Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2036–2044, DOI 10.12694/scpe.v26i5.4714

NEAR INFRARED SPECTROSCOPY DETECTION OF SOYBEAN FATTY ACID CONTENT BASED ON NEURAL NETWORK COMBINED WITH GENETIC ALGORITHM

LI MIN*, QI XIAOCUI†, LI RONGYAO‡, WANG RUFENG§, AND YU DANDAN¶

Abstract. In order to solve the problem of poor performance in rapid analysis of soybean fatty acid content using traditional methods, the author proposes a near-infrared spectroscopy detection method for soybean fatty acid content based on neural network combined with genetic algorithm. The author first collected sample spectra and preprocessed them using mean centralization and Savitzky Golay smoothing differentiation method. Then, the optimal band was selected through segmented combination modeling, and genetic algorithm was used to further screen wavelength points sensitive to content prediction modeling. Finally, the spectral data was decomposed using principal component analysis (PCA), and the score matrix was input into a 3-layer neural network for training. The optimal model was established through parameter optimization. The experimental results show that the calibration relative analysis error RPD of the model based on genetic algorithm combined with neural network is, indicating that the model has good prediction accuracy and stability. To further validate the model's reliability in practical applications, the correlation coefficient R between the predicted values and the standard values within the model's prediction range was found to be 0.986, indicating a strong linear relationship between them. The relative deviation of each value is less than 2%, and the standard deviation of the difference is 0.0091. Paired t-test =1.2706 was performed on both methods, which is less than =2.354. This indicates that there is no significant difference between the analysis results of this method and the standard method, verifying the strong predictive ability of the model in practical applications. The model designed by the author can accurately and efficiently complete near-infrared spectroscopy detection.

Key words: Neural network, Genetic algorithm, near-infrared spectroscopy, Soybeans, fatty acid

1. Introduction. Soybean Glycine max (L.) Merr. is an important source of plant protein and oil for humans, with an oil content of around [1]. It is the main raw material for the plant oil processing industry. Soybean oil contains a large amount of unsaturated fatty acids, which are beneficial to human health and have attracted widespread attention. However, due to the generally high content of linolenic acid in soybean oil, reducing its storage period, and improving the fatty acid composition from a breeding perspective is an effective method. Therefore, improving the fatty acid composition has become an important part of soybean quality breeding [2]. In the past, gas chromatography and liquid chromatography techniques were mainly used for the analysis of soybean fatty acids, which not only consumed a lot of manpower and material resources, but also caused time waste and reduced breeding efficiency. It can be seen that how to quickly and accurately determine the fatty acid composition in soybean seeds to accelerate soybean quality breeding is an urgent problem to be solved [3].

Near infrared spectroscopy analysis technology is an emerging qualitative and quantitative analysis technique that uses sample spectral data. The rapid and non-destructive measurement method has made this technology increasingly widely used in the determination of physicochemical properties and state monitoring of soybean fatty acids. Near infrared spectroscopy mainly refers to the harmonic and combined frequency absorption generated by the vibration of hydrogen containing groups in molecules. Due to the complex composition of lubricating oil, the spectral bands formed in the near-infrared spectral wavelength range overlap severely, contain a large amount of non target information, and are subject to external signal interference such as background and instrument noise. As a result, spectral preprocessing and calibration techniques greatly

*Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
†Admission and Employment Office, Cangzhou Medical College, HeBei, 061001, China.
‡Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
§Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
¶Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China (Corresponding author, dan5507803@126.com).

influence the model's predictive performance. Commonly used wavelength selection methods include the correlation coefficient method, interval partial least squares (iPLS), the successive projections algorithm (SPA), and genetic algorithm (GA). The genetic algorithm, a global optimization-based wavelength selection method, is particularly effective for selecting wavelengths in multi-component systems and can be integrated with various calibration methods to enhance and optimize the model's predictive performance [4,5,6].

2. Literature Review. Near-infrared spectroscopy (NIRS) is an innovative technique that leverages the vibrational absorption characteristics of organic substances in the near-infrared spectrum to quickly measure the content of various chemical components in samples. At present, there have been some research reports abroad on using NIRS technology to determine fatty acid composition in different crops. Deng, Z. et al. investigated the use of fatty acid composition and near-infrared spectroscopy for tracing the geographical origin of camellia oil in Hainan. They measured the relative content of 16 fatty acids in camellia oil samples using gas chromatography and gathered spectral data through near-infrared spectroscopy. The study demonstrated that both fatty acid composition and near-infrared spectroscopy are effective tools for accurately determining the geographical origin of camellia oil [7]. Zhe, L. I. et al. devised a method using near-infrared reflectance spectroscopy (NIRS) to rapidly and approximately assess the nutritional composition of Pacific oysters (Crassostrea gigas) [8]. Tonolini, M. et al. explored the application of near-infrared (NIR) spectroscopy combined with chemometrics to monitor the degumming process by measuring diglyceride levels in soybean oil in both offline and online environments. To develop an NIR spectroscopy prediction model, they studied 15 lab-scale batches of enzyme degumming, utilizing various water, acid, and enzyme doses from the final screening design. The findings suggest that glycerol diester content serves as a reliable indicator of enzyme effectiveness, and that NIR spectroscopy is an appropriate analytical technique for real-time quantification of glycerol diesters [9].

The author studied the use of genetic algorithm for wavelength screening in the determination of soybean fatty acid content by near-infrared spectroscopy, and combined it with a highly nonlinear prediction ability backpropagation neural network to establish an analysis and correction model, effectively improving the accuracy and adaptability of the model prediction.

3. Method.

3.1. Basic Framework. The author presents a convolutional neural network (CNN) learning method enhanced by genetic algorithm optimization, as illustrated in Figure 3.1. The CNN framework includes input layers, convolutional layers, pooling (sampling) layers, fully connected layers, and output layers. The number of convolutional and pooling layers is not fixed, but rather adjusted based on computational tasks and available resources [10]. A significant challenge with traditional CNNs is that their learning performance heavily depends on the initial weight settings of the convolutional and fully connected layers. To address this issue, the author employed a genetic algorithm to optimize the weights of each convolutional and fully connected layer during the training process. The basic idea is to construct the chromosomes of the genetic algorithm based on these weights, and obtain the optimal solution through selection, crossover, and mutation operations.

3.2. Configuration and Training of Convolutional Neural Networks.

(1) Configuration of Convolutional Neural Networks. Convolutional neural network classifiers typically consist of two parts: convolutional neural networks used for feature extraction and classifiers used for feature classification. The hierarchical framework of convolutional neural networks is shown in Figure 3.2. Assuming that the feature dimension of each sample is d and the total number of samples in the training set is m, storing the features of each sample in rows, where each row represents a sample, the input layer size of the convolutional neural network is $m \times d$ [11].

During the training process, continuously learn convolutional masks to identify the most discriminative local patterns in the classification process. Due to the fact that in the input layer, the features of each d-dimensional sample are stored in rows, the convolution mask only uses linear convolution masks in the horizontal direction [12]. To minimize the complexity of pattern recognition and enhance computational efficiency, the author applied a convolution mask in the initial convolutional layer. In the following convolutional layers, the number of convolution masks sequentially doubles—for instance, the second convolutional layer has 2 masks, the third layer has 4 masks, and so forth.

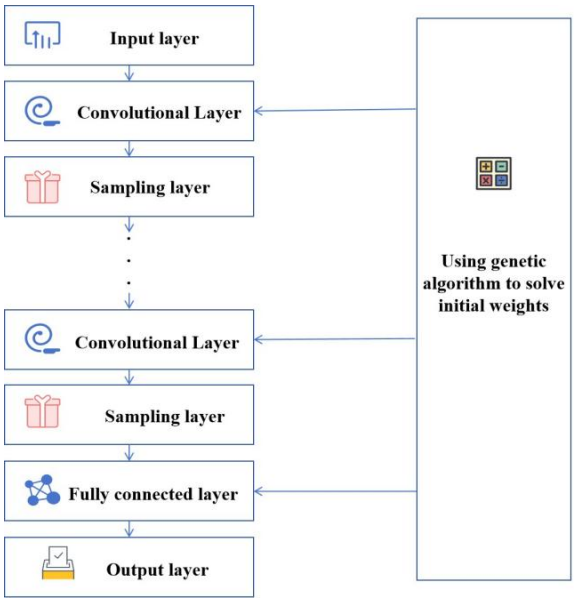

Fig. 3.1: Convolutional Neural Network Learning Framework Optimized with Genetic Algorithm

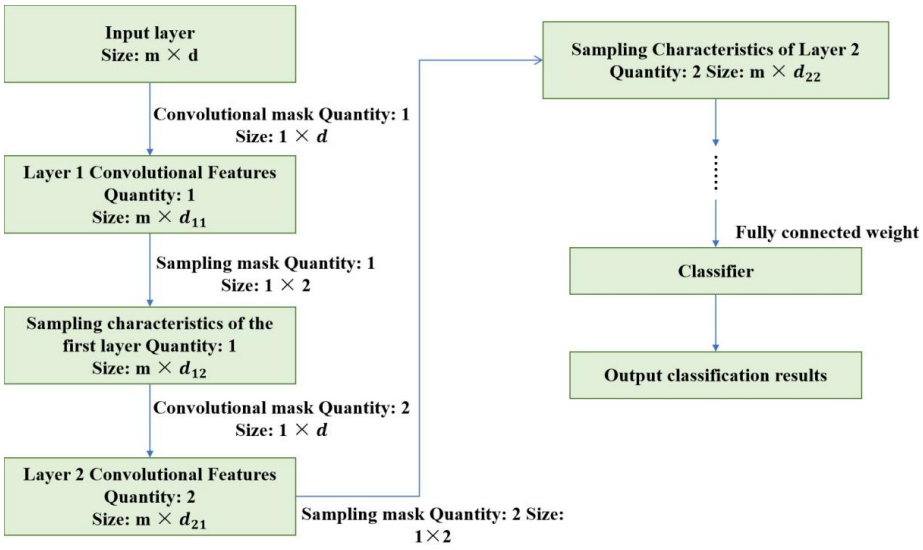

Fig. 3.2: Convolutional Neural Network Layered Framework

In addition, for ease of calculation, each convolutional mask uses the same mask size, denoted as $1 \times d'$. The sampling layer contains the same number of masks as the preceding convolutional layer, with each mask having a size of 1 * 2. After passing through the first convolutional layer, the dimension of the features is

$$d_{11} = d - d' + 1 \tag{3.1}$$

The feature dimension after the first sampling layer is

$$d_{12} = \frac{d - d' + 1}{2} \tag{3.2}$$

Similarly, the feature dimension after passing through the second convolutional layer and sampling layer is

$$d_{22} = \frac{d_{12} - d' + 1}{2} = \frac{1}{4}d - \frac{3}{4}d' + \frac{3}{4} \tag{3.3}$$

If the classifier to be trained only uses two convolutional layers and two sampling layers to extract features, then the dimension d' of each convolutional mask satisfies the condition:

$$\frac{1}{4}d - \frac{3}{4}d' + \frac{3}{4} > 0 \tag{3.4}$$

The final layer of the convolutional neural network classifier is the fully connected layer, which uses a classifier to categorize the features extracted by the convolutional and sampling layers. The author uses a commonly used logistic regression classifier [13,14].

(2) Training of Convolutional Neural Networks. The main problem with using the steepest descent method to train convolutional neural networks is that they are prone to getting stuck in local optima. To address this issue, the author employs a genetic algorithm, with the core concept being to determine the initial optimal weights for the convolutional neural network classifier using the genetic algorithm. Specifically, the weights of the convolutional and fully connected layers in the network are treated as the population for the genetic algorithm, and the weight combinations for each group are binary encoded to form the chromosomes of the genetic algorithm [15].

Assuming the number of convolutional layers in the convolutional neural network classifier to be trained is k, according to the aforementioned convolutional neural network configuration, the number of convolutional masks in the i-th convolutional layer is 2^{i-1}, the total number of convolutional masks in the convolutional neural network is $\sum_{i=1}^{k} 2^{i-1}$, the total number of weights in the convolutional layers is $d' \times \sum_{i=1}^{k} 2^{i-1}$, and the weight of a fully connected layer is added. Therefore, the number of bits of the genetic algorithm chromosome generated by binary encoding is

$$B = d' \times \sum_{i=1}^{k} 2^{i-1} + 1 \tag{3.5}$$

In this way, the first B-1 bit of the chromosome is used to encode the convolutional mask, and the last bit is used to initialize the fully connected layer [16]. Conventional convolutional neural networks typically employ the steepest descent algorithm for training, and their learning performance is significantly influenced by the choice of initial weights in the network. The author uses genetic algorithms to generate different initial weights, so that the optimal weight obtained through selection, crossover, and mutation operations is significantly better than the initial weight obtained through random selection.

3.3. Genetic algorithm solution for initial weights. Genetic algorithm is an adaptive heuristic search method, which originates from the natural law of biological evolution, namely survival of the fittest. In this method, each chromosome in the population undergoes a series of selection, crossover, and mutation changes to obtain a new chromosome with stronger adaptability. The implementation process of this algorithm is shown in Figure 3.3.

The author treats the weights of both the convolutional and fully connected layers in the neural network as the genetic algorithm's population. Initially, several sets of weights are chosen, and each set is encoded into chromosomes [17]. By performing selection, crossover, and mutation operations on these chromosomes, various weight combinations are produced. Then, the fitness value of chromosomes is calculated, which is the classification accuracy of the convolutional neural network under various weight combination methods.

Based on this, the final population, that is, the optimal weight combination, is selected.

The solving process of genetic algorithm is shown in Figure 3.3, and the specific steps will not be repeated by the author. The author focuses on introducing the fusion method of genetic algorithm and convolutional neural network. Correspondingly, the core of the above algorithm is the encoding of chromosomes and the solution of chromosome fitness values. The encoding method of chromosomes has been explained in the previous section. Here, we will specifically introduce the steps used by the author to solve the chromosome fitness values, which are as follows:

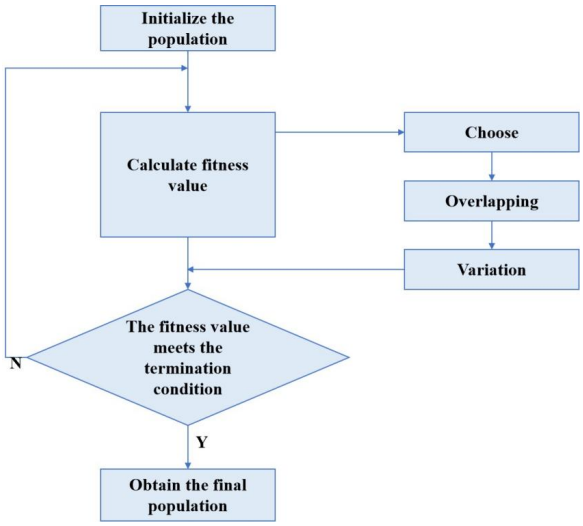

Fig. 3.3: Process of Genetic Algorithm

Step 1: Decode the encoded weight set to retrieve the initial weights;
Step 2: Apply these initial weights to the corresponding convolutional and fully connected layers of the neural network;
Step 3: Train the convolutional neural network classifier using the PL steepest descent algorithm;
Step 4: Assess the classification accuracy of the trained network and use this accuracy as the fitness score for the corresponding chromosome.

Among them, the parameter pl involved in step 3 refers to the number of iterations for training the convolutional neural network. In order to avoid overfitting of the data, the number of iterations should not be set too large [18].

The use of genetic algorithms can label many local valleys, which facilitates the steepest descent algorithm to quickly find local optimal values from these valleys. For a given population size n, after executing multiple rounds of genetic algorithm, the final population can be obtained, which contains n sets of initial weights. These n initial weights can be used to train n convolutional neural network classifiers. Below is an introduction on how to combine these n convolutional neural network classifiers for feature classification.

3.4. Joint classifier. Assuming the output value of a classifier is $o_1, o_2, \cdots, o_c$, where c is the total number of categories. Generally, the output value is binary, which means that for $i = 1, 2, \cdots, c$, there are

$$o_i = \begin{cases} 1, \text{The sample belongs to category i} \\ 0, \text{The sample does not belong to category i} \end{cases} \tag{3.6}$$

Moreover, assuming that each input sample belongs to only one of the c categories, that is

$$\sum_{i=1}^{c} o_i = 1 \tag{3.7}$$

For any test sample, the output category after classification by the classifier is

$$p = arg \max_{1 \leqslant i \leqslant s} o_i \tag{3.8}$$

For the n classifiers generated earlier, the author used the maximum value criterion to fuse the classification results of multiple classifiers and searched for the binary decoding output in the joint classifier model.

Specifically, assuming that the output value of the i-th classifier in the joint classifier model is $o_{1i}, o_{2i}, \cdots, o_d$, the jth output value of the joint classifier model can be represented as

$$f_j = \max_{1 \leqslant i \leqslant n} o_j \tag{3.9}$$

For any test sample x, if its set of output values in the joint classifier model is denoted as $\{f_j | j = 1, 2, \cdots, c\}$, then the category of the test sample should be

$$p = arg \max_{1 \leqslant j \leqslant i} f_j \tag{3.10}$$

In this way, a joint classifier composed of multiple classifiers can achieve higher classification accuracy [19].

3.5. Experimental verification.

3.5.1. Instrument device. The soybean fatty acid content detection device used in the experiment is the Perten8620 near-infrared spectrometer produced by the Swedish company Borton. This instrument has 20 filters corresponding to 20 wavelengths, and is equipped with NIR software program. The instrument used for chemical determination of soybean fatty acid content is the GC9A gas chromatograph produced by Shimadzu, Japan.

3.5.2. Near infrared spectroscopy acquisition. The near-infrared spectroscopy of the sample is collected after the instrument is preheated for 1 hour and the working state is stable. The instrument sample cell is a 5mm quartz flow colorimetric dish. Before spectral collection, it should be cleaned with n-hexane at least three times and air dried. Using air as a reference, perform sample spectral scanning in the 1000-1800nm wavelength range, scan 10 times, and take the average spectrum for modeling.

3.5.3. Spectral preprocessing and calibration methods. A total of 50 samples were taken, and 40 samples were selected as the calibration set and 10 samples as the validation set using the Kemand Stone classification method. In order to reduce random noise and background interference in the original spectrum, mean centering and Savitzky Golay smoothing differentiation method were used for spectral preprocessing. To identify spectral bands with significant informational content for modeling, the spectrum is divided into several subintervals for band selection, and a genetic algorithm is employed to refine the choice of wavelength points. Within the chosen wavelength range, a backpropagation neural network (BP-ANN) is then utilized to develop a calibration model, optimizing parameters such as the number of hidden layer nodes, learning rate, momentum factor, and the number of training iterations. The constructed model is evaluated and optimized through calibration set correlation coefficient (R_c), validation set correlation coefficient (R_p), cross validation root mean square error (SEC), and prediction root mean square error (SEP). Spectral analysis modeling is implemented using Unscrambler 10.4 and Matlab 7.0 software [20].

4. Results and Discussion. The mean centrality method was used to reduce background interference in the near-infrared spectra of the collected soybean fatty acid samples in use, reflecting the information of spectral differences. The author uses the S-G smooth derivative method to filter out spectral noise and eliminate the influence of baseline drift. The window width for smooth differentiation is optimized between 13 and 21 points, with an interval of 2. The modeling results SEC and SEP are used for evaluation, as shown in Figure 4.1. The smooth window width is set to 17 points, and the modeling effect is ideal when taking the first-order derivative of the spectrum.

The genetic algorithm, which selects wavelengths based on optimizing the objective function, offers superior global optimization capabilities compared to the correlation coefficient method. In addition, the determination of the threshold in the correlation coefficient method generally relies on experience and is highly subjective. Using genetic algorithm to screen wavelengths is beneficial for improving the prediction accuracy and stability of the model. Select wavelengths within the measurement range of 1000-1800nm for soybean fatty acid near-infrared spectroscopy. First, divide the entire spectrum into multiple sub intervals to investigate the effectiveness of different interval combinations in modeling, and screen for band combinations that are sensitive to predicting kinematic viscosity. Then, genetic algorithm is used to further optimize the selected bands and enhance the

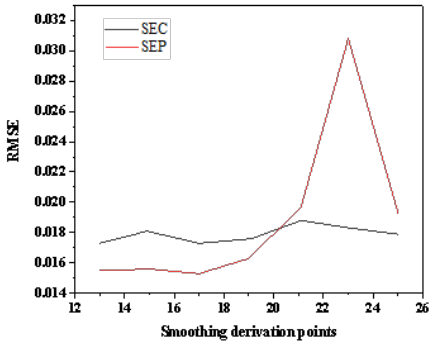

Fig. 4.1: Optimization of Smooth Derivative Points

Table 4.1: Modeling wavelength optimization for multi interval combinations

Sub-region Number	Main Factor	Selected Regions	SEC	SEP
21	5	[1 4 5 17]	0.02756	0.02645
23	7	[1 6 16 18]	0.02347	0.02424
25	6	[1 5 8 18]	0.02157	0.02201
27	7	[1 4 6 20]	0.02341	0.02406
29	7	[1 7 15 21]	0.02545	0.02447

Table 4.2: Modeling effect of wavelength selection using genetic algorithm and correlation coefficient method

Wavelength Selection Method	Wavelength Points	R_c	R_p	SEC	SEP
Genetic Algorithm	87	0.96114	0.97368	0.01712	0.01508
Correlation Coefficient	267	0.93871	0.92617	0.01981	0.0206

predictive ability of the model. The full spectrum is divided into 21-29 sub intervals, with band combinations of 3 and 4. SEC and SEP are used as evaluation criteria to preliminarily optimize the modeled wavelengths. The results are shown in Table 4.1.

According to Table 4.1, when the spectrum is divided into 25 sub intervals, the selected spectral bands are 1000-1031nm, 1128-1159nm, 1224-1255nm, and 1544-1575nm, respectively, resulting in the best modeling effect. Adopting R The improved genetic algorithm proposed by Leardi screens 128 wavelength points in the combined spectral regions of the four bands mentioned above, removing wavelength points that are insensitive to viscosity prediction and further optimizing the modeling wavelength. The genetic algorithm's objective function is defined as the root mean square error from cross-validation, with the algorithm configured to an initial population size of 32, a crossover rate of 0.05, a mutation rate of 0.01, and a maximum of 100 evolutionary generations. Table 4.2 displays the selected wavelength points and model evaluation metrics for wavelength modeling using both the genetic algorithm and the correlation coefficient method.

The author developed a 3-layer neural network model using a combination of principal component analysis (PCA) and backpropagation (BP). They first applied principal component decomposition to the preprocessed spectrum, then selected the principal components that accounted for 99% of the cumulative contribution. The resulting principal component scores were used as input for the BP network modeling. The GA algorithm selects 88 input elements for spectral data, which are reduced to 10 after principal component analysis, reducing the

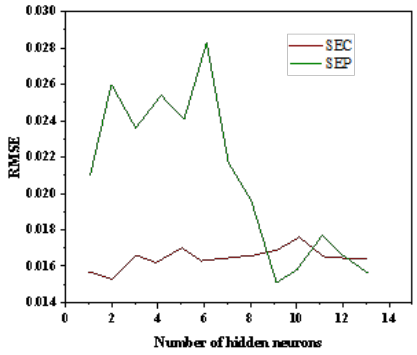

Fig. 4.2: The influence of the number of hidden layer nodes on modeling performance

Table 4.3: Evaluation of Correction Method Analysis Results

Correction Algorithm	R_c	R_p	SEC	SEP
PLS	0.94701	0.91212	0.01701	0.01625
BP-ANN	0.96014	0.97268	0.01612	0.01408

complexity of the training process. The number of hidden layer nodes is optimized based on the root mean square error of the modeling results, ranging from 1 to 13 points, as shown in Figure 4.2.

As observed in the figure above, the SEC and SEP values stabilize at lower levels when the number of hidden layer nodes reaches 9. Consequently, the optimal number of hidden layer nodes is determined to be 9. Additionally, the model uses the tansig function as the transfer function for the hidden layer and the purelin function for the output layer. When the number of training iterations reached 33, the training errors of the modeling set and prediction set reached the lowest value, and the predictive ability of the model tended to stabilize, establishing the optimal prediction model. The comparison between the constructed model and the results of linear partial least squares (PLS) modeling is shown in Table 4.3.

Table 4.3 reveals that the correlation coefficients for both the calibration and validation sets of the BP-ANN model are high and show minimal variation. This indicates that the model has strong prediction accuracy and stability. Compared with PLS method, the BP-ANN method effectively integrates the nonlinear information present in the spectrum, improving the predictive performance of the model. The calibration relative analysis error RPD of the model based on genetic algorithm combined with neural network is 5.01>3, indicating that the model has good prediction accuracy and stability. In order to further verify the reliability of the model in practical applications, within the prediction range of the model, the correlation coefficient R between the predicted value and the standard value is 0.986, indicating a good linear relationship between the two. The relative deviation S_d of each value is less than 2%, and the standard deviation of the difference is 0.0091. Paired t-test t_{solid}=1.2706 was performed on both methods, which is less than $t_{0.05,7}$=2.354. This indicates that there is no significant difference between the analysis results of this method and the standard method, verifying the strong predictive ability of the model in practical applications.

5. Conclusion. The author introduces a method for detecting soybean fatty acid content using near-infrared spectroscopy, enhanced by a neural network and genetic algorithm. This approach employs a genetic algorithm-based wavelength selection technique to identify optimal wavelengths for modeling, thereby maximizing the use of spectral information while reducing the number of variables in the model. By integrating the BP neural network, the method effectively addresses the nonlinear relationship between spectral data and fatty acid content, significantly boosting the accuracy and reliability of predictions. The analysis of actual samples

using optimization models yielded ideal prediction results, further verifying the reliability of the method in practical applications.

Acknowledgements. This study is supported by 2021's Cangzhou science and technology planning and self-financing project Research on Cangzhou Nutritional Food Informalization Evaluation under the ""Internet +"" Background. Cangzhou science and technology planning and self-financing project research result (213206018).

REFERENCES

[1] Anai, T. (2023). Improvement of soybean seed oil quality using fatty acid mutants. Nippon Shokuhin Kagaku Kogaku Kaishi, 70(1), 47-51.
[2] Bukowski, M. R., & Goslee, S. (2024). Climate-based variability in the essential fatty acid composition of soybean oil. The American journal of clinical nutrition., 119(1), 58-68.
[3] Song, H., Taylor, D. C., & Zhang, M. (2023). Bioengineering of soybean oil and its impact on agronomic traits. International Journal of Molecular Sciences, 24(3), 2256.
[4] Sim, J., Kuwabara, C., & Yamada, A. T. (2023). Recent advances in the improvement of soybean seed traits by genome editing. Plant Biotechnology, 40(3), 193-200.
[5] Johnson, J. B., Thani, P. R., & Naiker, M. (2023). Through?container detection of tea tree oil adulteration using near?infrared spectroscopy (nirs). Chemical papers, 77(4), 2009-2017.
[6] Palou, A., Patricia Jiménez, Casals, J., & Ingrid Masaló. (2023). Evaluation of the near infrared spectroscopy (nirs) to predict chemical composition inulva ohnoi. Journal of Applied Phycology, 35(5), 2007-2015.
[7] Deng, Z., Fu, J., Yang, M., Zhang, W., Y.-H., Y., & Zhang, L. (2024). Geographical origin identification of hainan camellia oil based on fatty acid composition and near infrared spectroscopy combined with chemometrics. Journal of Food Composition and Analysis, 125(1), 342-351.
[8] Zhe, L. I., Haigang, Q. I., Ying, Y. U., Liu, C., Cong, R., & Li, L. I., et al. (2023). Near-infrared spectroscopy method for rapid proximate quantitative analysis of nutrient composition in pacific oyster crassostrea gigas. Journal of Oceanology and Limnology(1), 150(3), 1-19.
[9] Tonolini, M., Wawrzynczyk, J., & Nielsen, M. E. S. B. (2023). On-line monitoring of enzymatic degumming of soybean oil using near-infrared spectroscopy. Applied Spectroscopy: Society for Applied Spectroscopy, 77(12), 1333-1343.
[10] Tonolini, M., Wawrzynczyk, J., & Nielsen, M. E. S. B. (2023). On-line monitoring of enzymatic degumming of soybean oil using near-infrared spectroscopy. Applied Spectroscopy: Society for Applied Spectroscopy, 77(12), 1333-1343.
[11] Liang, Y., Lin, H., & Kang, H. C. Q. (2023). Application of colorimetric sensor array coupled with machine-learning approaches for the discrimination of grains based on freshness. Journal of the Science of Food and Agriculture, 103(14), 6790-6799.
[12] Zhang, Q., Pizzi, A., Lei, H., Xi, X., Cao, M., & Cao, L. (2023). Maldi tof investigation of the reaction of soy protein isolate with glutaraldehyde for wood adhesives, 11(3), 12.
[13] Dai, C., Xu, X., & Huang, R. M. H. (2023). Monitoring of critical parameters in thermophilic solid-state fermentation process of soybean meal using nir spectroscopy and chemometrics. Journal of Food Measurement and Characterization, 17(1), 576-585.
[14] Bartalne-Berceli, M., Izso, E., & Gergely, S. S. A. (2022). Monitoring of soybean germination process by near-infrared spectroscopy. Acta Alimentaria: An International Journal of Food Science, 51(2), 204-217.
[15] Da Costa, G. B., De Fernandes, D. D. S., & Veras, GermanoDiniz, Paulo Henrique Goncalves DiasGondim, Amanda Duarte. (2024). Combining nir spectroscopy with dd-simca for authentication and ispa-pls-da for discrimination of ethyl route and oil feedstocks of biodiesels in biodiesel/diesel blends. Journal of the American Oil Chemists' Society, 101(2), 187-196.
[16] Kuang, J., Luo, N., Hao, Z., Xu, J., He, X., & Shi, J. (2022). Ni-raman spectroscopy combined with bp-adaboost neural network for adulteration detection of soybean oil in camellia oil. Journal of Food Measurement and Characterization, 16(4), 3208-3215.
[17] Liu, W., Sun, S., Xia, Y., Zhao, P., Liu, C., & Zheng, L. (2022). Rapid determination of benzo(a)pyrene concentration in soybean oil by terahertz transmission spectroscopy with chemometrics. Journal of Infrared, Millimeter, and Terahertz Waves, 43(7), 695-708.
[18] Santos, C. S., Clara, S., & Bagheri, M. S. M. W. (2024). Predicting iron deficiency and oxidative stress in glycine max through fourier transform infrared spectroscopy in a time-course experiment. Plant and Soil, 496(1/2), 161-177.
[19] Wang, J., He, J., Yang, J., & Gai, J. (2023). A novel procedure for identifying a hybrid qtl-allele system for hybrid-vigor improvement, with a case study in soybean (glycine max) yield. The Crop Journal, 11(1), 177-188.
[20] Petravicius, D., Binelo, M. O., & Binelo, Marcia de F. B.Faoro, Vanessada Silva, Jose A. G. (2023). Genetic algorithm in the design of soybean silos for airflow homogenization 1. revista brasileira de engenharia agricola e ambiental, 27(7), 531-538.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Aug 25, 2024
Accepted: Oct 5, 2024