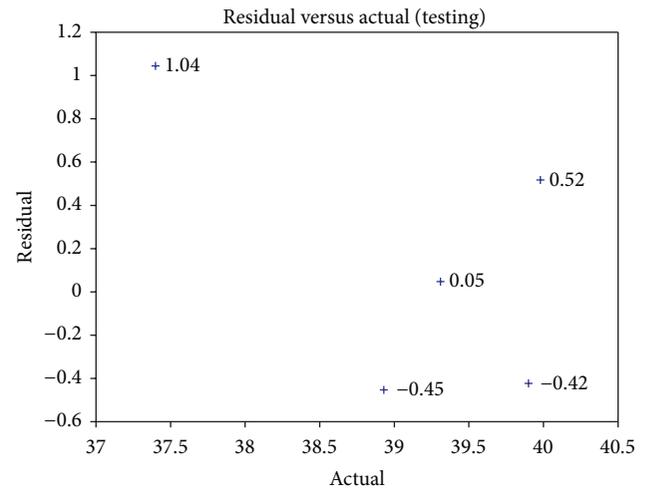


FIGURE 6: Comparison between residual values and actual values during testing process.

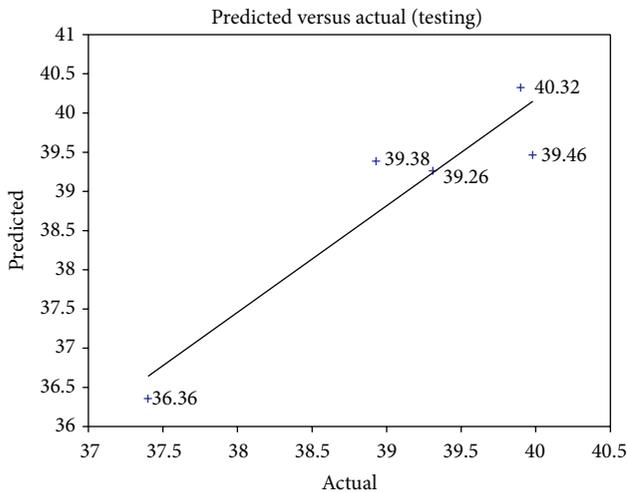


FIGURE 5: Comparison between predicted values and actual values during testing process.

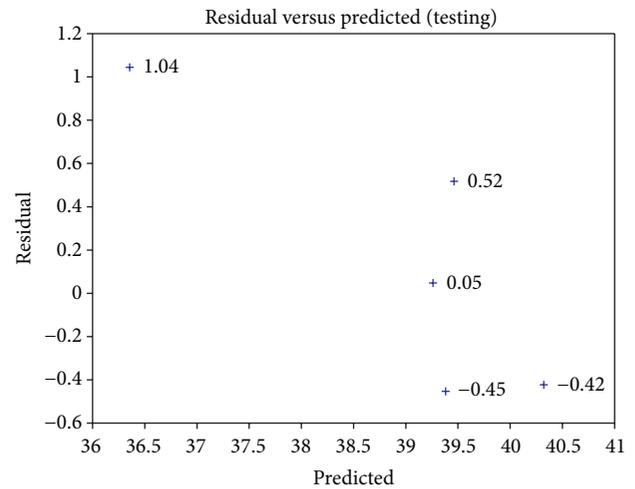


FIGURE 7: Comparison between residual values and predicted values during testing process.

In order to confirm the robustness of comparison between residual values and actual values as well as the comparison between residual values and predicted values, we plotted the comparison between residual values and these two kinds of values, which are shown in Figures 6 and 7.

Figures 5, 6, and 7 depict the average testing process of the MLFN-8 model. All the values shown in the three figures are the average values, from which we can draw a conclusion that the model is accurate and robust.

According to the results presented above, MLFN-8 model is proved to be a suitable and rational model in determining the oil content of rapeseed.

3.3. Discussion. There are several previous studies that are relative to the field we studied [12, 17–20]. Gu and Wang [12] analyzed the oil content of rapeseed by multiple linear regression based on near spectral data, which is the chief inspiration

of our work. In contrast, our work has a higher robustness and precision since the core we paid attention to is the well-fitted nonlinear function. Besides, Madsen [17] established a quick determination approach of oil content in rapeseed by a commercial nuclear magnetic resonance spectrometer. Tkachuk [18] utilized a near infrared reflectance technique to determine oil, protein, chlorophyll, and glucosinolate content in whole rapeseed kernels. In addition, Velasco and relative coworkers [19] used near-infrared reflectance spectroscopy to estimate the seed weight, oil content, and fatty acid composition in intact single seeds of rapeseed. Shafii and his coworkers [20] analyzed the interaction effects on the winter rapeseeds yield and oil content. These researches can analyze the oil content and other properties of rapeseeds effectively, which can be seen as the great references. However, these analytical approaches still need complex manual operation and the process is intricate to some extent. Our study has successfully proved that the oil content of rapeseed can

be analyzed by artificial neural networks, which is a quick and easy method that can be calculated automatically by computer.

In the field of food science and analytical chemistry, oil content of rapeseed reveals the yield of the relative products in practical applications. Taking one of the production steps as an example, people should estimate and evaluate the oil content of the rapeseed samples before mass run. Therefore, using artificial neural networks can achieve this step in a high effective way.

4. Conclusion

Oil content of rapeseed is a crucial aspect on practical applications of food science and chemistry. In this paper, instead of using traditional analytical methods, we successfully used artificial neural networks (ANNs) method to analyze the oil content of 29 rapeseed samples based on near spectral data with different wavelengths. Results show that the multilayer feed-forward neural networks with 8 nodes (MLFN-8) are the most suitable and reasonable mathematical model during experiments. In future research, we will aim at looking for the explicit nonlinear functions of near spectral data in the analysis of rapeseed's oil content.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was funded by the National Marine Public Welfare Research Project (nos. 201305002 and 201305043) and the Natural Science Foundation of Dalian (no. 2012003219).

References

- [1] K. Motobayashi, K. Minami, N. Nishi et al., "Hysteresis of potential-dependent changes in ion density and structure of an ionic liquid on a gold electrode: in situ observation by surface-enhanced infrared absorption spectroscopy," *The Journal of Physical Chemistry Letters*, vol. 4, no. 18, pp. 3110–3114, 2013.
- [2] L. V. Brown, K. Zhao, N. King, H. Sobhani, P. Nordlander, and N. J. Halas, "Surface-enhanced infrared absorption using individual cross antennas tailored to chemical moieties," *Journal of the American Chemical Society*, vol. 135, no. 9, pp. 3688–3695, 2013.
- [3] Y.-T. Su, Y.-H. Huang, H. A. Witek, and Y.-P. Lee, "Infrared absorption spectrum of the simplest criegee intermediate CH_2OO ," *Science*, vol. 340, no. 6129, pp. 174–176, 2013.
- [4] J. M. Parnis and K. B. Oldham, "Beyond the Beer-Lambert law: the dependence of absorbance on time in photochemistry," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 267, pp. 6–10, 2013.
- [5] K. Fuwa and B. L. Vallee, "The physical basis of analytical atomic absorption spectrometry: the pertinence of the Beer-Lambert law," *Analytical Chemistry*, vol. 35, no. 8, pp. 942–946, 1963.
- [6] N. Gupta, "Artificial neural network," *Network and Complex Systems*, vol. 3, no. 1, pp. 24–28, 2013.
- [7] H. Li, X. F. Liu, S. J. Yang et al., "Prediction of polarizability and absolute permittivity values for hydrocarbon compounds using artificial neural networks," *International Journal of Electrochemical Science*, vol. 9, no. 7, pp. 3725–3735, 2014.
- [8] Y. Wang, H.-C. Han, J. Y. Yang, M. L. Lindsey, and Y. Jin, "A conceptual cellular interaction model of left ventricular remodeling post-MI: dynamic network with exit-entry competition strategy," *BMC Systems Biology*, vol. 4, no. 1, article S5, 2010.
- [9] Y. Wang, T. Yang, Y. Ma et al., "Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction," *BMC Genomics*, vol. 13, supplement 6, article S21, 2012.
- [10] T. Yang, Y. A. Chiao, Y. Wang et al., "Mathematical modeling of left ventricular dimensional changes in mice during aging," *BMC Systems Biology*, vol. 6, supplement 3, article S10, 2012.
- [11] C. H. Aladag, A. Kayabasi, and C. Gokceoglu, "Estimation of pressuremeter modulus and limit pressure of clayey soils by various artificial neural network models," *Neural Computing and Applications*, vol. 23, no. 2, pp. 333–339, 2013.
- [12] W. Z. Gu and Y. X. Wang, "Analysis of rapeseed oil by linear regression using near spectral data," *Journal of the Chinese Cereals and Oils Association*, vol. 10, no. 2, pp. 57–64, 1995 (Chinese).
- [13] Ö. Polat and T. Yildirim, "FPGA implementation of a General Regression Neural Network: an embedded pattern classification system," *Digital Signal Processing*, vol. 20, no. 3, pp. 881–886, 2010.
- [14] C. H. Chen, T. K. Yao, C. M. Kuo et al., "Evolutionary design of constructive multilayer feedforward neural network," *Journal of Vibration and Control*, vol. 19, no. 16, pp. 2413–2420, 2013.
- [15] S. Mirjalili, S. Z. Mohd Hashim, and H. Moradian Sardroudi, "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm," *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 11125–11137, 2012.
- [16] R. Pahlavan, M. Omid, and A. Akram, "Energy input-output analysis and application of artificial neural networks for predicting greenhouse basil production," *Energy*, vol. 37, no. 1, pp. 171–176, 2012.
- [17] E. Madsen, "Nuclear magnetic resonance spectrometry as a quick method of determination of oil content in rapeseed," *Journal of the American Oil Chemists Society*, vol. 53, no. 7, pp. 467–469, 1976.
- [18] R. Tkachuk, "Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy," *Journal of the American Oil Chemists Society*, vol. 58, no. 8, pp. 819–822, 1981.
- [19] L. Velasco, C. Möllers, and H. C. Becker, "Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy," *Euphytica*, vol. 106, no. 1, pp. 79–85, 1999.
- [20] B. Shafii, K. A. Mahler, W. J. Price et al., "Genotype X environment interaction effects on winter rapeseed yield and oil content," *Crop Science*, vol. 32, no. 4, pp. 922–927, 1992.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

