Chemometrics and Intelligent Laboratory Systems 155 (2016) 145–150

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

A combination of chemometrics methods and GC–MS for the classification of edible vegetable oils

Xinhui Li, Wei Kong, Weimin Shi *, Qi Shen **

College of Chemistry and Molecular Engineering, Zhengzhou University, Zhengzhou, 450052, China

ARTICLE INFO

Article history:
Received 15 January 2016
Received in revised form 18 March 2016
Accepted 23 March 2016
Available online 1 April 2016

Keywords:
Edible vegetable oil
GA–SVM
Kennard–Stone algorithm
Fatty acid
GC–MS

ABSTRACT

The authenticity of edible vegetable oils is a very important issue due to consumer health and commercial reasons. Gas chromatography–mass spectrometry (GC–MS) was applied to analyze the fatty acid composition of sixty six samples from six different kinds of edible vegetable oils. The fatty acid profiles of these edible vegetable oils were used to classify the type of edible oils. For improving the classification accuracy of vegetable oils with respect to type, the support vector machine (SVM) technique, optimized using the genetic algorithm (GA), was employed to construct the classification model. The effectiveness of the GA–SVM combination in classification was compared with that of other well-known strategies for classification, such as minimum distance classification (MDC) and linear discriminant analysis (LDA). In addition, the Kennard–Stone algorithm was used to select the representative training samples and compared with the random sampling method. The misclassification rates were 8.48% and 3.03% for training and test set, respectively, by the GA–SVM model using the linear kernel. Only one or two samples will be misclassified in the process of GA–SVM classification. The classification task based on fatty acid data can be successfully achieved by the GA–SVM technique combined with the Kennard–Stone algorithm. The results reveal that this strategy is of great promise in flexible and accurate classification of edible vegetable oils.

1. Introduction

Edible vegetable oils have been an indispensable ingredient of our diet in daily life since they contain a variety of essential fatty acids that are necessary for the human body by accelerating the absorption of fat-soluble vitamins [1,2]. They can also provide the body with a direct source of energy. The most salient feature of vegetable oils is that their nutritional value is higher than animal oils. The long-term consumption of animal fats, comprising primarily saturated fatty acids, will increase the risk of hypertension and coronary heart disease [3], which makes edible vegetable oils especially attractive for people. The performances and qualities of different types of vegetable oils vary during homemade cooking and food production depending on their compositions [4]. Therefore, the authenticity of edible vegetable oils is a very important issue while considering commercial and consumer health reasons. In the present study, the continued need for improving the classification accuracy of edible vegetable oils is investigated.

Nowadays, routine methods of analysis of vegetable oils involve many instrumental analysis techniques, such as near and mid-infrared spectrometry [5,6], fluorescence [7], chemiluminescence [8], chromatography [9,10], nuclear magnetic resonance spectroscopy [11], and mass spectrometry [12]. Among these instrumental techniques mentioned above, gas chromatography–mass spectrometry (GC–MS) is frequently used to identify the vegetable oil type by analyzing their fatty acid composition [13–16]. However, due to the complex composition of the vegetable oil, the resulting chromatograms may be formed by the overlapping of several analytical peaks. Furthermore, the chromatograms of different vegetable oils may be too similar to be distinguished directly. To further the improvement of previous methods, many researchers have proposed the spectroscopic techniques in combination with chemometrics methods as an alternative method that can be used for the discrimination of different categories and the detection of adulterants in oils [17–19]. The commonly used chemometrics methods include principal component analysis (PCA) [20], linear discriminant analysis (LDA) [21], and minimum distance classification (MDC). In the present study, a support vector machine (SVM) technique [22] was employed to construct the classification model with genetic algorithm (GA) [23] to get the optimized solution for edible vegetable oil classification.

SVM is a promising machine learning technique with comprehensive theoretical foundation. Because of the powerful ability in interpreting the linear or nonlinear relationships between the sample information and their properties, SVM has exhibited desirable generalization performance in numerous applications. GA simulates the Darwinian evolution of natural selection and genetic mechanism natural evolution. As a popular global stochastic optimization technique,

* Corresponding author.
** Corresponding author. Tel.: +86 37167767957.
E-mail addresses: shiweimin@zzu.edu.cn (W. Shi), shenqi@zzu.edu.cn (Q. Shen).

http://dx.doi.org/10.1016/j.chemolab.2016.03.028
0169-7439/

GA has been successfully used for global search and optimization problem [24]. Here, GA is invoked to seek the optimal parameters for the classification model, including penalty constant and kernel parameter in a kernel transform of SVM. Using GA to get the optimal solution readily makes SVM an adaptive parameter-free method for edible vegetable oil identification, without any parameters to be adjusted. A synergetic optimization of the parameters also enables a flexible modeling approach for SVM according to the performance of the total model. The proposed strategy has been applied to the classification of 66 samples from six different kinds of edible vegetable oils. The effectiveness of the GA–SVM in classification ability was compared with that of other well-known strategies for classification, such as minimum distance classification (MDC) and linear discriminant analysis (LDA). To further improve the classification accuracy, the Kennard–Stone algorithm was employed for the selection of samples for each type of vegetable oil.

2. Experimental section

2.1. Chemical reagents and sample collection

All reagents used in the experiment were of analytical grade. Petroleum ether (boiling point 30–60 °C), methanol, and methylbenzene were provided by Sinopharm Chemical Reagent Co. Ltd (shanghai, China), potassium hydroxide from Shenghao chemical reagent Co. Ltd (Guangzhou, China), hydrochloric acid (12 M) from Kaixin chemical reagent Co. Ltd (Hengyang, China).

All solutions were prepared using ultrapure water, which was obtained through a Millipore Milli-Q water purification system (Billerica, MA, USA) and had an electric resistance > 18.3 MΩ.

A total of 66 samples of edible vegetable oils were acquired from local supermarkets, including 14 samples of soybean oil, 14 samples of rapeseed oil, 10 samples of peanut oil, 12 samples of sesame oil, 8 samples of corn germ oil, and 8 samples of camellia oil. Different brands or different batches of pure edible oil samples were used in the experiment rather than simply replicates of same oil product. The brands of edible oils include fulinmen, luhua, jinlongyu, ganhua, knife, shengzhou, duoli, xinyuan, xiangmanyuan, and huishan. These samples were split into two groups for training and prediction sets; 7 soybean oils, 7 rapeseed oils, 5 peanut oils, 6 sesame oils, 4 corn germ oils, and 4 camellia oils were selected and combined into a training set to create a classification model. All the rest of the samples, which were not involved in constructing the classification models, were used as a prediction set.

2.2. Sample preparation

Edible vegetable oil of 100 mg was added into a mixture (2 mL) of petroleum ether and methylbenzene (v/v 1:1) and shaken until the oil completely dissolved (~10 min). Then 2 mL KOH (0.5 M in methanol) was added and shaken for another 5 min to simultaneously saponify and methyl esterify the oil. After that, the solution was incubated with 2 mL HCl (2 M) for 3 min to neutralize the excessive potassium hydroxide. The final organic solution on the top was filtered using a 0.45 μm filter for GC–MS analysis.

2.3. GC–MS analysis

The fatty acid profiles of vegetable oil samples were determined using GC–MS (Leco Pegasus® 4D, USA) equipped with a DB-5MS capillary column (30 m × 0.25 mm × 0.25 μm) and a time-of-flight mass spectrometer. Helium (99.99%) was used as the carrier gas at a constant flow rate of 1 mL/min. The oven temperature was programmed from 60 ºC to 215 ºC at 15 ºC/min, to 250 ºC at 10 ºC/min, and to 260 ºC at 2 ºC/min. Then it was finally increased at 280 ºC at 5 ºC/min and held for 2 min. Injections (1 μL each) used a 40:1 with the injector inlet temperature is 250 ºC.

Time of flight mass spectrometer was operated by electron ionization (EI), transfer line and EI temperatures, 270 ºC and 250 ºC, respectively. The solvent was delayed for 10 min.

The masses to display were recorded in total ion chromatogram (TIC). The m/z ranged from 30 to 500 amu, and the acquisition rate was 20 spectra per second. The electron energy was 70 eV and detector voltage was 1500 V.

In our experiments, no internal standard was used, but all the oil samples were treated and repeated for three times according to the described sample preparation procedure, and GC–MS analyses were also performed three times for every prepared sample to record the final GC–MS data. We found there were no significant changes between the GC–MS data obtained from a same oil sample, suggesting the saponification and methylation procedures [25,26].

3. Chemometrics methods

3.1. Training and prediction set selection

The set of samples of edible oils was partitioned into two groups: a training set and a prediction set. Half of the samples were used for training purposes in classification studies, and the rest, constituting the prediction set, were used to evaluate the prediction capability of the classification model. There are several algorithms that can be used for the selection of samples for training and prediction sets. Among these algorithms, random sampling (RS) is widely used because of its simplicity. Using this algorithm, a training set is randomly extracted from the original data, and there is no need to select the representation of the data set. Another alternative approach is the Kennard–Stone (KS) algorithm [27,28], which aims to select a representative subset to ensure training samples spread evenly throughout the sample space. In this paper, both algorithms were employed for partitioning data in classification.

3.2. Genetic algorithm optimized support vector machine

Support vector machine (SVM) is a promising machine learning technique with a comprehensive theoretical foundation and usually displays desirable generalization performances in giving the solutions for both linear and nonlinear problems. SVM classified the samples by constructing a hyperplane, which maximizes the distance between the two classes. For the problem of nonlinearity, SVM generally uses a kernel function, by which the data are transformed from the original variables to a feature space in which the model becomes linear. There are two most commonly used kernel functions, which are Gaussian radial basis function (RBF) kernel and the linear kernel. A Gaussian radial basis function transform is frequently utilized if the preknowledge of the problem dealt with is lacking. Both of the two kernels are conducted and compared in the present investigation. The penalty constant C is introduced to adjust the confidence limit of machine learning and the proportion of empirical risk; parameter σ of the Gaussian kernel function defines the non-linear mapping from the original space to the high-dimensional feature space and influences the properties of the SVM classifier. GA was used to automatically determine the optimal parameters, C and σ, and to ensure SVM with the highest predictive accuracy and generalization ability simultaneously.

GA is a parameter searching and optimization technique based on emulating the evolutionary process of the nature, such as breeding, mating, and mutation phenomenon in natural selection and genetic evolution. GA starts from any initial population and then produces a group of individuals that are new, more adaptive to the environment by random selection, crossover, and mutation operation. Binary strings were adopted to encode the chromosome. The first and the second half of a binary string were converted to two decimal values that stand for the parameter C and parameter σ, respectively. The problem

of optimal solution can be achieved by its continuous reproduction and evolution from generation to generation. For a classification problem, the fitness value is usually the percentage of correct classification, which is determined by cross validation. In this paper, the fitness value is estimated using the classification error of SVM over 5-fold cross validation for the training set. The fitness function could be expressed as follows:

$$\text{Fitness} = \frac{N_e}{N_T} \times 100\% \quad (1)$$

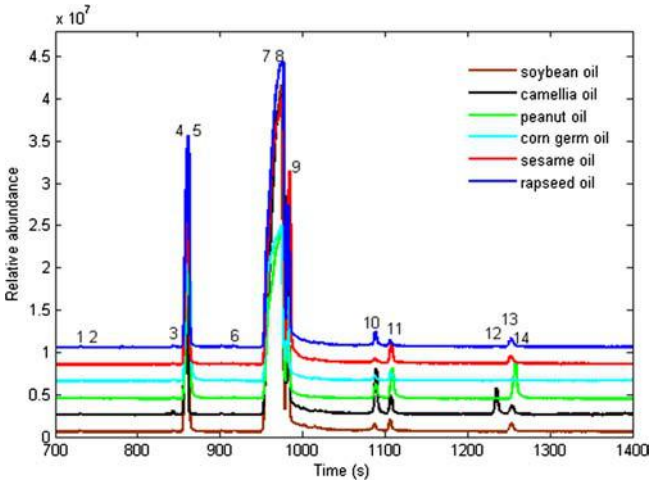
where N_e is the number of misclassified samples, N_T express the total number of the samples in the training set. The prediction set is never used for any optimization of parameters by the GA. The aim of the GA algorithm is to minimize this value. In 5-fold cross validation, the training samples are randomly partitioned into 5 subsamples. Of the 5 subsamples, a single subsample was retained as the validation data, and the remaining 4 subsamples were used for construction model. Optimizing the parameters of the SVM model by GA keeps the model from getting trapped into local optima and improves the model performance; it also enables SVM to be an adaptive parameter-free modeling technique for determining edible oil type.

All the calculations were implemented using the MATLAB version 7.10. The LibSVM toolbox was used for SVM classifications.

4. Results and discussion

4.1. Characteristics of fatty acid profiles of various edible vegetable oils

Fig. 1 shows an overlay of the typical total ion chromatographic profiles of the derivatives of fatty acids for different edible vegetable oils obtained using GC–MS, covering a range of time from 700 s to 1400 s. The main composition of edible vegetable oils is triglyceride, which is composed of a variety of different fatty acids. It can be found that the types of fatty acids included in the six different kinds of vegetable oils were quiet similar. Typically, tetradecanoic acid (C14:0), palmitic acid (C16:0), heptadecanoic acid (C17:0), hexadecenoic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and octadecatrienoic acid (C18:3) could be found in all these six edible vegetable oils. Tridecanoic acid (C13:0) and docosenoic acid (C22 :1) were contained in five kinds of the collected edible vegetable oils, the former including rapeseed oil, peanut oil, corn germ oil, sesame oil, and camellia oil, the latter containing soybean oil, rapeseed oil, corn germ oil, sesame oil, and camellia oil. Although different vegetable oils shared several kinds of fatty acids, the fatty acid contents may vary. Camellia oil had the highest oleic acid, which may be more than 75%. Contents of tetradecanoic acid, stearic acid, linoleic acid, and octadecatrienoic acid in soybean oil were higher than that in other oils. There was more palmitic acid in corn germ oil than in sesame oil, while the contents of linoleic acid and oleic acid were quite similar. Rapeseed oil was high in stearic acid. Peanut oil and rapeseed oil were high in oleic acid and low in linoleic acid. As a whole, the highest percentage level of monounsaturated fatty acid (MUFA), polyunsaturated fatty acid (PUFA), and saturated fatty acid (SFA) belong to camellia oil, soybean oil, and soybean oil, respectively. The lowest percentage level of PUFA and SFA belong to camellia oil, respectively. Among the fatty acids mentioned above, palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2) were the main fatty acids in vegetable oil. The percentage level of unsaturated fatty acids in the six kinds of vegetable oils was 80% and above. Furthermore, the composition and the content of each component of corn germ oil and sesame oil were exactly similar. More detailed information about the distribution of fatty acids in the six kinds of vegetable oils is presented in Table 1. The relative peak area is used from the FA profiles in our study. Despite these differences, it is still hard for one to distinguish the vegetable oil type accurately by directly using their TIC profiles. Thus, the potentials of chemometrics methods in distinguishing the edible vegetable oil type were investigated.



Note: 1, 2, 3… 14 stands for C12:0, C13:0, C14:0… C24:0 (Listed in Table 1), respectively

Fig. 1. Overlay of the typical total ion chromatogram of the derivatives of fatty acids for different edible vegetable oils obtained using GC–MS.

4.2. Clustering different kinds of edible vegetable oils using PCA

Before building the classification model, the relationship between the six kinds of edible vegetable oils was investigated using PCA. For chemometrics analysis, the relative contents of fourteen common fatty acids obtained using GC–MS were used to constitute the data matrix. The GC–MS data are scaled into (0, 1) for chemometrics analysis. Principal components (PC) are uncorrelated variables, which are linear combinations of the variables in the measurement matrix. The first PC possesses the most information and accounts for the largest variation in the original data. Then the second, third, fourth … principal components in turn are calculated, which account for successively smaller amounts of variation. Each PC is orthogonal to each other. Fig. 2A displays the projection results of all the 66 edible vegetable oil samples on the first two PCs. The first two PCs explained 64.88% and 23.26% of the total variation, respectively. The plot shows the obvious cluster tendency of the different classes of samples. It can be seen that the samples of camellia oil had their feature distributions and can be separated from other oils. Rapeseed oil and peanut oil were partially overlapping. The samples of corn germ oil cannot be distinguished from those of sesame oil or soybean oil in the projection map. It may be that the composition of corn germ oil is very similar to that of soybean oil and sesame oil. As shown in Table 1, the SFA, MUFA, and PUFA contents in corn germ oil and sesame oil are fairly similar, and the major components, oleic acid (C18:1) and linoleic acid (C18:2), are nearly equal. The content of linoleic acid (C18:2) is also quite similar in soybean oil and corn germ oil. The loading plot from the PCA based on the extracted concentration of analytes is shown Fig. 2B. The loading plot is useful to extract the implied analyte that has a higher effect on the PC score. According to the loading plot, three analytes (palmitic acid (C16:0), oleic acid (C18:1), and linoleic acid (C18:2)) distributed far from the center are the analytes contributed most to differentiation among different oils. The unsatisfied clustering results seem to indicate that PCA algorithm can only be employed for initially exploration the distribution trends of the oil samples. In order to classify these vegetable oils, three classification methods, including LDA, MDC, and GA–SVM, were investigated to predict the identity of the samples in the prediction set.

Table 1
The relative contents of the fatty acids in different edible vegetable oils.[a]

Fatty acid	Percentage level (%)					
	A[b]	B[b]	C[b]	D[b]	E[b]	F[b]
C12:0	–	0.08 ± 0.01	0.27 ± 0.03	0.05 ± 0.01	–	–
C13:0	–	0.45 ± 0.00	0.63 ± 0.09	1.72 ± 0.21	0.78 ± 0.02	0.54 ± 0.07
C14:0	1.26 ± 0.11	0.34 ± 0.03	0.04 ± 0.01	0.62 ± 0.06	0.06 ± 0.01	0.90 ± 0.20
C16:0	5.80 ± 1.01	3.73 ± 0.31	6.23 ± 0.30	9.14 ± 0.94	5.14 ± 0.78	1.51 ± 0.02
C17:0	1.84 ± 0.05	0.55 ± 0.03	2.97 ± 0.04	2.05 ± 0.07	0.73 ± 0.04	0.62 ± 0.01
C16:1	0.56 ± 0.06	0.30 ± 0.03	0.07 ± 0.02	0.14 ± 0.03	0.26 ± 0.05	2.81 ± 0.09
C18:0	8.28 ± 0.24	3.49 ± 0.32	2.95 ± 0.11	1.29 ± 0.08	5.14 ± 1.09	3.90 ± 0.35
C18:1	31.55 ± 0.77	50.95 ± 1.03	68.49 ± 2.06	39.65 ± 2.03	43.44 ± 1.05	79.38 ± 3.41
C18:2	45.00 ± 0.31	39.76 ± 1.04	18.28 ± 0.79	44.73 ± 1.02	43.81 ± 2.06	4.66 ± 0.69
C18:3	5.49 ± 0.57	0.34 ± 0.05	0.05 ± 0.01	0.56 ± 0.08	0.11 ± 0.02	3.64 ± 0.33
C20:2	–	–	–	0.12 ± 0.02	0.01 ± 0.01	0.14 ± 0.03
C20:3	–	–	–	–	–	1.03 ± 0.11
C22:1	0.47 ± 0.05	0.02 ± 0.01	–	0.15 ± 0.01	0.03 ± 0.03	0.10 ± 0.02
C24:0	–	–	0.02 ± 0.01	–	–	–
SFA[c]	17.18	8.64	13.11	14.87	11.85	7.47
MUFA[d]	32.58	51.27	68.56	39.94	43.73	82.29
PUFA[e]	50.49	40.10	18.33	45.41	43.93	9.47

[a] The data are presented as the mean ± SD.
[b] A: soybean oil; B: rapeseed oil; C: peanut oil; D: corn germ oil; E: sesame oil; F: camellia oil.
[c] SFA: saturated fatty acid.
[d] MUFA: monounsaturated fatty acid.
[e] PUFA: polyunsaturated fatty acid.

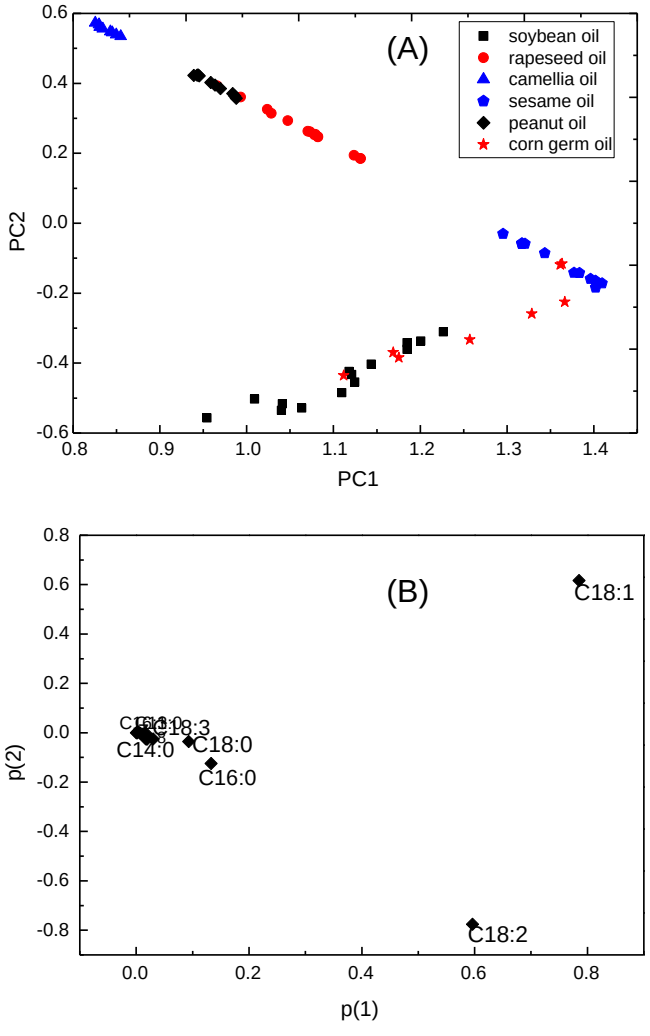


Fig. 2. Principal component analysis score plot (A) and loading plot (B).

4.3. Classifying edible vegetable oils based on random sampling (RS)

To accurately identify the edible vegetable oil samples, the GA-optimized SVM was firstly employed to build classification models. All the 66 samples were randomly divided into two independent data sets, including a training set of 33 samples and a prediction set of 33 samples. Both training and prediction sets contain six kinds of edible vegetable oils. The training set was used to construct the classification model, and the prediction set was used to demonstrate the performance of the classification model. Because of the arbitrariness of partition of data set, the classification error rate of a model at each iteration is not necessarily the same. When the data set was split into training and prediction data sets by random sampling (RS), the above procedure was repeated 100 times to evaluate the predictive ability and reliability of these models. This means that the total edible oil samples were randomly partitioned into training and prediction sets 100 times, and the classification errors for each classification model were calculated and averaged. In the present study, the population size of GA is 20, the maximum number of generations is 200, and the crossover and mutation rates are 0.9 and 0.1. In GA, the penalty constant C and parameter σ are modified in the range from 0 to 100. In SVM, the two kernel functions were both employed independently and compared to determine which kernel was the most suitable for this system. To estimate the proposed method more accurately, both the error over 5-fold cross validation for the training set and the error on the test data set were used for estimator of model performance in our study. The result obtained by the GA–SVM using Gaussian kernel gives the cross validation misclassification rate as 9.64% for the training set. The average error rate for the test set is 11.55%. When the GA–SVM classifier using the linear kernel was employed, there was a slight difference, with a cross validation classification error of 9.64% and a testing classification error of 12.03%. The results obtained by the GA–SVM with linear kernel are similar to that using Gaussian kernel; there was no distinct difference. When all GA–SVM classification terminates, one may count the number of times for a particular category of edible oil is misclassified in 100 individual classification models. The most frequently misclassified edible oil categories were sesame oils and corn germ oils. Once again, sesame oils and corn germ oils are not easily distinguishable and more than 90% of the error-prone vegetable oils are sesame and corn germ oil.

Table 2
The classification results of the classification models based on random sampling

Models	Misclassification rate (%)		
	Cross validation	Training set	Prediction set
LDA	10.84	2.15	39.45
MDC	9.78	4.67	15.58
GA–SVM1[a]	9.64	0.18	11.55
GA–SVM2[b]	9.64	0.48	12.03

[a] GA–SVM1: the model with the RBF kernel.
[b] GA–SVM2: the model with the linear kernel.

To further demonstrate the performance of GA–SVM, other classification algorithms, including LDA and MDC, were also used for the classification of these edible vegetable oils. Table 2 shows the statistical classification results of the classification models. Using LDA, the average error rate for training and prediction set were 10.84% and 39.45%, respectively. The predictive performance by LDA is poor. The average misclassification rate for the test set was 15.58% by MDC. Comparing the classification models, the GA-optimized SVM model has a better predictive performance. Flexible feature mapping using Gaussian kernel transform and the linear kernel, with kernel width and penalty constant optimized synergistically using the GA, guaranteed the SVM technique to generate an excellent classification model. Furthermore, it can be seen from Table 2 that the average error rate for test sets for the four models were obviously higher than that for training set. This may be caused by lack of the representative training sets. The unrepresentative training set may result in underestimation of model classification error.

4.4. Classification of edible oils based on KS algorithm

To improve the classification model, the KS algorithm was employed to select the representative training set. Table 3 shows the misclassification rate of different classification models using the KS algorithm. The two GA–SVM classifiers both achieved a cross validation classification error of 8.48% for the training set. The error rate for the prediction set was 4.61 % by the GA–SVM model on the Gaussian kernel transform. From Table 3, it can be observed that the prediction performance of the SVM classifier using the linear kernel is slightly better than using the other kernel, with a classification error of 3.03% for the test set. Only one or two samples will be misclassified in the process of GA–SVM classification. Compared with the random sampling, the misclassification rate reduced about 7% using the KS algorithm. The error rate by LDA has dropped from 39.45% by RS to 9.09% for the test set by the KS algorithm. The LDA model by the KS algorithm also gave a higher accuracy than that obtained by the RS technique. The MDC model presented a classification error of 6.06% for the prediction set when using the KS algorithm. Compared with 15.58% acquired by RS, there was a 9.52% improvement in classification accuracy. By selecting the representative training set using the KS algorithm, the classification performance of all the models was improved. The above results demonstrate that the classification task of edible oil using fatty acid data can be successfully solved by GA–SVM techniques combined with the Kennard–Stone algorithm.

Table 3
The classification results of the classification models based on KS algorithm

Models	Misclassification rate (%)		
	Cross validation	Training set	Prediction set
LDA	9.13	6.06	9.09
MDC	8.56	3.03	6.06
GA–SVM1[a]	8.48	2.06	4.61
GA–SVM2[b]	8.48	0	3.03

[a] GA–SVM1: the model with the RBF kernel.
[b] GA–SVM2: the model with the linear kernel.

5. Conclusion

In the present study, MDC, LDA, and GA–SVM were employed to construct classification models for edible vegetable oils using a fatty acids data set of oil samples obtained from the GC-MS. It was verified that a combination of chemometrics methods and GC-MS could be a suitable tool for the classification of edible vegetable oils. Compared with the LDA and MDC models, the GA–SVM techniques combined with the Kennard–Stone algorithm can successfully classify the edible vegetable oils. The performance of the proposed strategy in classifying six different edible vegetable oils using the fatty acid data of GC–MS is of great promise in flexible and accurate classification of edible vegetable oils. As SVM technique can be used for not only classification but also regression analysis, the proposed method can be easily extended to the detection of adulterants in oils. The strategy of fatty acid profiles might be also useful to detect the adulteration with low quality oils.

Disclaimer

The material presented in this publication reflects the opinions of the authors, and not of their institutional affiliations.

Conflict of interest

There is no conflict of interest.

Acknowledgement

The work was financially supported by the National natural Science Foundation of China (Grant No. 21575131).

References

[1] M.M. Gui, K.T. Lee, S. Bhatia, Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock, Energy 33 (2008) 1646–1653.
[2] T. Aoyama, N. Nosaka, M. Kasai, Research on the nutritional characteristics of medium-chain fatty acids, J. Med. Investig. 54 (2007) 385–388.
[3] G. Fang, J.Y. Goh, M. Tay, H.F. Lau, S.F. Li, Characterization of oils and fats by 1H NMR and GC/MS fingerprinting: classification, prediction and detection of adulteration, Food Chem. 138 (2013) 1461–1469.
[4] J. Kim, D.N. Kim, S.H. Lee, S.H. Yoo, S. Lee, Correlation of fatty acid composition of vegetable oils with rheological behaviour and oil uptake, Food Chem. 118 (2010) 398–402.
[5] L. Rasterlli, S. Passi, F. Ippolito, G. Vacca, F. De Simone, Rate of degradation of tocopherol, squalene, phenolics, and polyunsaturated fatty acids in olive oil during different storage conditions, J. Agric. Food Chem. 50 (2002) 5566–5570.
[6] R. Japón-Luján, J. Ruiz-Jiménez, M.D. Luque de Castro, Discrimination and classification of olive tree varieties and cultivation zones by biophenol contents, J. Agric. Food Chem. 54 (2006) 9706–9712.
[7] S. Morais, J. Pratoomyot, J.B. Taggart, J.E. Bron, D.R. Guy, J.G. Bell, D.R. Tocher, Genotype-specific responses in Atlantic salmon (*Salmo salar*) subject to dietary fish oil replacement by vegetable oil: a liver transcriptomic analysis, BMC Genomics 12 (2011) 255–259.
[8] J.A. Murillo Pulgarín, L.F. García Bermejo, A. Carrasquero Durán, Evaluation of the antioxidant activity of vegetable oils based on luminol chemiluminescence in a microemulsion, Eur. J. Lipid Sci. Technol. 112 (2010) 1294–1301.
[9] Y.L. Chu, S.J. Jiang, Speciation analysis of arsenic compounds in edible oil by ion chromatography–inductively coupled plasma mass spectrometry, J. Chromatogr. A 1218 (2011) 5175–5179.
[10] R. Aparicio, R. Aparicio-Rúiz, Authentication of vegetable oils by chromatographic techniques, J. Chromatogr. A 881 (2000) 93–104.
[11] G.P. Mambrini, C. Ribeiro, L.A. Colnago, Nuclear magnetic resonance spectroscopic analysis of ethyl ester yield in the transesterification of vegetable oil: an accurate method for a truly quantitative analysis, Magn. Reson. Chem. 50 (2012) 1–4.
[12] S. Vichi, N. Cortés-Francisco, J. Caixach, Ultrahigh resolution mass spectrometry and accurate mass measurements for high-throughput food lipids profiling, J. Mass Spectrom. 47 (2012) 1177–1190.
[13] C. Ruiz-Samblás, L. Cuadros-Rodríguez, A. González-Casado, P. de la Mata-Espinosa, F. García, J.M. Bosque-Sendra, Multivariate analysis of HT/GC-(IT)MS chromatographic profiles of triacylglycerol for classification of olive oil varieties, Anal. Bioanal. Chem. 399 (2011) 2093–2103.
[14] M.J. Lerma-García, E.F. Simó-Alfonso, A. Méndez, J.L. Lliberia, J.M. Herrero-Martínez, Classification of extra virgin olive oils according to their genetic variety using linear discriminant analysis of sterol pro files established by ultra-performance liquid chromatography with mass spectrometry detection, Food Res. Int. 44 (2011) 103–108.

[15] S. Laroussi-Mezghani, P. Vanloot, J. Molinet, N. Dupuy, M. Hammami, N. Grati-Kamoun, J. Artaud, Authentication of Tunisian virgin olive oils by chemometric analysis of fatty acid compositions and NIR spectra. Comparison with Maghrebian and French virgin olive oils, Food Chem. 173 (2015) 122–132.
[16] M. Casale, P. Oliveri, C. Casolino, N. Sinelli, P. Zunin, C. Armanino, M. Forina, S. Lanteri, Characterisation of PDO olive oil Chianti Classico by non-selective (UV-visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques, Anal. Chim. Acta 712 (2012) 56–63.
[17] N. Dupuy, O. Galtier, D. Ollivier, P. Vanloot, J. Artaud, Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis, Anal. Chim. Acta 666 (2010) 23–31.
[18] N. Dupuy, L. Duponchel, J.P. Huvenne, B. Sombret, P. Legrand, Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra, Food Chem. 57 (1996) 245–251.
[19] J.M. Bosque-Sendra, L. Cuadros-Rodríguez, C. Ruiz-Samblás, A.P. de la Mata, Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data, Anal. Chim. Acta 724 (2012) 1–11.
[20] R. Aparicio, R. Aparicio-Rúiz, Authentication of vegetable oils by chromatographic techniques, J. Chromatogr. A 881 (2000) 93–104.
[21] H. Lizhi, K. Toyoda, I. Ihara, Discrimination of olive oil adulterated with vegetable oils using dielectric spectroscopy, J. Food Eng. 96 (2010) 167–171.
[22] Y. Lee, C.K. Lee, Classification of multiple cancer types by multicategory support vector machines using gene expression data, Bioinformatics 19 (2003) 1132–1139.
[23] H. Li, C. Nantasenamat, T. Monnor, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, Genetic algorithm search space splicing particle swarm optimization as general-purpose optimizer, Chemom. Intell. Lab. 128 (2013) 153–159.
[24] O. Devos, G. Downey, L. Duponchel, Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils, Food Chem. 148 (2014) 124–130.
[25] D. Wang, J. Cai, B.Q. Zhu, G.F. Wu, C.Q. Duan, G. Chen, Y. Shi, Study of free and glycosidically bound volatile compounds in air-dried raisins from three seedless grape varieties using HS-SPME with GC-MS, Food Chem. 177 (2015) 346–353.
[26] M.R. Brunetto, S. Clavijo, Y. Delgado, W. Orozco, M. Gallignani, C. Ayala, V. Cerda, Development of a MSFIA sample treatment system as front end of GC-MS for atenolol and propranolol determination in human plasma, Talanta 132 (2015) 15–22.
[27] R.K.H. Galvão, M.C.U. Araujo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, A method for calibration and validation subset partitioning, Talanta 67 (2005) 736–740.
[28] W. Wu, B. Walczak, D.L. Massart, S. Heuerding, F. Erni, I.R. Last, K.A. Prebble, Artificial neural networks in classification of NIR spectral data: design of the training set, Chemom. Intell. Lab. 33 (1996) 35–46.