

## Research Article

# Near-Infrared Spectroscopy Combined with Multivariate Calibration to Predict the Yield of Sesame Oil Produced by Traditional Aqueous Extraction Process

Yong-Dong Xu,<sup>1</sup> Yan-Ping Zhou,<sup>2</sup> and Jing Chen<sup>3</sup>

<sup>1</sup>College of Educational Science, Tongren University, Tongren, Guizhou 554300, China

<sup>2</sup>Key Laboratory of Pesticide and Chemical Biology of Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, China

<sup>3</sup>College of Material and Chemical Engineering, Tongren University, Tongren, Guizhou 554300, China

Correspondence should be addressed to Yong-Dong Xu; 2194694058@qq.com

Received 26 March 2017; Accepted 28 May 2017; Published 5 July 2017

Academic Editor: Hai-Yan Fu

Copyright © 2017 Yong-Dong Xu 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.

Sesame oil produced by the traditional aqueous extraction process (TAEP) has been recognized by its pleasant flavor and high nutrition value. This paper developed a rapid and nondestructive method to predict the sesame oil yield by TAEP using near-infrared (NIR) spectroscopy. A collection of 145 sesame seed samples was measured by NIR spectroscopy and the relationship between the TAEP oil yield and the spectra was modeled by least-squares support vector machine (LS-SVM). Smoothing, taking second derivatives (D2), and standard normal variate (SNV) transformation were performed to remove the unwanted variations in the raw spectra. The results indicated that D2-LS-SVM (4000–9000  $\text{cm}^{-1}$ ) obtained the most accurate calibration model with root mean square error of prediction (RMSEP) of 1.15 (% w/w). Moreover, the RMSEP was not significantly influenced by different initial values of LS-SVM parameters. The calibration model could be helpful to search for sesame seeds with higher TAEP oil yields.

## 1. Introduction

Sesame (*Sesamum indicum* L., Pedaliaceae family) has been one of the main oil crops in China and other Asian countries for many years [1]. Globally, sesame is cultivated in tropical and subtropical regions over 7.54 million hectares and its annual seed yield has exceeded 3 million tons [2]. As the largest producer of sesame, China accounts for approximately a quarter of the world's total production [2]. Sesame has been widely used as a flavor ingredient in sweets and desserts. As an important source for human nutrition, sesame seeds contain oil (44–58%), protein (18–25%), carbohydrate (<13.5%), and ash (<5%) [3]. By experimental investigations, sesame seeds have demonstrated various healthy and bioactive properties, including hypocholesterolaemic, hepatoprotective, antimutagenic, antiproliferative, antihypertensive, anti-inflammatory, and anticarcinogenic effects [4–6].

Sesame oil is rich in oleic, linoleic, palmitic, and stearic acids and also has low contents of palmitoleic, linolenic, and

eicosenoic acids [4–6]. Recently, sesame oil has attracted much attention due to its reported oxidative stability as well as its role in the prevention of hypertension and reducing the incidence of certain cancers [7]. In China, sesame oil is very popular in cooking and flavoring due to its unique and pleasant aroma and has been recognized as a top-grade vegetable oil.

Sesame oil can be extracted by various procedures, depending on the materials and equipment available. Mechanical pressing or/and chemical solvent extraction have been the commonly used techniques for large-scale production of sesame oil. Although mechanical pressing or/and chemical solvent extraction can generally obtain a high oil yield (80–99%, w/w), this is at the expense of a degraded oil quality [8]. During the mechanical pressing, heat treatment and the high temperature during expressing not only reduce the oil quality but also denature the proteins in the meal which is often served as a protein source for animals. Moreover, the

chemical solvent tends to coextract undesired components from the cell walls of seeds.

In China, a traditional aqueous extraction process (TAEP) for sesame oil has been used since ancient times [9]. In TAEP, the dried and roasted sesame seeds are ground by a small-scale miller under low-temperature conditions (60~65°C) and then pure water is added to replace the oil from the sesame sauce. Compared with mechanical pressing or/and solvent extraction, TAEP avoids serious heat damage to the oil and proteins and does not require the chemical refining steps; therefore, the oil can retain its nutritive components and natural flavor as far as possible. Nevertheless, low oil yield is the main challenge to TAEP, which is responsible for the high price of the TAEP oil. Besides the extraction conditions, the TAEP oil yield is mainly influenced by the varieties and quality of sesame seeds. Some investigations have been devoted to predictions of the oil contents in sesame seeds [9]. However, few researches have been reported on modeling the relationship between the seed quality and the oil yield by the TAEP. Currently, the oil yield can just be roughly estimated by the producers based on their experience and survey of the seeds. Therefore, a rapid and effective method to search for high-yield sesame seeds will be profitable to this traditional industry.

This work was aimed at developing a multivariate calibration model to predict the TAEP oil yield of sesame seeds by combining the nondestructive near-infrared (NIR) spectroscopy technique and chemometrics. A wide range of sesame seeds were collected from different producing areas of China to ensure the generalization performance of the model. In order to improve the prediction accuracy, different data preprocessing techniques were investigated to reduce unwanted spectral variations. The least-squares support vector machine (LS-SVM) [10, 11] was used to develop the nonlinear relationship between the TAEP oil yield and the measured NIR spectra.

## 2. Materials and Methods

**2.1. Sesame Seeds.** A set of 145 sesame seed samples of different producing areas of China were collected from domestic markets, including Henan (22), Jiangxi (15), Anhui (16), Hubei (16), Hebei (20), Jiangsu (12), Shandong (21), Shanxi (13), and Zhejiang (10). All of the sesame seeds were fresh seeds harvested in 2014 and the colors of seeds include off-white, buff, tan, brown, and gray. Black sesame seeds were excluded from this study because they usually have a higher price and are rarely used for oil extraction. All the sesame seeds were dried in the sun and stored in a cool and dark area before NIR measurement.

**2.2. NIR Analysis.** The NIR diffuse reflectance spectra of sesame seed samples were measured in the wavelength range of 4000–12,000  $\text{cm}^{-1}$  on a Bruker-TENSOR37 FTIR spectrometer (Bruker Optics, Ettlingen, Germany) using OPUS software. Dried sesame seeds were analyzed in a quartz cuvette with a PbS detector and an internal gold background as the reference. Each sample was scanned three times at

room temperature with being stirred before each scanning and the average spectrum was used. For each measurement, 64 scans were performed and more scans did not improve the signal quality significantly. The resolution was  $4 \text{ cm}^{-1}$ , and the scanning interval was  $1.929 \text{ cm}^{-1}$ . Therefore, each raw spectrum had 4148 individual wavelengths for multivariate analysis. The temperature was kept around  $25^\circ\text{C}$  and the humidity was kept at a steady level during NIR analysis. The sequence of NIR measurements was randomly designed for all the samples.

**2.3. TAEP.** After NIR spectroscopy analysis, all the seed samples were randomly coded and extracted in an oil mill using the TAEP method. The steps of TAEP were as follows: (1) the sesame seeds were sieved, rinsed, and roasted ( $150^\circ\text{C}$  for 40 min and then  $200^\circ\text{C}$  for 20 min); (2) the roasted seeds were then ground using a stone mill to produce sesame slurry; and (3) water was added to the slurry, which was then stirred for 30 min and left to stand for 1 h to separate the oil and water. The oil yield by TAEP was defined as

$$\text{yield}_{\text{TAEP}} = \frac{A_{\text{oil}}}{A_{\text{seed}}} \times 100\%, \quad (1)$$

where  $\text{yield}_{\text{TAEP}}(\%)$  is the oil yield,  $A_{\text{oil}}(\text{g})$  is the net weight of oil obtained by TAEP, and  $A_{\text{seed}}(\text{g})$  is the net weight of sesame seed.

**2.4. Data Preprocessing and Splitting.** The data analysis was performed on MATLAB 7.10.0 (R2010a) (MathWorks, Sherborn, MA). To remove the unwanted spectral variations in the measured NIR data, smoothing, taking second-order derivatives (D2) [12], and standard normal variate (SNV) transformation [13] were performed. In order to compare the performances of different models and data preprocessing methods, all the models were developed using the same training and test sets. The DUPLEX algorithm [14] was performed on the raw NIR data to divide the total objects into representative training and test sets. DUPLEX works as follows: (1) the two objects having the largest Euclidean distance were selected from the objects pool as two training objects; (2) the two objects with the largest distance among the remaining objects were put in the test set; (3) repeat (1) and (2) until one has sufficient testing objects; and (4) put all the remaining objects in the training set. The training and test sets obtained by DUPLEX are uniformly distributed in the entire experimental space.

**2.5. Least-Squares Support Vector Machines (LS-SVMs).** The multivariate calibration model between the measured NIR spectra and the TAEP oil yield was developed using LS-SVM. LS-SVM is a simplified version of support vector machines (SVMs) [15]. LS-SVM uses equality type of constraints instead of the quadratic programming as in the ordinary SVMs and has a much faster computation speed. In this study, the Gaussian radial basis function (RBF) was used as a nonlinear kernel transformation in LS-SVM. LS-SVM has two parameters,  $\gamma$  and  $\sigma$ , to be optimized. The kernel width,  $\gamma$ , can be adjusted to control the nonlinear nature of the RBF. The

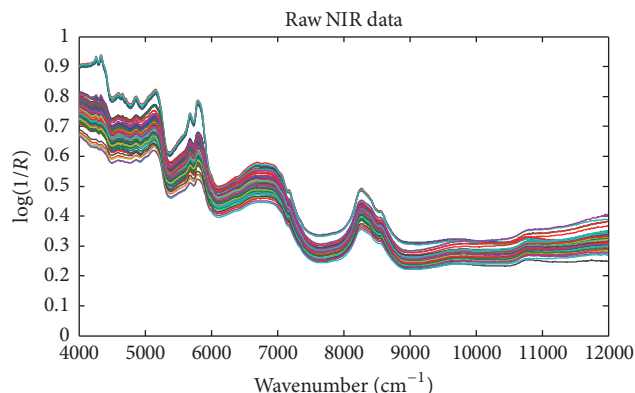


FIGURE 1: The raw NIR spectra of the 145 sesame seed samples.

regularization parameter,  $\sigma$ , controls the tradeoff between minimization of model structural risk and the learning error. The parameters of LS-SVM models ( $\gamma$  and  $\sigma$ ) were optimized by minimizing the prediction errors of leave-one-out cross validation (LOOCV) using the Simplex optimization. The tuning, training, and prediction of LS-SVM models were performed using the LS-SVMLab v1.8 MATLAB toolbox [10, 11].

### 3. Results and Discussion

The raw NIR spectra of the 145 sesame seed samples are shown in Figure 1. The raw spectra in the spectral interval of 9000–12000  $\text{cm}^{-1}$  carried little chemical information and were influenced by baseline shifts, so only the spectral range of 4000–9000  $\text{cm}^{-1}$  was used for developing multivariate calibration models. Due to band overlapping, it is difficult to accurately assign the peaks to specific chemical components. The peak around 4300  $\text{cm}^{-1}$  is the combination absorbance of C–H stretching and the deformation of  $\text{CH}_2$ ; and those at 4500–5000  $\text{cm}^{-1}$  can be attributed to the overlapping of combination absorbances of C=O stretching and the deformation of –CO–NH– and the combination absorbance of N–H stretching and the deformation of –CO–NH–. Other peak assignments are as follows: 5200  $\text{cm}^{-1}$  (combination of O–H stretching and H–O–H deformation), 5500–5800  $\text{cm}^{-1}$  (the second overtone C–H stretching in  $\text{CH}_2$ ), 6000–7000  $\text{cm}^{-1}$  (the second overtone of N–H stretching in amides and aromatic amines), 8000–8800  $\text{cm}^{-1}$  (the second overtones of C–H stretching in various groups), and 10700  $\text{cm}^{-1}$  (the third overtones of C–H stretching in various groups).

The NIR spectra preprocessed by smoothing, taking D2, and SNV transformation are demonstrated in Figure 2. Both smoothing and taking D2 were performed using the S-G polynomial fitting algorithm. The order of polynomial was two and the window size for smoothing and taking D2 was 11 and 19 wavenumbers, respectively. Comparison of the smoothed and the raw spectra indicates few differences, which can be attributed to the high SNR of the raw data and

TABLE 1: The training and prediction results of LS-SVM with different data preprocessing methods.

Preprocessing	Optimized parameters $\ln(\gamma)$ , $\ln(\sigma^2)$	RMSEC	RMSEP
Raw data	3.9190, 8.2156	2.18	2.80
Smoothing	4.2536, 8.3497	2.15	2.72
D2	17.5932, 15.0775	1.09	1.15
SNV	1.1017, 6.8987	1.79	2.26

the fact that little improvements of the spectral resolution can be obtained by enhancing SNR. Taking D2 of the raw spectra could remove most of the baseline shifts and a much higher peak resolution was obtained without reducing the SNR significantly. SNV transformation could remove some baselines but the baseline shifts along the wavelengths were still retained in the SNV-preprocessed data. The actual influence of data preprocessing should be evaluated by comparing the performance of calibration models.

The DUEPLX algorithm was then performed on the raw NIR spectra of the 145 sesame seed samples to obtain a training set of 105 objects and a test set of 40 objects. To compare the influences of different data preprocessing methods on calibration models, all the LS-SVM models are trained and validated using the same training and test sets. The parameters ( $\gamma$  and  $\sigma$ ) of LS-SVM models were simultaneously optimized using the Simplex method to obtain the minimized mean squared errors (MSE) in leave-one-out cross validation (LOOCV). With different preprocessing methods, the training and prediction results by LS-SVM were demonstrated in Table 1.

As seen from Table 1, preprocessing can generally reduce the training and prediction errors of LS-SVM in terms of root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP). Compared with the models based on the raw and smoothed spectra, models with D2 and SNV obtained significantly lower errors. The best model was obtained with D2-LS-SVM with RMSEC of 1.09 and RMSEP of 1.15. D2 preprocessing had superior performance than SNV, indicating linear baseline may exist along the wavelengths in the raw data, which can be removed by D2 but can only be partially removed by SNV. The correlation plot between the actual and predicted TAEP sesame oil yield by D2-LS-SVM is shown in Figure 3, indicating the prediction results were well consistent with the reference values. In order to investigate the stability of parameter optimization of LS-SVM, the RMSEP values of 500 repeated model developments with random data splitting were computed and the box and whisker plots are demonstrated in Figure 4. The results indicate that, with smoothing, taking D2, and SNV transformation, the uncertainty in model optimization was reduced and its influence on prediction was insignificant.

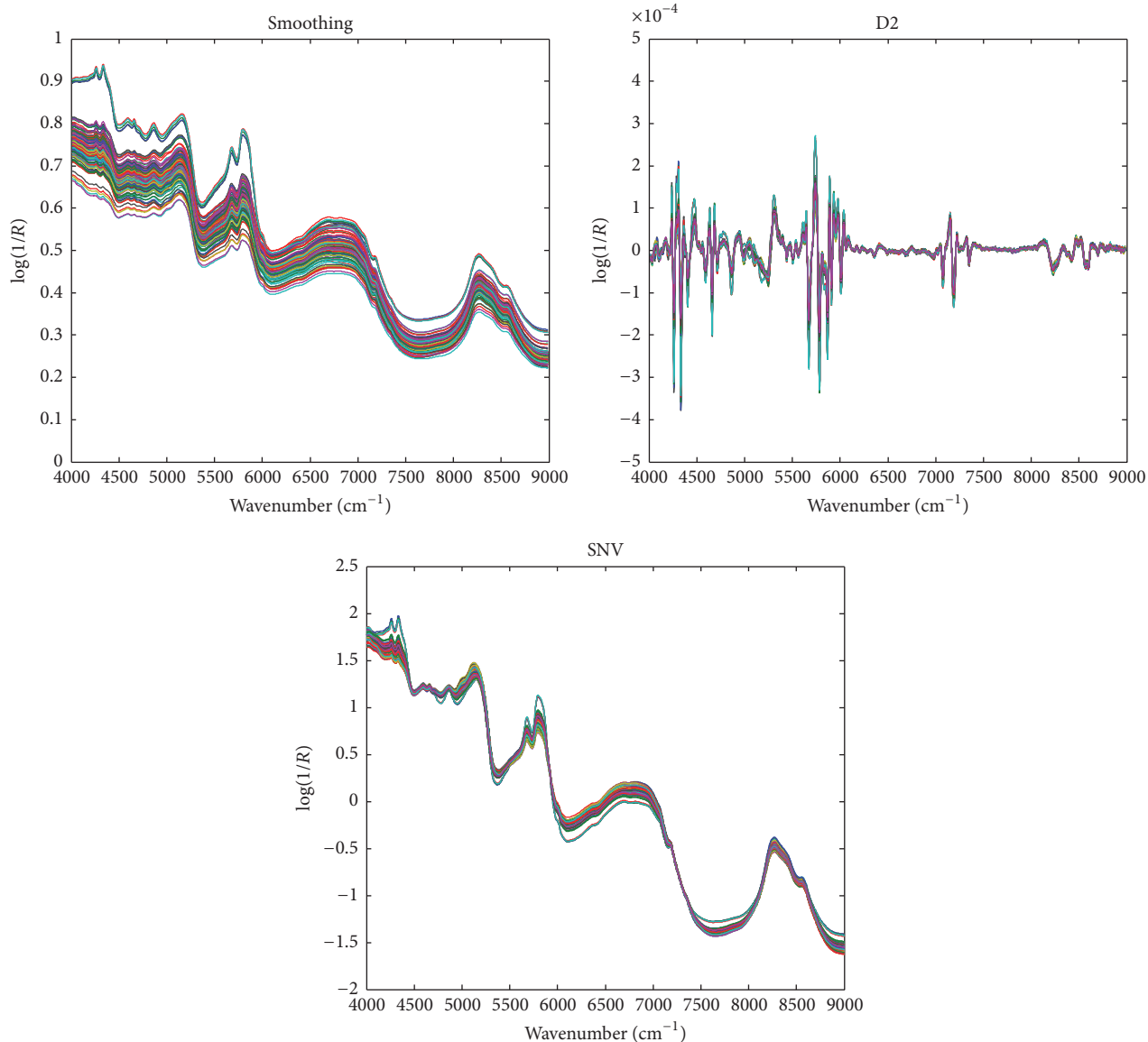


FIGURE 2: The preprocessed spectra of sesame seeds by smoothing, taking D2, and SNV transformation.

#### 4. Conclusions

Rapid and nondestructive analysis methods for the oil yield of sesame oil samples for TAEP were developed by FTIR spectroscopy and chemometric multivariate calibration. The effects of different data preprocessing on model accuracy were investigated. The analysis results indicate spectra preprocessing by taking D2 and SNV can obtain accurate and stable calibration models. RMSEC of 1.09 and RMSEP of 1.15 were obtained by D2-LS-SVM. This work demonstrates that FTIR spectroscopy combined with LS-SVM provides an accurate and practical method to predict the sesame oil yield for TAEP. It is useful to search for high-quality sesame seeds TAEP of sesame oil. Our further study will be focused on the relationship between the NIR spectra of sesame seeds and the quality or/and flavor of TAEP oil.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

The authors are grateful for the financial support from the National Natural Science Foundation of China (Grants nos. 21665022, 21665002, and 21675059). The authors appreciate the financial support from the Open Research Program (no. GCTKF2014007) of State Key Laboratory Breeding Base of Green Chemistry Synthesis Technology (Zhejiang University of Technology), the Research Fund for the Doctoral Program of Tongren University (no. trxyDH1501), the Open Research Program (no. 2015ZY006) from the Modernization

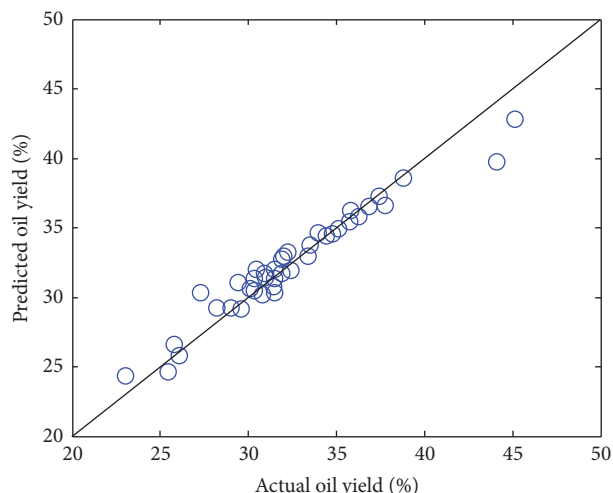


FIGURE 3: Correlation plot between the actual and predicted TAFP sesame oil yield by D2-LS-SVM.

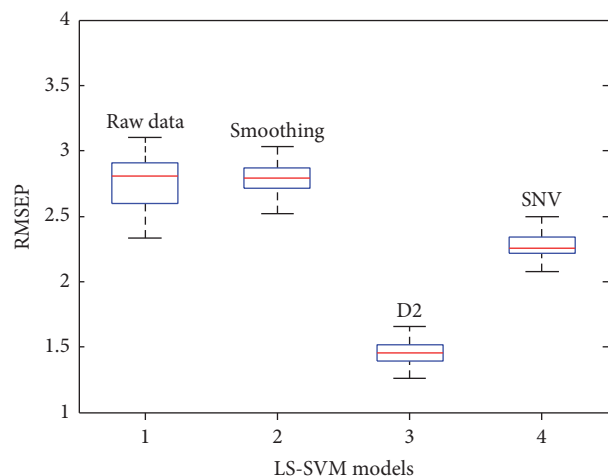


FIGURE 4: Box and whisker plots of RMSEP obtained for 500 repeated parameter optimizations for LS-SVM models with different data preprocessing methods. Each plot indicates the minimum, lower quartile, median, upper quartile, minimum, and maximum of RMSEP.

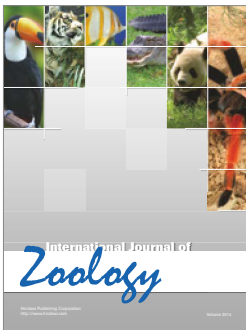
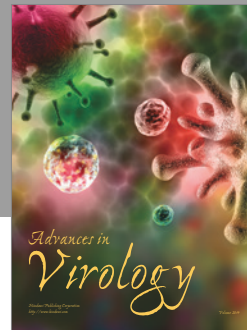
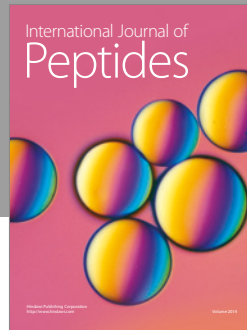
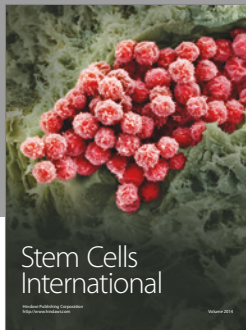
Engineering Technology Research Center of Ethnic Minority Medicine of Hubei Province (South-Central University for Nationalities), and the research funds from the Education Department of Guizhou Province (no. QJHKYZ[2015]498).

## References

- [1] H. A. Abou-Gharbia, A. A. Y. Shehata, and F. Shahidi, "Effect of processing on oxidative stability and lipid classes of sesame oil," *Food Research International*, vol. 33, no. 5, pp. 331–340, 2000.
- [2] S. Latif and F. Anwar, "Aqueous enzymatic sesame oil and protein extraction," *Food Chemistry*, vol. 125, no. 2, pp. 679–684, 2011.
- [3] B. A. Were, A. O. Onkware, S. Gudu, M. Welander, and A. S. Carlsson, "Seed oil content and fatty acid composition in East African sesame (*Sesamum indicum* L.) accessions evaluated

over 3 years," *Field Crops Research*, vol. 97, no. 2-3, pp. 254–260, 2006.

- [4] M. Elleuch, S. Besbes, O. Roiseux, C. Blecker, and H. Attia, "Quality characteristics of sesame seeds and by-products," *Food Chemistry*, vol. 103, no. 2, pp. 641–650, 2007.
- [5] A. A. Dar and N. Arumugam, "Lignans of sesame: purification methods, biological activities and biosynthesis: a review," *Bioorganic Chemistry*, vol. 50, pp. 1–10, 2013.
- [6] M. Namiki, "Nutraceutical functions of sesame: a review," *Critical Reviews in Food Science and Nutrition*, vol. 47, no. 7, pp. 651–673, 2007.
- [7] J.-C. Zhou, D.-W. Feng, and G.-S. Zheng, "Extraction of sesamin from sesame oil using macroporous resin," *Journal of Food Engineering*, vol. 100, no. 2, pp. 289–293, 2010.
- [8] P. Willems, N. J. M. Kuipers, and A. B. De Haan, "Hydraulic pressing of oilseeds: Experimental determination and modeling of yield and pressing rates," *Journal of Food Engineering*, vol. 89, no. 1, pp. 8–16, 2008.
- [9] L. Hou, L. S. Xiao, X. Wang, J. Liu, and L. Zhou, "Studies on the aqueous extraction method of sesame oil," *Advanced Materials Research*, vol. 634-638, no. 1, pp. 1542–1547, 2013.
- [10] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [11] J. A. K. Suykens, J. de Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, pp. 85–105, 2002.
- [12] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [13] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.
- [14] R. D. Snee, "Validation of regression models: methods and examples," *Technometrics*, vol. 19, no. 4, pp. 415–428, 1977.
- [15] H. Li, Y. Liang, and Q. Xu, "Support vector machines and its applications in chemistry," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 2, pp. 188–198, 2009.



# Hindawi

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

