Identification of Edible-vegetable-oil Types Based on Multi-kernel Learning and Multi-spectral Fusion

Yang Chen, Jie Wang, Qiang Xu, Qingsong Luo and Xiao Zheng*

School of Mechanical Engineering, Wuhan Polytechnic University, Wuhan, Hubei, 430023

*Corresponding author e-mail: kjiwj74135@163.com

Abstract. Aiming at the identification of edible oil quality, this study proposed a multi-spectral data fusion combined with multi-kernel learning support vector machine (SVM) method. This method used serial and wavelet fusion approaches to fuse Raman and near infrared spectral data, and established an identification model for edible-oil types, with the aid of the multi-kernel learning support vector machine (MKL-SVM). The performances of the single spectral model and spectral fusion model were compared, demonstrating that the spectral fusion could effectively improve the prediction accuracy and generalization ability of the model.

1. Introduction

With the development of economy and living standard, the quality of edible oil has drawn more and more attention. To gain more profits, many immoral providers fake high-grade edible oil with low-grade oil, jeopardizing the interests and health of consumers [1-2]. Therefore, it is of great significance to explore a detection technique for the fast and accurate identification of quality and composition of various commercial edible oils.

Currently, spectroscopy has been extensively applied in the quality identification of edible oils. In this field, the support vector machine (SVM) technology [3] has been widely applied to the analysis of spectral data. Compared to the traditional single-kernel support vector machine, the multi-kernel learning support vector machine (MKL-SVM), possessing the concept of multi-kernel learning, applies different projection methods to the different types of data in sample space, reducing the influence of original data on the parameters of a model and improving the generalization ability of the model [4]. On the other hand, data fusion is a multilevel data processing process, which consists of three levels: data, feature and decision layers. Each layer has its specific advantages and disadvantages [5]. This work adopted the data fusion technique, of which the information compression degree was appreciable. This merit is beneficial to real-time processing, and the fusion results can provide critical information for the decision analysis to a great extent because the extracted features are directly related to the decision analysis.

In this paper, the MKL-SVM models were established based on Raman and near infrared spectral data by using the data fusion and multi-kernel learning support vector machine technologies, for the fast identification of edible oil types.

2. Experimental

2.1. Materials

The performances of the MKL-SVM models established would greatly depend on the number of types of samples. Hence, we purchased eight types of vegetable oils including soybean oil, peanut oil, rapeseed oil, rice oil, corn oil, sunflower seed oil, camellia oil and olive oil, provided by famous providers in the globe. In addition, we purchased the feedstock of these eight types of vegetable oils, and extracted oil samples from the feedstock, to ensure the authenticity of the relative samples.

After the preparation work, a total of 468 edible oil samples were obtained, and is summarized in Table 1. The SPXY algorithm was used to separate these oil samples into calibration and prediction sets at a ratio of 3:1.

2.2. Instrument and spectral acquirement

An RamTraceer-200 laser Raman spectral instrument (OptoTrace Technologies, Inc., China) was used in this study. The laser wavelength was 785nm, the resolution was ≤ 8cm-1, the wavenumber studied was in the range of 250–2340cm-1, and the maximum laser power was 320mW. The integration time of the Raman spectrometer was set to be 5 seconds and the laser power was 220mW in this study.

A home-made laser near-infrared spectral instrument with an AxsunXL410-type host (AXSUN, USA) was used for the fast detection of the quality of edible oil. The spectra were scanned 32 times in the range of 1350–1800nm, the resolution was 3.5 cm-1, the wavelength repeatability was 0.01nm, and the signal-to-noise ratio (250ms, RMS) was higher than 5500:1. In this measurement, 2-, 5-, and 10-mm cuvettes could be selected. The temperature was in the range of 20–100 ℃. Each sample was tested 3 times at room temperature and the average spectrum was used in the following steps. The original Raman spectra acquired in the range of 780–1800cm-1 with a high signal to noise ratio are shown in Figure 1; the original near infrared (NIR) spectra acquired are shown in Figure 2.

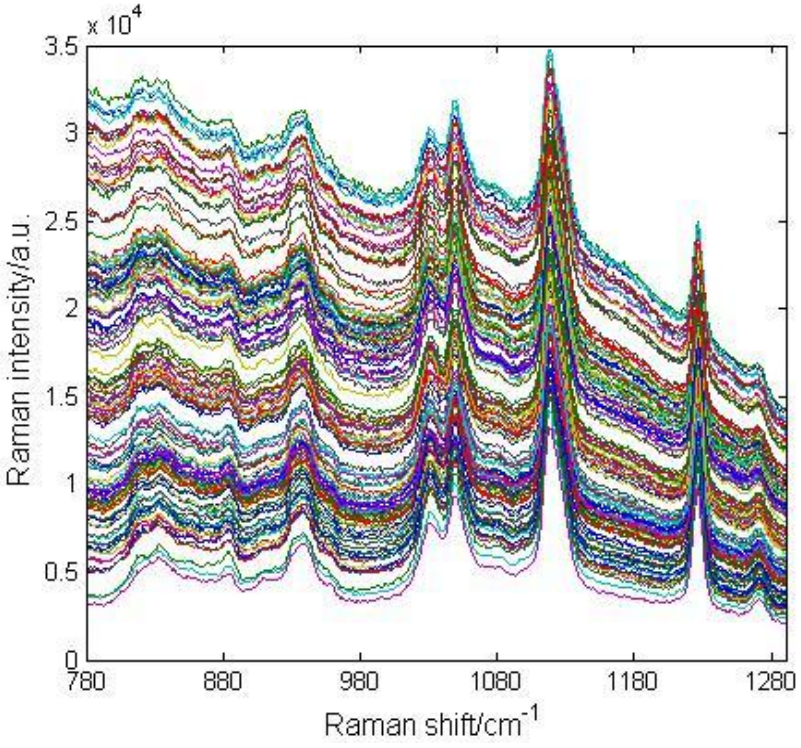

Figure 1. Original Raman spectra.

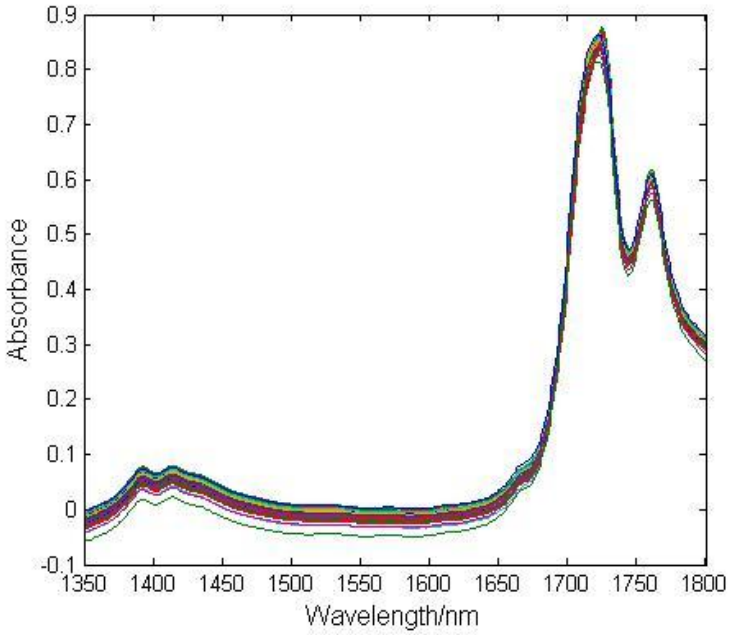

Figure 2. Original near-infrared spectra.

2.3. Data preprocessing methods

In the experiment, the Raman spectra were separately preprocessed with the moving average 11-point method, adaptive iterative reweighted-penalty least square method, and the normalization method based on the intensity of characteristic peak at 1454 cm-1 (MA11-airPLS-Nor). The Raman spectra preprocessed are shown in Figure 3. On the other hand, the NIR spectra were preprocessed following the standard normal variable transformation algorithm combined with detrending technique (SNV_DT). The NIR spectra preprocessed are shown in Figure 4.

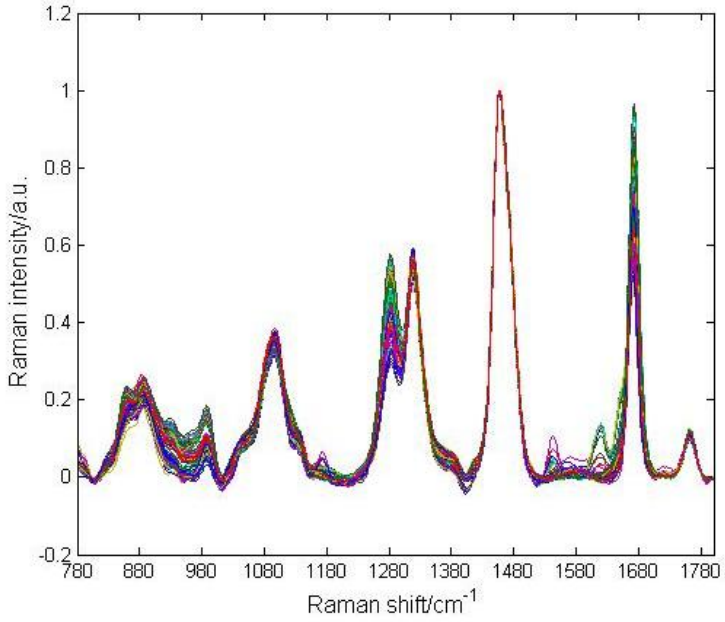

Figure 3. Raman spectra preprocessed with MA11-airPLS-Nor.

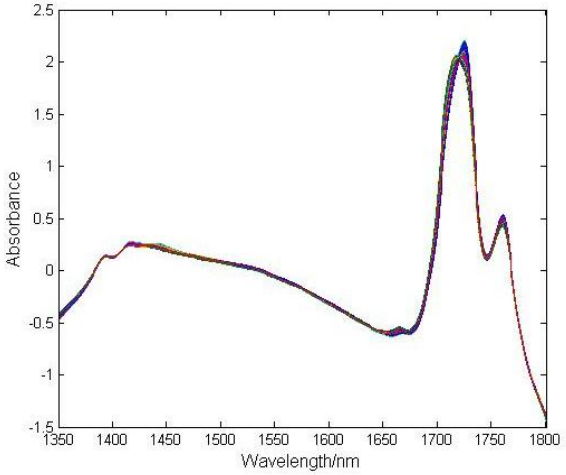

Figure 4. Near-infrared spectra preprocessed with SNV_DT.

2.4. Data fusion methods

In this study, the Raman and NIR spectra were separately fused on the feature level with the serial fusion and wavelet fusion approaches.

The serial fusion transformed the Raman and NIR spectra on the feature level into the same coordinate system. Then, the dimensions of feature were reduced by using the competitive adaptive reweighted sampling (CARS) method, for the extraction of the characteristic information after the fusion of Raman and NIR spectra [6].

The wavelet fusion technique decomposes the spectral information at different frequency ranges. The different spectral information at different frequency ranges was fused with different information fusion strategies, for the retention of the effective information of the spectra [7].

The wavelet fusion was performed according to the following fusion procedure:

First, two groups of spectra were separately transformed following the discrete wavelet transform to derive the low- and high-frequency details of the spectra. Then, the spectral information was fused following the principle: a larger coefficient was selected for high-frequency details, and mean coefficient was selected for low-frequency details. Eventually, the fused image was rebuilt through the inverse wavelet transform [8].

The preprocessed Raman and NIR spectra are shown in Figures 5–7.

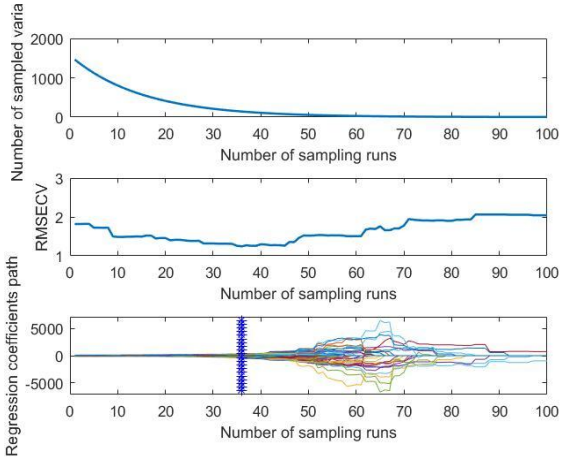

Figure 5. NIR spectra preprocessed with SNV-DT and CARS-optimized-Raman-spectra preprocessed with MA11-airPLS-Nor.

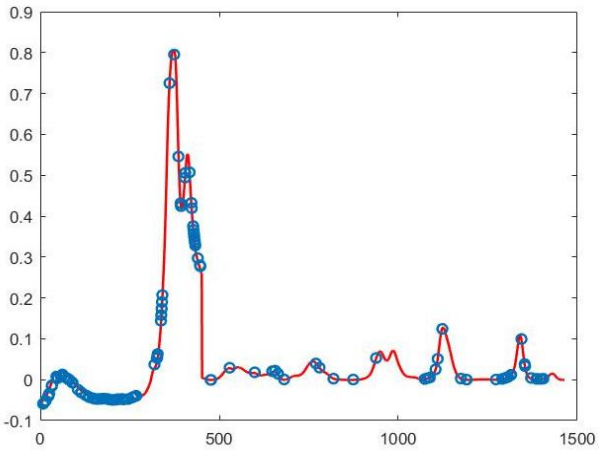

Figure 6. NIR spectra preprocessed with SNV-DT and CARS-optimized-Raman-spectra preprocessed with MA11-airPLS-Nor.

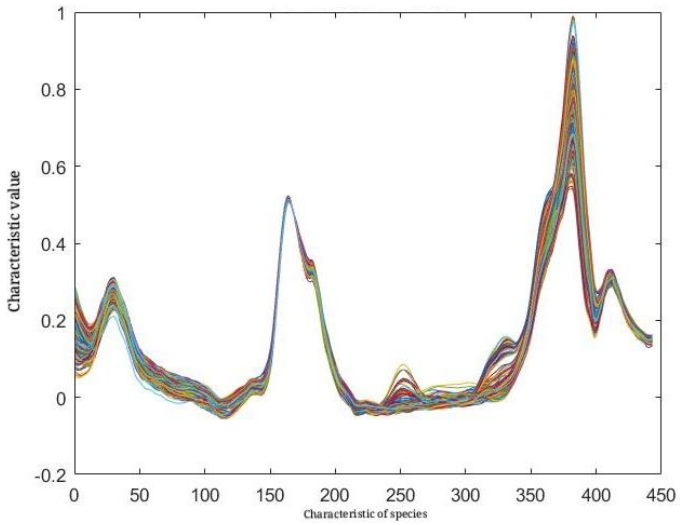

Figure 7. NIR spectra preprocessed with SNV-DT and wavelet-fusion-Raman-spectra preprocessed with MA11-airPLS-Nor.

3. MKL-SVM models for type identification

The identification model was established by the MKL-SVM. Based on the idea of multi-kernel learning, the spectral data were divided into 10 groups according to the feature dimensions. The features of these 10 groups were the Gauss kernels (RBF). The particle swarm optimization algorithm (PSO) was employed to optimize the parameters (C, g) of each group. A total of 10 groups of (C, g) were obtained. The kernels of these 10 different (C, g) values were processed with the weighted voting, making a kernel with the better classification performance more powerful in the classification process. Thus, the final parameters (C, g) were 10×8 matrixes ([C], [g]).

The SNV-DT-preprocessed NIR, MA11-airPLS-Nor-preprocessed Raman, SNV-DT-MA11-airPLS-Nor-CARS-fused, and SNV-DT-MA11-airPLS-Nor-DWT-fused spectral data were used as the input variables, for the separate establishment of MKL-SVM models. The prediction results of these

MKL-SVM models are shown in Table 2. The parameters in Table 2 are shown in Figures 8–15. The PSO optimization processes and prediction results of these models are shown in Figures 16–23.

C1 =

833	833	880	448	73	226	380	91
899	756	622	804	217	711	245	392
600	920	963	658	532	284	581	830
274	596	614	618	271	630	379	866
251	718	182	550	695	317	521	404
955	1000	985	967	550	933	919	87
544	159	109	978	55	319	634	845
101	64	840	363	415	287	997	626
33	74	181	789	8	240	575	428
971	780	43	837	715	947	498	59

Figure 8. Parameter C1 of the SNV-DT-NIR-MKL-SVM.

g1 =

336	896	90	857	393	675	696	440
645	364	531	807	796	213	957	618
914	772	770	577	239	304	873	552
616	1000	877	870	920	557	929	310
921	921	893	658	611	680	570	466
439	424	921	357	789	868	901	613
545	837	996	792	388	319	563	962
54	748	436	246	44	526	196	162
52	596	653	674	938	651	119	566
104	177	546	143	721	506	491	44

Figure 9. Parameter g1 of the SNV-DT-NIR-MKL-SVM.

C2 =

143	57	138	946	785	1000	817	72
57	16	795	1000	18	912	541	985
81	403	1000	46	564	324	497	655
20	1000	4	592	131	498	348	209
99	692	214	224	443	1000	39	0
356	1000	457	1000	348	774	935	41
32	312	655	1000	987	613	605	515
887	534	75	720	924	1000	241	407
94	1000	187	269	659	514	936	927
97	522	497	158	72	427	22	471

Figure 10. Parameter C2 of the MA11-airPLS-Nor-Raman-MKL-SVM.

g2 =

59	586	261	168	112	72	41	134
204	268	15	7	995	753	34	6
396	123	93	664	120	200	402	30
280	10	994	18	495	20	54	32
645	17	140	66	57	76	610	648
277	84	453	64	64	39	16	467
19	43	183	25	19	11	40	55
111	559	727	307	54	42	730	279
173	25	174	108	30	33	4	2
378	573	854	810	984	12	761	423

Figure 11. Parameter g2 of the MA11-airPLS-Nor-Raman-MKL-SVM.

C3 =

766	1000	997	1000	955	1000	946	580
246	1000	817	1000	1000	729	269	932
775	906	1000	704	848	228	845	803
989	1000	1000	634	1000	932	848	603
956	837	935	982	503	480	920	770
405	311	312	272	884	551	136	49
384	1000	699	542	732	10	991	142
1000	563	824	214	928	592	330	174
496	731	875	888	696	864	915	914
1	697	1000	269	436	1000	974	998

Figure 12. Parameter C3 of the NIR-Raman-CARS-MKL-SVM.

g3 =

87	913	1000	478	998	956	1000	769
1000	1000	558	777	1000	83	710	1000
661	900	1000	851	661	976	1000	259
966	1000	1000	684	1000	1000	765	599
783	820	551	1000	663	690	392	383
937	102	418	788	969	476	942	943
355	938	589	780	920	287	720	64
307	1000	778	66	141	206	497	957
904	368	353	56	345	799	770	780
909	17	97	993	1000	910	693	468

Figure 13. Parameter g3 of the NIR-Raman-CARS-MKL-SVM.

C4 =

25	118	728	22	208	4	274	111
0	858	13	23	0	58	15	291
20	28	118	30	367	1	47	39
67	192	3	1000	5	26	232	5
569	146	398	4	0	45	56	922
11	113	968	672	37	452	28	35
265	38	509	6	413	19	1000	767
12	86	468	0	226	47	53	320
2	41	41	11	0	17	4	9
10	165	131	164	15	13	431	9

Figure 14. Parameter C4 of the NIR-Raman-DWT-MKL-SVM.

g4 =

74	5	957	35	83	214	1	5
321	1	647	380	372	186	240	0
400	55	40	86	6	1000	37	0
23	6	1000	2	563	502	0	1000
3	1	16	865	0	342	167	495
756	17	3	1	354	12	238	278
16	605	737	39	7	181	2	175
1000	913	867	987	792	672	8	7
1000	695	813	15	124	742	482	330
1000	5	32	4	128	346	484	839

Figure 15. Parameter g4 of the NIR-Raman-DWT-MKL-SVM.

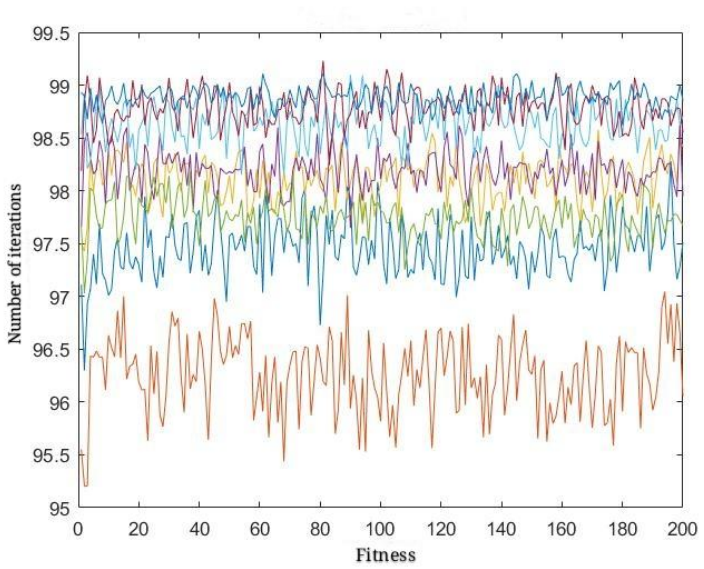

Figure 16. Parameters optimization process of the NIR-MKL-SVM.

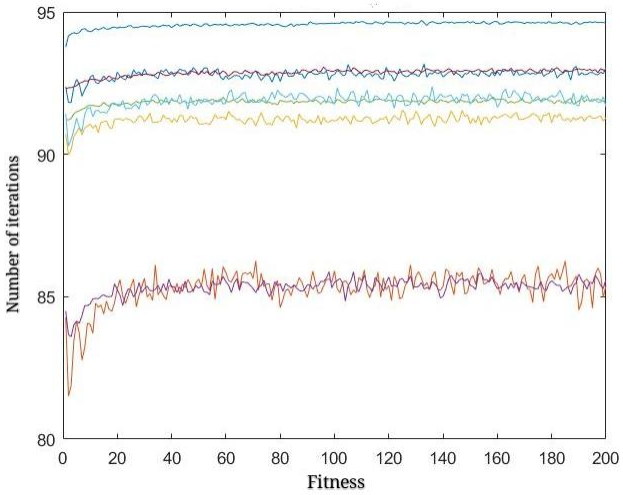

Figure 17. Parameters optimization process of the Raman-MKL-SVM.

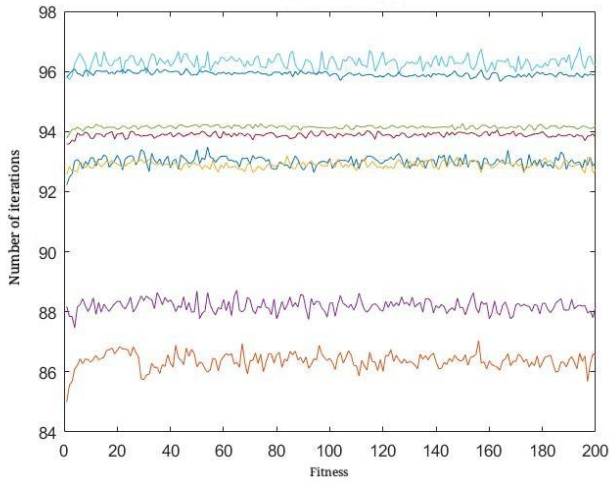

Figure 18. Parameters optimization process of the CARS-MKL-SVM.

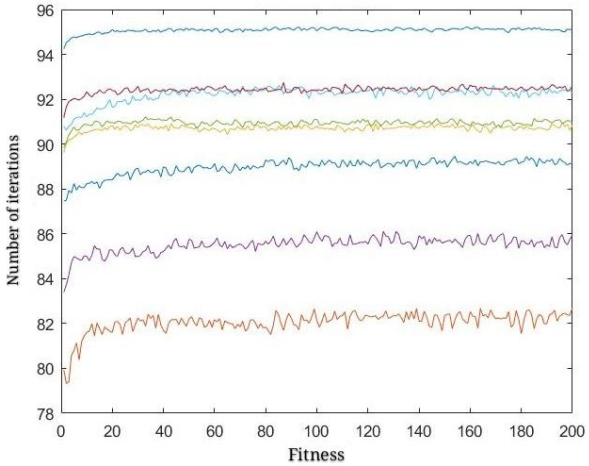

Figure 19. Parameters optimization process of the DWT-MKL-SVM.

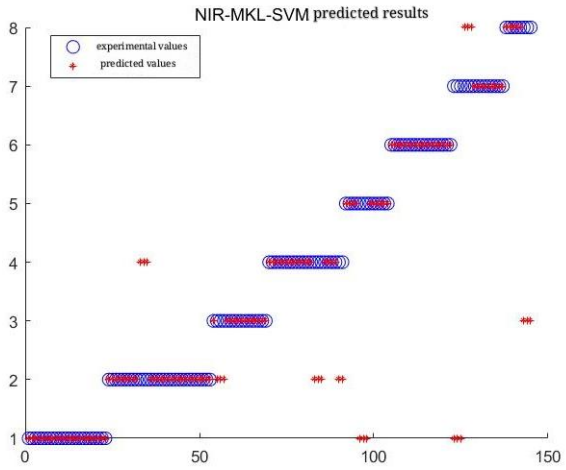

Figure 20. Prediction results of the NIR-MKL-SVM.

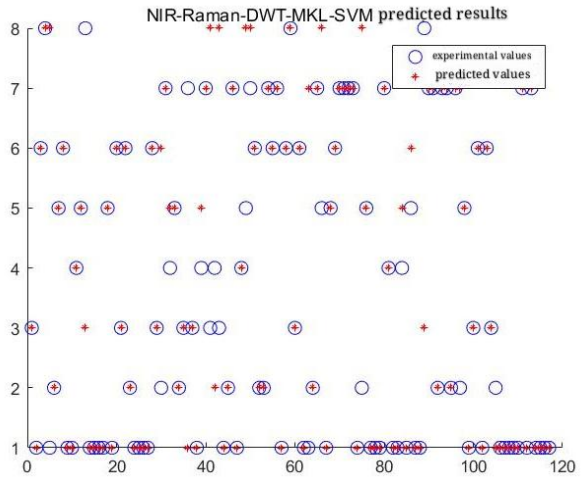

Figure 21. Prediction results of the Raman-MKL-SVM.

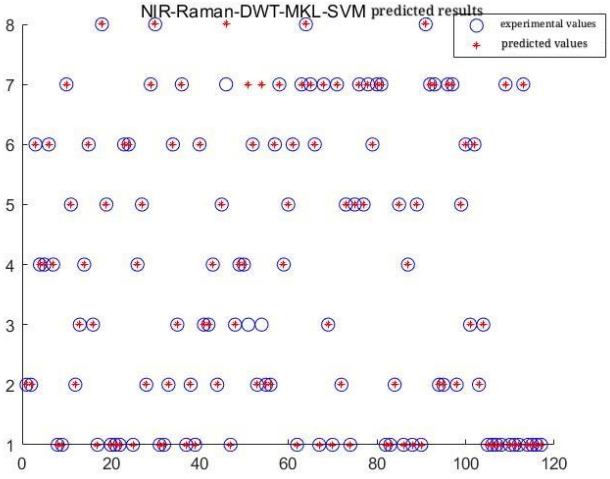

Figure 22. Prediction results of the CARS-MKL-SVM.

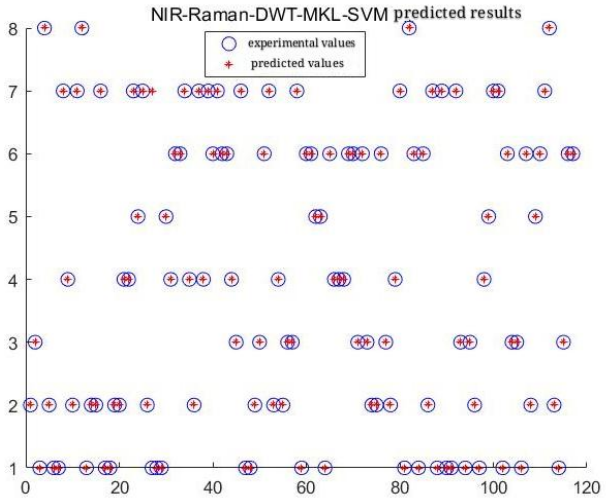

Figure 23. Prediction results of the DWT-MKL-SVM.

As shown in Table 2 and Figures 16–23, the prediction models based on MKL-SVM had better classification performances, and the accuracy was up to 99.15%. In detail, the prediction performance of the NIR-based MKL-SVM classification model was better than that of the Raman-based one. And, the prediction performance of the spectra-fusion-based MKL-SVM classification model was better than that of the single-spectra-based one. It is shown that the fusion of multi-source spectral data can improve the performance of training models for the establishment of multi-kernel learning support vector machine classification models. As shown in Figures 8–15, those models with smaller parameters [C] and [g] had better prediction performances, higher stability, and greater generalization ability.

4. Conclusion

The classification models based on Raman and near infrared spectroscopy combined with the MKL-SVM method can fast identify the authenticity of edible vegetable oil. The highest accuracy of these models was up to 99.15%. The prediction accuracy of the spectra-fusion-based MKL-SVM model was higher than that of the single-spectra-based NIR-MKL-SVM and Raman-MKL-SVM models, indicating that the data fusion can effectively improve the performance of MKL-SVM. Between both spectra-fusion-based models, the wavelet-fusion-based DWT-MKL-SVM had the highest prediction accuracy, the smallest parameters, and greatest generalization ability.

Acknowledgments

2016 Innovation and Transformation of Grain Science and Technology of Hubei Province (20165104); Key Science and Technology Project of Wuhan City (2013010501010147).

References

[1] Wang Ruiyuan. Consumption status of edible vegetable oils in China [J]. Journal of Heilongjiang Grain, 2017 (5): 11-13.
[2] Li Chang, Dan Hao, Wang Xingguo. Review of detection methods for adulteration of edible oils [J]. Agricultural Industrialization, 2007 (5): 30-35.
[3] Chapelle O, Vapnik V, Bousquet O. Choosing Multiple Parameters for Support, Parameters, O.,
[4] Liu Zhiqiang, Jiang Wanlu, Tan Wenzhen, et al. Fault Identification Method for Hydraulic Pumps Based on Multi-feature Fusion and Multiple Kernel Learning SVM [J]. China Mechanical Engineering, 2016, 27 (24): 3355-3361.

[5] Sun Quansen, Zeng Sheng Ji, Yang Maolong, et al. Combined Feature Extraction Based on Canonical Correlation Analysis and Face Recognition [J]. Journal of Computer Research and Development, 2005, 42 (4): 614-621.
[6] Sun Quansen, Zeng Sheng Ji, Wang Pingan, et al. The Theory of Canonical Correlation Analysis and Its Application to Feature Fusion [J]. Chinese Journal of Computers, 2005, 28 (9): 1524-1533.
[7] Wang Zheng. Research on multi-focus image fusion algorithm [D]. Tianjin University, 2008.
[8] Deng Lei, Li Jing, Chen Yunhao, et al. Comparison and Application of Several Wavelet-based Fusion Methods in Remote Sensing Image Fusion [J]. Remote Sensing Information, 2007 (6): 23-27.