RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

One-class classification based authentication of peanut oils by fatty acid profiles

Liangxiao Zhang[1,4,5,6,*], Peiwu Li[1,3,4,5,*], Xiaoman Sun[1,5], Jin Mao[1,5], Fei Ma[1,2,5], Xiaoxia Ding[1,4,5], Qi Zhang[1,2,3]

[1] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China

[2] Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China

[3] Key laboratory of Detection for Mycotoxins, Ministry of Agriculture, Wuhan 430062, China

[4] Laboratory of Risk Assessment for Oilseeds Products (Wuhan), Ministry of Agriculture, Wuhan 430062, China

[5] Quality Inspection and Test Center for Oilseeds Products, Ministry of Agriculture, Wuhan 430062, China

[6] Hubei Collaborative Innovation Center for Green Transformation of Bio-Resources, Wuhan 430062, China

∗Corresponding authors at: Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China

Tel.: +86 27 86812943; Fax: +86 27 86812862.

E-mail addresses: liangxiao_zhang@hotmail.com (L. Zhang);

peiwuli@oilcrops.cn (P. Li)

The abbreviation of the title: One-class classification based authentication of peanut oils

Abstract

Developing a method of identifying oil authenticity becomes critical for protecting customers' rights as adulteration of edible oils is a particular concern in food quality. Since adulterants in edible oils are usually unknown, the authenticity identification technique is a one-class classification model in chemometrics. In this study, a one-class classification model was built to identify the authenticity of peanut oils by fatty acid profiles. Based on the previous studies, 28 fatty acids were identified and quantified for peanut oils. The authenticity identification model was built by the one-class partial least squares (OCPLS) classifier for peanut oils. Subsequently, the established model was validated by independent test sets. The results indicated that the OCPLS classifier could completely detect adulterated oils and were therefore employed for authenticity assessment. Moreover, counterfeit oils adulterated with different levels of other edible oils were simulated by the Monte Carlo method and employed to test the lowest adulteration level of this one-class classifier. As a result, the model could identify peanut oils and sensitively detect adulteration of edible oils with other vegetable oils at the adulteration level of 4%.

Keywords: One-class partial least squares (OCPLS); Fatty acid profiles; Authentication; Peanut oil; GC/MS; Chemometrics

1. Introduction

Peanut oil has a pleasant flavor and is non-transgenic, experiencing a gradual increase in the market share in China and other Asian countries[1, 2]. Similar with adulteration of olive oil in western countries, adulteration of peanut oil is also serious. Therefore, seeking a reliable method to identify the authenticity of peanut oil is greatly demanded. Recently, different targets have been employed to identify the authenticity of edible oils including: (a) genetic markers of adulterants[3]; (b) characteristic metabolites of adulterants[4]; (c) spectra of entire edible oils detected by nuclear magnetic resonance (NMR)[5], near-infrared spectroscopy (NIR)[6], infrared spectroscopy (IR)[7], fluorescence spectroscopy[8], Raman spectroscopy[9], electronic nose[10] and ion mobility spectrometry[11-12]; (d) metabolomics or metabolite profiles[13-16]. With explicit chemical significance and advantages of multivariate analysis, the authentication methods based on metabolomics become the most promising for edible oils[14, 17].

In the spectroscopy- and metabolomics-based authentication methods, chemometrics is a powerful tool to detect adulterated edible oils when used qualitatively for classifying unknown samples with similar characteristics and quantitatively for determining adulterant analytes in samples[18]. Recently, chemometric methods, such as self-organizing maps based on chaotic parameters, cluster discriminant analysis (CDA), support vector machine (SVM) and random forests (RF), were used to distinguish edible oils from refined recycled cooking oils, identify edible oils from different regions, or detect adulteration of extra virgin olive oil with inferior edible oils[19-22]. Generally, adulteration detection of edible oils is considered as a two- or multi-class classification method to determine whether the target edible oil is adulterated with known oil. However, since the adulterants in edible oils are usually unknown, the authenticity identification of the target edible oil might be the best choice, which is a technique of one-class classification in chemometrics. Therefore, the key technologies of authenticity

identification include stable markers or metabolite profiles and the effective one-class classification method. Among the four kinds of targets, genetic markers and some metabolite profiles are relatively stable. Since genetic markers of adulterants significantly decrease after refinement, they are usually hard to detect in adulterated edible oils. Fatty acids are the dominant components of edible oils, and their composition is relatively stable in the oilseeds from different producing areas and edible oils produced by different processing methods. Though these low-abundance fatty acids are not so attractive to nutritionists, our previous study indicates that they are highly sensitive in adulteration identification of edible oils[14]. The fatty acid profiles are therefore taken as a key marker and quality parameter of different oilseeds and their products[17].

Lately, partial least squares density modeling (PLS-DM) was proposed and employed in the identification of adulterated peanut oils by mid-infrared spectroscopy (MIR)[23] and authentication of olives in brine by NIR[24], which could achieve a better balance between sensitivity and specificity than the typical SIMCA method[24]. However, the IR and NIR spectra of edible oils reflect the whole chemical fingerprint instead of the detailed chemical compositions, and an effective one-class classification model therefore depends on a large number of training samples in practice. If not, many authentic edible oils out of the training set might be mistaken as adulterated edible oils, which is unacceptable in actual applications of quality supervision and inspection. Therefore, in this study, an authenticity identification model was built by the one-class partial least squares (OCPLS) classifier for peanut oil based on fatty acid profiles. Moreover, adulterated oils were simulated by adulterating peanut oil with different levels of other oils or mixtures thereof and employed to test the sensitivity of the one-class classifier.

2. Experimental

2.1 Materials and reagents

To ensure that the oil samples could represent the actual edible oils, 80 peanut samples were collected from different production areas (see Supporting material Table S1) and employed to prepare edible oils by TEN GUARD oil mill machinery (TZC-0502, China). Supelco 37 component fatty acid methyl esters (FAME) mix (No. 47885-U) was purchased from Sigma (St. Louis, MO, USA). 11-octadecenoic acid (C18:1n-7, > 97.0 purity) and 7-hexadecenoic acid methyl ester were also purchased from Sigma (St. Louis, MO, USA).

2.2 Fatty acid analysis

As descried in the previous study[14, 25], fatty acids in edible oils were first derived to produce FAMEs and subsequently analyzed by GC-MS in selected ion monitoring (SIM) mode. The detailed procedure of fatty acid analysis was described in the Supporting materials.

Identification of fatty acids in SIM mode was conducted according to the protocol in our previous study[25]. The fatty acid percentage composition (percentage of the peak area) of edible oils was employed as quantitative results.

2.3 Multivariate analysis

Data matrix includes the relative contents of fatty acids in edible oils. Since the chemical properties of fatty acids are relatively stable, the weight of fatty acid composition in blended oil is equal to the sum of the weights of fatty acid composition in individual oils. To establish a more accurate adulteration detection model, adulterated oils were simulated by the Monte Carlo method. In this study, the adulterated oils were simulated by adulterating with low proportions of one or more vegetable oils[14]. For example, we could obtain an adulterated peanut oil by adding 5% of soybean, sesame, sunflower and rapeseed with random proportions to 95% of peanut oil.

A one-class partial least squares (OCPLS) classifier was proposed based on partial least squares (PLS) using a distance-based sample density measurement as the response variable[26-29]. In OCPLS, the potential function probability density is calculated on PLS scores and residual Q statistics to develop an efficient one-class classifier. The detailed algorithm of OCPLS modeling is described elsewhere[24, 27-29]. Initially, a PLS model was developed by analytical data and response vectors with all elements being 1 using the SIMPLS algorithm[30]. The number of PLS components was estimated by cross validation (CV). Then, the PLS scores on the first several latent variables (LVs) were used to estimate the PFM probability density of the class to obtain the critical value. In addition, the PLS residuals were used to compute the critical value of Q statistics. The predicted residual sum of squares (PRESS) obtained by Monte Carlo cross validation (MCCV) or leave-one-out cross validation (LOOCV) can be used to estimate the number of significant LVs. Finally, the score distance (SD) of an object in the space spanned by the primary OCPLS components and the absolute centered residual (ACR) were calculated and plotted to screen for outliers. OCPLS outlier diagnosis depended on the ACR of response variables and the OCPLS scores of primary LVs. Four regions at the bottom left, bottom right, top left and top right denote the regular points, good leverage points, class outliers and bad leverage points, respectively. The samples falling in the top left and top right regions were detected as adulterated oils. In this study, the Gaussian radial basis function (GRBF) transformation was placed in the position of the training objects, and the number of RBFs equaled the number of training objects to conduct nonlinear GRBF-OCPLS[28].

The programs of Monte Carlo simulation of adulterated oils were coded in Matlab 2011a for Windows (The Mathworks, Natick, MA), and the programs of OCPLS were kindly provided by Dr. Xu[29].

3. Results and discussion

3.1 Fatty acids profiles of peanut oils

In this study, fatty acid profiles of peanut oils were obtained by GC/MS. To build robust authentication modeling, the SIM mode was used to detect more fatty acids in edible oils. FAMEs were identified by combining the selective ions with retention indices based on equivalent chain lengths (ECL)[24]. As a result, 28 fatty acids were identified and quantified (see Supporting material Table S2), which were significantly more than in full scan mode (23 fatty acids)[22]. Fatty acid profiles of peanut oils were described by the percentage content and used in subsequent multivariate analysis.

3.2. One-class partial least squares (OCPLS) classifier

After determination and quantification of fatty acids in 80 peanut samples, the data matrix of relative contents was used in one-class classification modeling. The Kennard and Stone (K-S) algorithm[31-32] was employed to divide authentic edible oil samples into a training set and a test set at a ratio of 6:4. The OCPLS model was developed based on the fatty acid profiles of peanut oils. MCCV with a sampling ratio of 0.8 was conducted to select significant OCPLS components and estimate the standard deviation of prediction residuals. The number of significant components was estimated by examining the PRESS values obtained by MCCV according to the previous study[28-29].

The OCPLS model was built using 80 peanut oils. According to MCCV with a sampling ratio of 0.8 (Figure 1a), when the number of significant OCPLS components was set to 8 ($\delta^2 = 10$) , the lowest standard deviation of residual of cross validation was obtained. According to the OCPLS score distance

(SD) and ACR of the predicted responses, four regions at the bottom left, bottom right, top left and top right denote the regular points, good leverage points, class outliers and bad leverage points, respectively. Generally, the samples at the bottom left, bottom right regions were identified as pure peanut oils, while the samples at the top left and top right regions were identified as impure peanut oils or other oil. The prediction results of the samples in training and test sets by OCPLS with the fatty profiles of pure peanut oils are shown in Figure 1b, 1c and Table 1. The samples in the training (Figure 1b) and test sets (Figure 1c) are predicted as regular points with a small SD and a small ACR or good leverage points with a large SD and a small ACR (3 samples in the training set). Meanwhile, the reference edible oils from our previous study[14] including 62 sunflower seed oils, 63 rapeseed oils and 80 sesame oils were also used to test this model. The results in Figure 1d show that this model could effectively predict these three kinds of edible oils as bad leverage points. The validation by the independent test sets of peanut oils and other three kinds of edible oils indicated that the OCPLS model with fatty acid profiles could effectively determine the authenticity of peanut oils.

(Figure 1)

(Table 1)

3.3. Adulteration detection of edible oils

In the OCPLS model for peanut oils, we found that the one-class model could completely differentiate peanut oil from other kinds of edible oils. However, the adulteration of edible oils with other cheap edible or inedible oils is more common in practice. Therefore, it is essential to detect these adulterated edible oils. Generally, adulteration detection of edible oils is taken as a two- or multi-class classification method to classify authentic edible oils from fake oils adulterated with one or more known oils. However, since the adulterants in edible oils are usually unknown, the adulteration

detection falls into the one-class classification field in chemometrics. Meanwhile, the combination of multiple adulterants is usually ignored in the establishment and validation of adulteration detection methods. Recently, based on the fact that fatty acids are relatively stable and no chemical reaction happens in physical blends of vegetable oils, the Monte Carlo simulation of adulterated oils was proposed to test the adulteration detection model by blending a fixed proportion (Q%) of other adulterant oils with random proportions to (1 - Q%) of the current edible oils[14].

(Figure 2)

In this study, counterfeit oils with different adulteration levels (3%-12%, mol/mol) were simulated to check the lowest adulteration level of this model. As results, 40 adulterated oils were obtained at each adulteration level. At the beginning, 80 peanut oils were used to build the OCPLS model. Then, the simulated adulterated oils were predicted by this model. As shown in Figure 2a, the OCPLS model could correctly identify all of 80 authentic peanut oils in the training set. From Figure 2b and Table 1, we just misidentified three adulterated peanut oils as authentic peanut oils at the adulteration level of 3%, indicating that the accuracy rate of this model equals 92.5% (37/40) for the adulterated peanut oils with the adulteration level of 3%. When the adulteration level is higher than 3%, this OCPLS model could completely detect these adulterated peanut oils. Thus, the lowest adulteration level of this OCPLS model is set to 4%.

The results in this work indicated that the OCPLS model with fatty acid profiles could be a good strategy to identify authenticity of peanut oils and other edible oils. Compared with the previous authentication methods, this method possesses the following advantages: (1) It can detect counterfeit oils adulterated with any unknown oils; (2) it can also detect the multivariate adulteration.

4. Conclusion

In this study, a one-class classification model was built to identify the authenticity of peanut oils by fatty acid profiles. The authentication identification model was built by the one-class partial least squares (OCPLS) classifier for peanut oils. Subsequently, the established model was validated by independent test sets. The results indicated that the OCPLS classifier could completely detect adulterated oils and were therefore employed for authenticity assessment. Moreover, the oils adulterated with different levels of other edible oils were simulated by the Monte Carlo method and employed to test the lowest adulteration level of this one-class classifier. Compared with the studies in Ref. 14, the OCPLS model for peanut oil is more robust in detecting all kinds of adulteration including known adulterants and multiple adulterants. As a result, this model could identify peanut oils and sensitively detect adulteration of edible oils with other vegetable oils at the adulteration level of 4%.

Acknowledgements

This work was supported by the Project of National Science & Technology Pillar Plan (2012BAK08B03), the Special Fund for "Quality Inspection" Research in the Public Interest (2012104010-4), the National Major Project for Agro-product Quality & Safety Risk Assessment (GJFP2015006), the National Natural Science Foundation of China (21205118), and the earmarked fund for China Agriculture research system (CARS-13).

Conflict of interest

No authors declared any potential conflicts of interest.

References

[1] J. K. Kim, H.J. Lim, D.H. Shin, E.C. Shin, Comparison of nutritional quality and thermal stability between peanut oil and common frying oils. J. Korean Soc. Appl. Biol. Chem. 58 (2015) 527-532.

[2] F. Zhao, J. Liu, X. Wang, P. Li, W. Zhang, Q. Zhang, Detection of adulteration of sesame and peanut oils via volatiles by GC×GC–TOF/MS coupled with principal components analysis and cluster analysis. Eu. J. Lipid Sci. Tech. 115 (2013) 337-347.

[3] H. Zhang, Y. Wu, Y. Li, B. Wang, J. Han, X. Ju, Y. Chen, PCR-CE-SSCP used to authenticate edible oils. Food Control 27 (2012) 322-329.

[4] X. Zhao, F. Ma, P.W. Li, G.M. Li, L.X. Zhang, Q. Zhang, W. Zhang, X.Q. Wang, Simultaneous determination of isoflavones and resveratrols for adulteration detection of soybean and peanut oils by mixed-mode SPE LC-MS/MS, Food Chem. 176 (2015) 465-471.

[5] P. Daisa, E. Hatzakisb, Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review, Anal. Chim. Acta 765 (2013) 1-27.

[6] O. Galtier, N. Dupuy, Y. Le Dréau, D. Ollivier, C. Pinatel, J. Kister, J. Artaud, Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra, Anal. Chim. Acta 595 (2007) 136-144.

[7] L. Xu, C.B. Cai, D.H. Deng, Multivariate quality control solved by one-class partial least squares regression: identification of adulterated peanut oils by mid-infrared spectroscopy, J. Chemometrics 25 (2011) 568-574.

[8] F. Ge, C.Y. Chen, D.Q. Liu, S.L. Zhao, Rapid Quantitative Determination of Walnut Oil Adulteration with Sunflower Oil Using Fluorescence Spectroscopy, Food Anal. Methods 7 (2014) 146-150.

[9] D. Wei, Y. Q. Zhang, B. Zhang, X. P. Wang, Rapid prediction of fatty acid composition of vegetable oil by Raman spectroscopy coupled with least squares support vector machines, J Raman Spectrosc. 44 (2013) 1739-1745.

[10] A. Guadarrama, M.L. Rodríguez-Méndez, C. Sanz, J.L. Ríos, J.A. de Saja, Electronic nose based on

conducting polymers for the quality control of the olive oil aroma Discrimination of quality, variety of olive and geographic origin, Anal. Chim. Acta 432 (2001) 283-292.

[11] L.X. Zhang, Q. Shuai, P.W. Li, Q. Zhang, F. Ma, W. Zhang, X.X. Ding, Xiaoxia Ding, Ion mobility spectrometry fingerprints: A rapid detection technology for adulteration of sesame oil. Food Chem. 192 (2016) 60-66.

[12] Q. Shuai, L.X. Zhang, P.W. Li, Q. Zhang, X.P. Wang, X.X. Ding, W. Zhang, Rapid adulteration detection for flaxseed oil using ion mobility spectrometry and chemometric methods. Anal. Methods 6 (2014) 9575-9580.

[13] T. Řezanka, H. Řezanková, Characterization of fatty acids and triacylglycerols in vegetable oils by gas chromatography and statistical analysis, Anal. Chim. Acta 398 (1999) 253-261.

[14] L.X. Zhang, P.W. Li, X.M. Sun, X.F. Wang, B.C. Xu, W.P. Wang, F. Ma, Q. Zhang, X.X. Ding, Classification and adulteration detection of vegetable oils based on fatty acid profiles. J. Agr. Food Chem. 62 (2014) 8745-8751.

[15] B.C. Xu, L.X. Zhang, H. Wang, D.L. Luo, P.W. Li, Characterization and authentication of four important edible oils using free phytosterol profiles established by GC-GC-TOF/MS. Anal. Methods 6 (2014) 6860-6870.

[16] W. Hu, L. X. Zhang, P. W. Li, X. P. Wang, Q. Zhang, B.C. Xu, X.M. Sun, F. Ma, X.X. Ding, Characterization of volatile components in four vegetable oils by head space two-dimensional comprehensive chromatography time-of-flight mass spectrometry, Talanta129 (2014) 629-635.

[17] L.X. Zhang, X.Y. Ji, B.B. Tan, Y.Z. Liang, N.N. Liang, X.L. Wang, H. Dai, Identification of the composition of fatty acids in Eucommia ulmoides seed oil by fraction chain length and mass spectrometry, Food Chem.121 (2010) 815-819.

[18] J.C. Moore, M. Lipp, J.C. Griffiths, Preventing the adulteration of food protein. Food Technol. 62 (2011) 46-50.

[19] F. Mümtaz, H. Dıraman, D. Özdemir, Classification of Turkish Monocultivar (Ayvalık and Memecik cv.) Virgin Olive Oils from north and south zones of Aegean region based on their triacyglycerol profiles. J. Am. Oil Chem. Soc. 90 (2013) 1661-1671.

[20] T.B. Liu, L.J. Zhou, Z.W. Chen, B.B. Li, Y. Shi, Authentication of edible vegetable oil and refined recycled cooking oil using a Micro-UV spectrophotometer based on chemometrics. J. Am. Oil Chem. Soc. 90 (2013) 1599-1606.

[21] J.S. Torrecilla, J.C. Cancilla, G. Matute, P. Díaz-Rodríguez, A.I. Flores, Self-organizing maps based on chaotic parameters to detect adulterations of extra virgin olive oil with inferior edible oils. J. Food Eng. 118 (2013) 400-405.

[22] F.F. Ai, J. Bin, Z.M. Zhang, J.H. Huang, J.B. Wang, Y. Z. Liang, L. Yu, Z.Y. Yang, Application of random forests to select premium quality vegetable oils by their fatty acid composition, Food Chem. 143 (2014) 472-478.

[23] L. Xu, C.B. Cai, D. H. Deng, identification of adulterated peanut oils by mid-infrared spectroscopy, J. Chemometrics 25 (2011) 568-574.

[24] P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, S. Lanteri, Partial least squares density modeling (PLS-DM)-A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy, Anal. Chim. Acta 851 (2014) 30-36.

[25] L.X. Zhang, P.W. Li, X.M. Sun, W. Hu, X.P. Wang, Q. Zhang, X.X. Ding, Untargeted fatty acid profiles based on the selected ion monitoring mode. Anal. Chim. Acta 839 (2010) 44-50.

[26] L. Xu, H.Y. Fu, N. Jiang, X.P. Yu, A new class model based on partial least square regression and its applications for identifying authenticity of bezoar samples. Chin. J. Anal. Chem. 38 (2010) 175-180.

[27] L. Xu, C.B. Cai, D. H. Deng, identification of adulterated peanut oils by mid-infrared spectroscopy, J. Chemometrics 25 (2011) 568-574.

[28] L. Xu, S.M. Yan, C.B. Cai, X.P. Yu, One-class partial least squares (OCPLS) classifier, Chemometr. Intell. Lab. 126 (2013) 1-5.

[29] L. Xu, M. Goodarzi, W. Shi, C.B. Cai, J.H. Jiang, A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers, Chemometr. Intell. Lab. 139 (2014) 58-63.

[30] S. de Jong, SIMPLS: An alternative approach to partial least squares regression, Chemom. Intell. Lab. Syst. 18 (1993) 251-263.

[31] R.W. Kemmrd, L.A. Stone, Technometrics 11 (1969) 137-148.

[32] M. Daszykowski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux C, B. Walczak, TOMCAT: a MATLAB toolbox for multivariate calibration techniques. Chemom. Intell. Lab. 85 (2007) 269-277.

Figure captions:

Figure 1 (a) Cross validation of robust GRBF-OCPLS for peanut oils; (b) Training of robust GRBF-OCPLS for peanut oils; (c) Prediction of robust GRBF-OCPLS for peanut oils; (d) Prediction of robust GRBF-OCPLS for sunflower seed, rapeseed and sesame oils

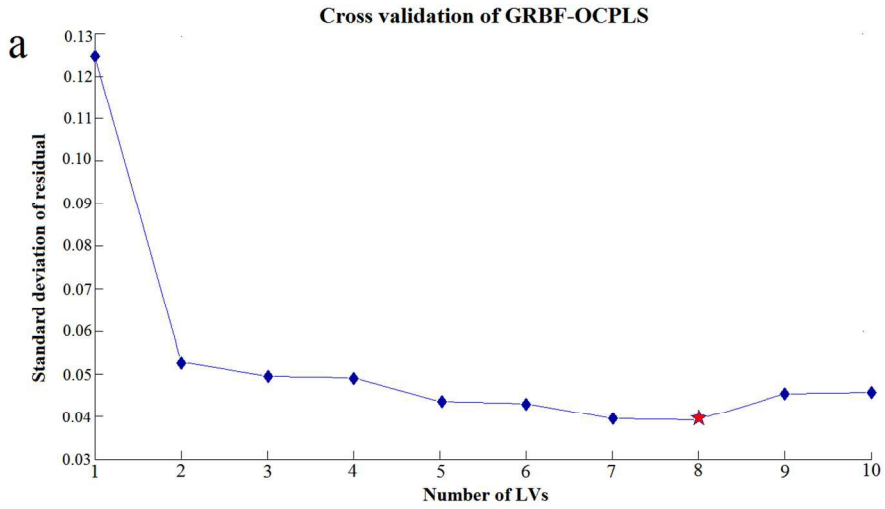
Figure 2 (a) Training of robust GRBF-OCPLS for peanut oils; (b) Prediction of robust GRBF-OCPLS for fake peanut oils with the adulteration levels of 3%-12%

Table title:

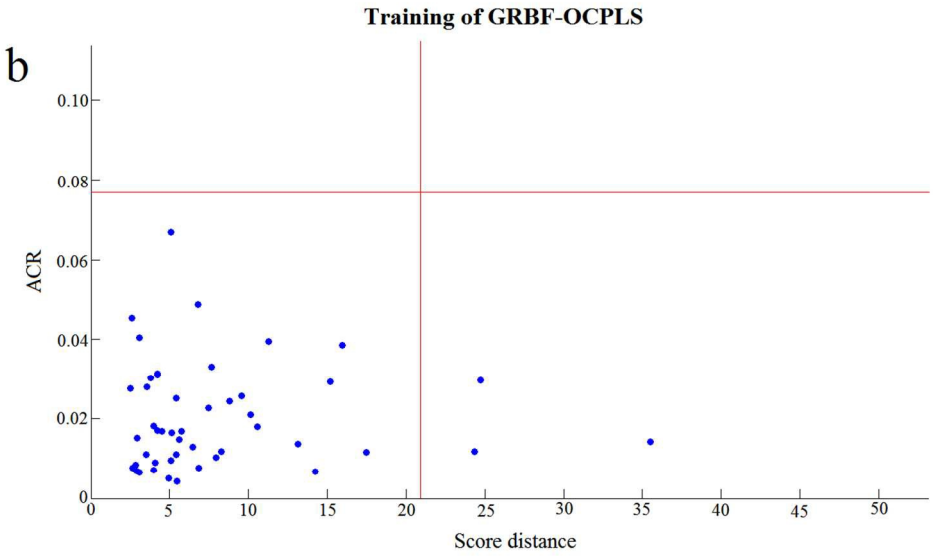
Table 1 Prediction results of pure and adulterated peanut oils with the OCPLS model

Table 1 Prediction results of pure and adulterated peanut oils with the OCPLS model

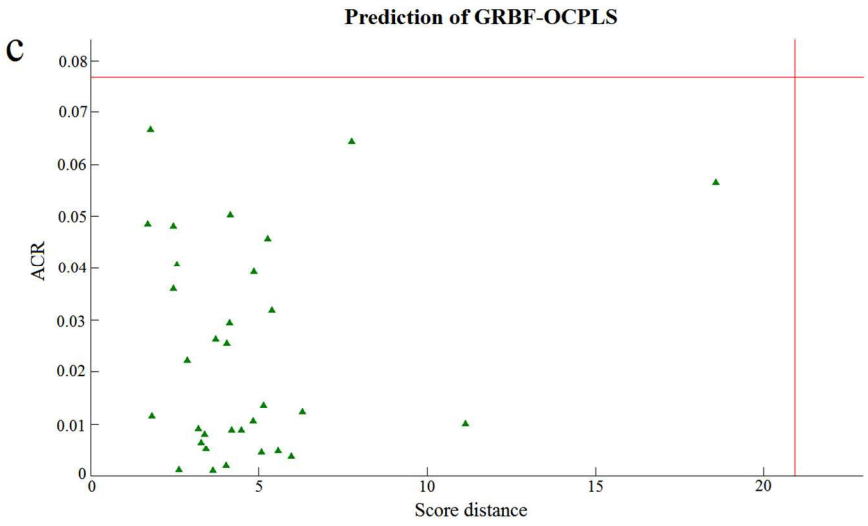
	Test set	Sensitivity	Specificity
1	Pure peanut oils	-	100% (32/32)
2	Other edible oils	100% (205/205)	-
3	Adulterated peanut oils (3%)	92.5% (37/40)	-
4	Adulterated peanut oils (4%)	100% (40/40)	-



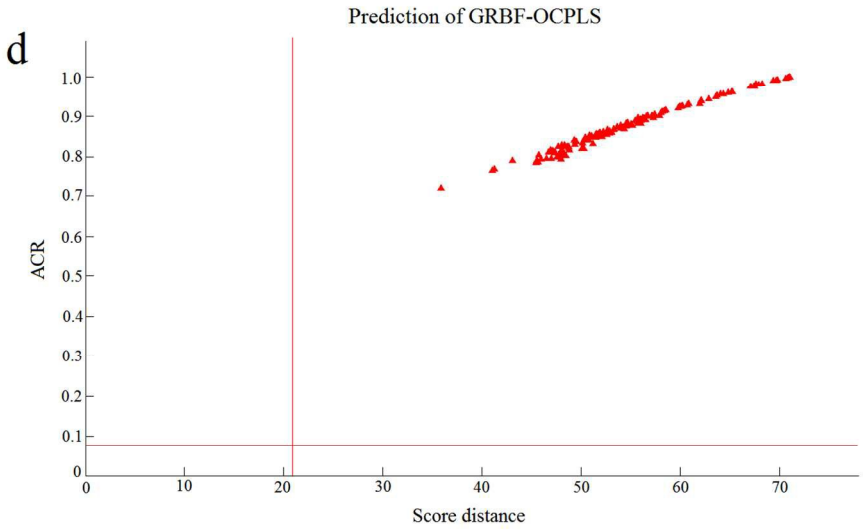
381x198mm (96 x 96 DPI)



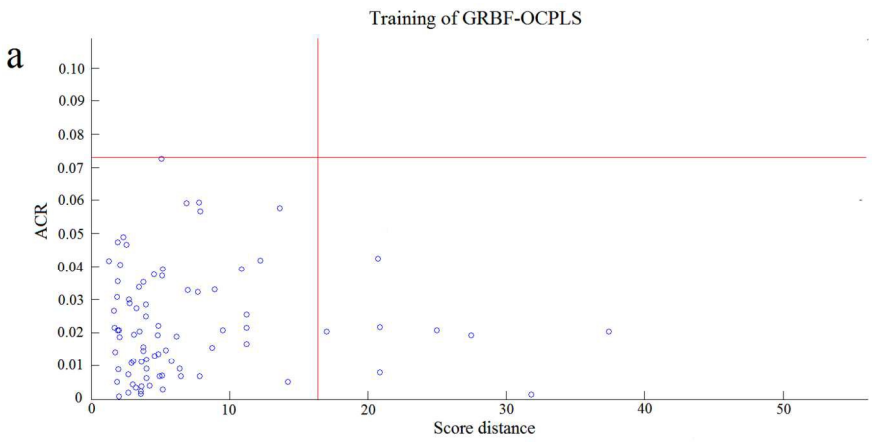
352x202mm (96 x 96 DPI)



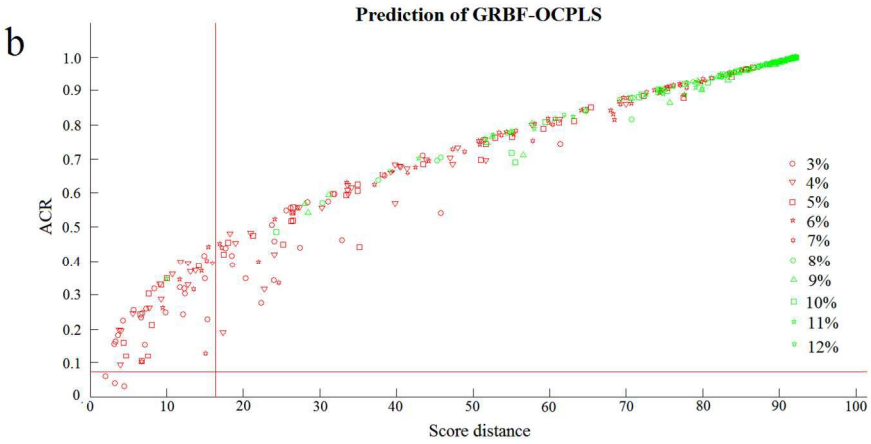
381x205mm (96 x 96 DPI)



381x205mm (96 x 96 DPI)



847x376mm (96 x 96 DPI)



847x388mm (96 x 96 DPI)