

# Brand Identification of Printing Oils and Paste Based on Laser Confocal Raman Spectrometer Raman Spectroscopy and Machine Learning

Zhe Zhang<sup>1\*</sup>

<sup>1</sup> Graduate School of the Chinese People's Police University, Langfang, China

## Email Address

z2391025688@163.com (Zhe Zhang)

\*Correspondence: z2391025688@163.com

Received: 25 July 2024; Accepted: 12 August 2024; Published: 29 August 2024

## Abstract:

To address the key issues in the authentication of seal text, this study focuses on the accurate identification of seal oil brands. Confocal Raman spectroscopy was used to scan the spectra of 16 samples of printing oil and printing clay, and combined with machine learning algorithms for classification and identification. The data quality was enhanced by pre-processing steps such as noise reduction, baseline correction and normalization of the spectral data. Subsequently, the support vector machine (SVM) and random forest (RF) algorithms were utilized to identify the test set, and after parameter tuning, the recognition rates of SVM and RF were improved from 57.63% and 77.08% to 90.27% and 90.97%, respectively, for the original data, which were significantly better than that of the original data. This study shows that the combination of confocal Raman spectroscopy and machine learning provides a new scientific method for the accurate identification of the brand of printing oil and printing clay, which has potential application value and can provide strong support for the document seal inspection and identification of document authenticity.

## Keywords:

Raman Spectroscopy, Machine Learning, Printing Oil and Clay Brand Identification, Seal Authenticity

## 1. Introduction

In recent years, the rapid development of electronic seals, however, due to the relevant laws and regulations are not yet perfect, coupled with the long-established habits of signing paper contracts and cultural background, seal imprints still play an irreplaceable role in economic activities and document review [1]. However, the illegal and criminal activities of forging seals and stamps continue to occur frequently [2], seriously infringing on the interests of citizens and disrupting social order. The traditional identification method based on the morphological characteristics of the seal has been difficult to meet the current rapid and accurate identification needs.

Therefore, through the physical and chemical properties of the printing oil printing mud test, to assist in the identification of the authenticity of the seal appears to be particularly important. At present, the accurate identification of the brand of printing oil, and then assist in the detection of the authenticity of the seal, for the fight against the illegal and criminal activities of forging seals and seals, and maintain the normal order of society has a vital significance [3].

Raman spectroscopy has been widely used in the classification and identification of printing oil and printing clay because of its non-destructive detection and no need for sample preparation, but it still faces many challenges. Specifically, the method needs to be compared with the standard spectral map corresponding to the excitation light source, and the complex composition of the printing oil and printing clay, it is difficult to find a perfect match of the standard spectrum. At the same time, changes in external physical conditions can easily lead to shifts in the absorption peaks, increasing the difficulty of identification [4]. In addition, manual comparison of spectral information is not only time-consuming, but also the reliability is easily affected by subjective factors, which makes the accuracy rate worrying. To overcome these challenges, this study attempts to introduce machine learning algorithms to realize the precise identification of the brand of printing oil and printing clay, and to improve the detection accuracy and efficiency.

Traditional brand identification techniques for printing oil and printing clay mostly rely on chromatography or mass spectrometry, but these methods usually require cumbersome sample pretreatment and complex experimental operations, and they are destructive to the printing color material, which makes it difficult to meet the demand for nondestructive analysis and detection in practical inspection [5]. In contrast, molecular spectroscopy, especially Raman spectroscopy, opens up new possibilities for the identification of the brand of printing oil and printing clay due to its significant advantages such as simple sample preparation, nondestructive detection, and fast analysis speed [6-12].

At present, Raman spectroscopy for ink identification has made some progress, Han used laser microscopic Raman spectroscopy to identify the types of a variety of different brands and models of printing oil stamping samples, and used the Raman standard spectral peaks of the dyes and additives in the printing oils to identify the species of printing oils [13]. Liu et al. used laser Raman spectroscopy to study 10 kinds of printing text colorants, and classified the tested printing text LI used laser Raman spectroscopy to study 10 types of printing pigments and classified the tested printing pigments into 6 categories [14]. Li determined whether the printing pigments were copied and pasted by printers by identifying the absorption peaks of alcohols in the Raman spectra of the printing pigments [15], which provided a new way of thinking about the application of Raman spectroscopy in the authenticity of the printing pigments, but there are still some challenges and limitations in accurately identifying the brand names of the printing oils and printing pastes.

Although Raman spectroscopy shows potential in evaluating the brand of printing oil and printing clay [16,17,18], its application is limited by the difficulty of ordinary Raman instruments to realize the precise selection of the detection area, which leads to the challenge of accurate identification of the imprint area. The emergence of confocal Raman spectrometry provides a fast, nondestructive and accurately selectable identification area detection program for the brand of printing oil and printing clay. However, in practical applications, confocal Raman spectroscopy is

susceptible to instrumental and experimental environment noise interference, baseline drift and spectral overlap, which leads to signal quality degradation and useful fingerprint information is difficult to extract [19]. To solve this problem, the combination of preprocessing algorithms and machine learning techniques is particularly important. By implementing preprocessing algorithms such as noise reduction smoothing and baseline correction, the data quality can be effectively improved [20]. Subsequently, machine learning algorithms such as random forest (RF) and support vector machine (SVM) can be utilized to further extract useful fingerprint information from the preprocessed spectra, improve the accuracy of spectral data analysis, and provide strong support for the accurate identification of the brand of printing oil and printing clay.

The aim of this paper is to investigate the feasibility of Raman spectroscopy combined with machine learning to recognize the brand of printing oil printing clay. By using a confocal Raman spectrometer to prepare and scan the samples of printing oil and printing clay, the acquired spectral data were pre-processed with noise reduction, baseline correction and normalization, and then machine learning models such as SVM and RF were used to classify the spectra of the printing oil and printing clay, and important variables in the RF algorithm that contributed more to the classification model were identified. This study provides a potentially feasible scientific method for future seal authentication.

## 2. Experimental Sample Preparation and Sampling

Experimental samples for the domestic market publicly sold 16 kinds of printing oil printing clay, as shown in Table 1. In the experimental process, printing oil printing clay through the seal stamped to A4 print white paper sheets, in order to control the influence of paper and seal factors, paper for the same brand with the same batch of production, customized production of seals, the material for the rubber, each seal engraved with the same text, in order to avoid contamination of the sample, the seal with a plastic sealing bag sealing package.

*Table 1. Experimental samples.*

| Sample number | Name of ink paste | Sample Model | Sample Type                |
|---------------|-------------------|--------------|----------------------------|
| 1             | QIXING            | 6-S          | quick-drying               |
| 2             | JINYU             | B            | quick-drying               |
| 3             | M&G               | BLS831       | quick-drying               |
| 4             | LIBERTY           | MS-50-S      | quick-drying               |
| 5             | XUEAO             | 0518         | ink paste                  |
| 6             | GongZiPai         | 684          | ink paste                  |
| 7             | Dahua             | 12#6#3#      | ink paste                  |
| 8             | Huhua             | 685          | ink paste                  |
| 9             | UNI               | HSN-S50K     | Photosensitive ink filters |
| 10            | Xima              | 21564        | Photosensitive ink filters |
| 11            | trodat            | 271          | Photosensitive ink filters |
| 12            | deli              | 9870         | atomic ink                 |
| 13            | HuiFeng           | 727#         | atomic ink                 |
| 14            | Arxin             | 013          | atomic ink                 |
| 15            | Xsamper           | CS-10N       | atomic ink                 |
| 16            | KaiYu             | WD-1         | atomic ink                 |

In order to take into account the actual inspection and monitoring process to test the samples must be stamped after the completion of a period of time to carry out, and to

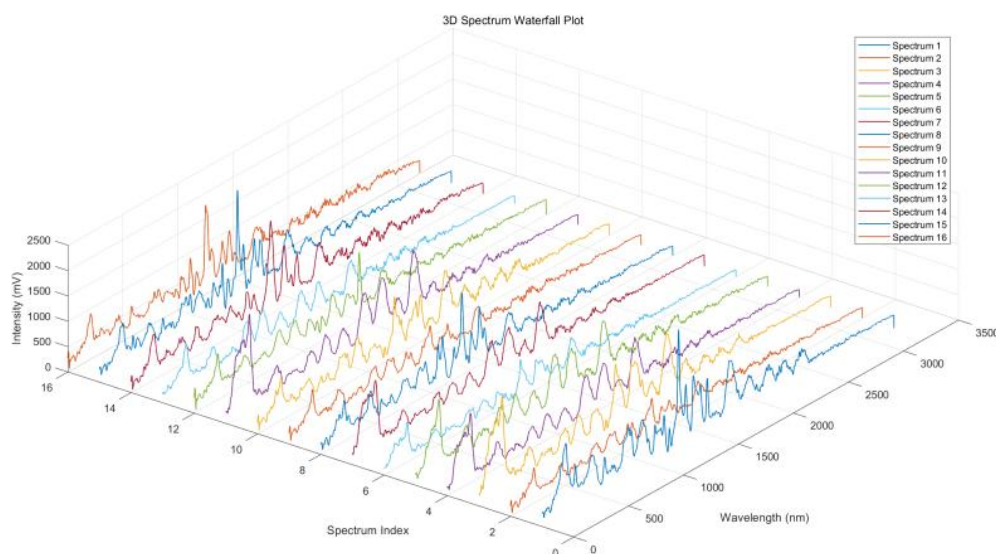
prevent just stamped after the evaporation of water and volatile substances on the experimental results have an impact on the printed paper to avoid light and ventilation preservation of more than 48h before testing. Confocal Raman spectrometer by laser irradiation of the samples to obtain the reflected Stokes and Stokes reflected light to determine the vibration of the molecular chemical bonding, reflected in the peak spectral image of the spectral peaks of the displacement and the intensity of the peaks, which in turn reflects the internal chemical bonding structure of the material, the formation of each sample corresponds to a unique fingerprint. As shown in Figure 1, the FORAM 685-3 detector excitation light source has a wavelength of 685 nm, a maximum excitation power of 45 mW, a spectral resolution of  $7\text{ cm}^{-1}$ , and a receiving light source range of  $200\sim 3200\text{ cm}^{-1}$ . In order to obtain Raman spectra of the samples at random positions, the samples were placed on a three-dimensional moving stage. The laser power for sample testing was uniformly chosen to be 11.25 mW, and the integration time was 1 second. Each collected spectrum was averaged using six scans to obtain stable, high signal-to-noise experimental results. All experiments were performed at room temperature and pressure. Thirty samples were collected for each sample from randomly selected sampling areas.



**Figure 1.** Raman spectrometer FORAM 685-3.

### 3. Preprocessing of Raman Spectroscopy Data

The Raman spectra of the collected samples are shown graphically in Figure 2.

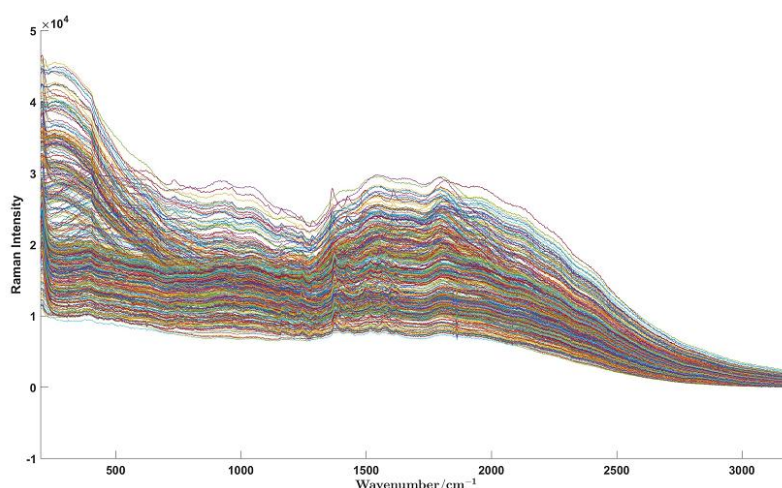


**Figure 2.** Waterfall plots of Raman spectra corresponding to the collection of 16 samples.

Savitzky-Golay (SG) filtering algorithm has been widely used in dealing with noise in recent years, MA Awal in the study of different filtering algorithms to solve the

problem of noise contamination of electrocardiogram (ECG) signals, SG filtering algorithm shows better performance [21], and compared with other algorithms, SG filtering algorithm can retain the original attributes of the signal as much as possible, and the smoothness and noise reduction performance can be optimized by adjusting the fitting order and window size. The performance of noise reduction can be optimized by adjusting the fitting order and window size. Therefore, the SG filtering algorithm can be adjusted according to different application requirements to achieve the best denoising effect. Zhang et al. used an automatic polynomial fitting algorithm based on adaptive iteration-adaptive iterative penalized least squares (airPLS)-to pre-process the Raman spectral data of the cervical adenocarcinoma and squamous carcinoma tissues. were preprocessed, and after comparing the combination of various spectral preprocessing methods and classification models, the results highlighted the important role of airPLS in improving the diagnostic accuracy [22]. The best method was airPLS preprocessing and this combination increased the diagnostic accuracy to 96.3%. We used the SG filtering algorithm and airPLS as the Raman spectral preprocessing technique for acquisition based on the previous research mentioned above.

The obtained Raman spectral data were processed using the SG filtering method with smoothing noise reduction and airPLS baseline correction, the SG filtering window size was 7, and the fitting order was selected as 6. The results before and after the processing are shown in Figure 3 and Figure 4.



**Figure 3.** Original Raman spectrum.

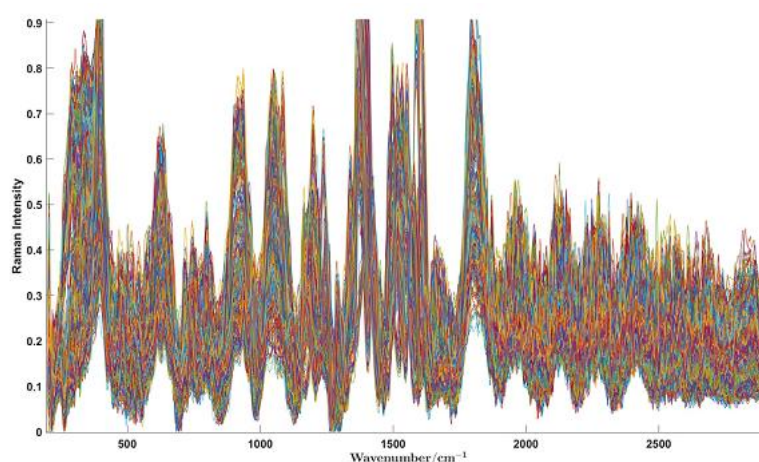


Figure 4. Raman spectrum after preprocessing.

4. Results and Analysis

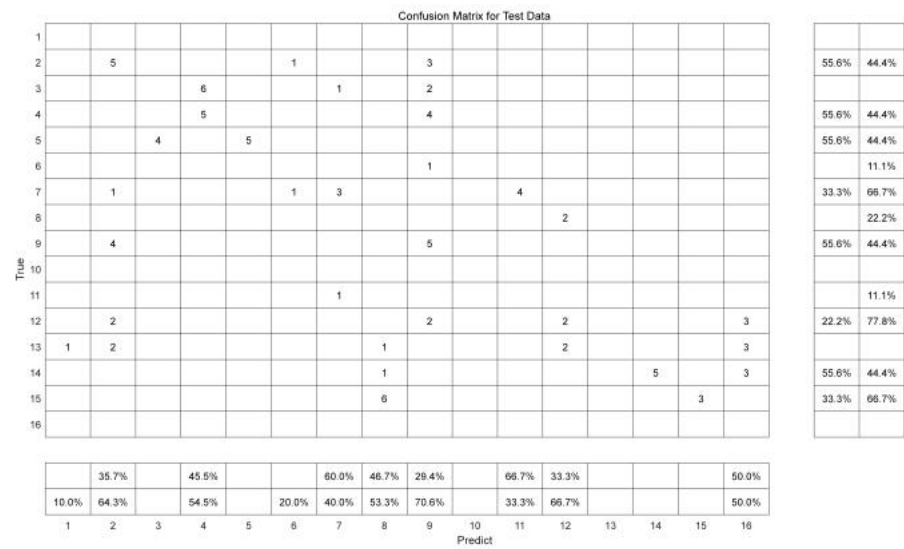


Figure 5. Test recognition rate corresponding to each brand for Raman raw data.

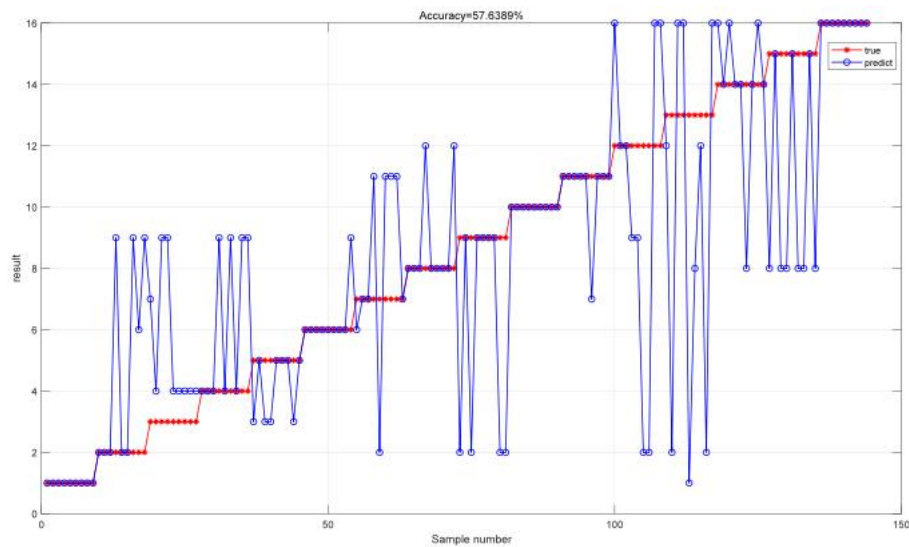
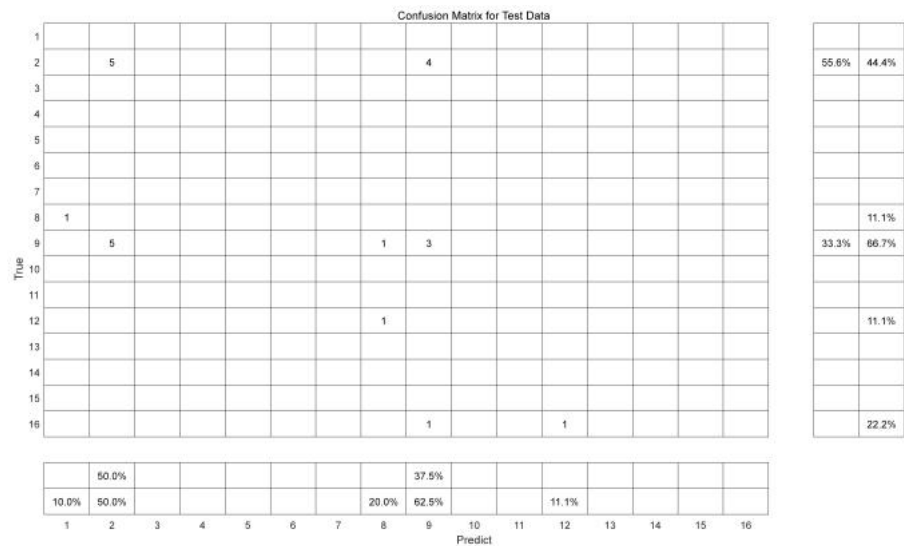
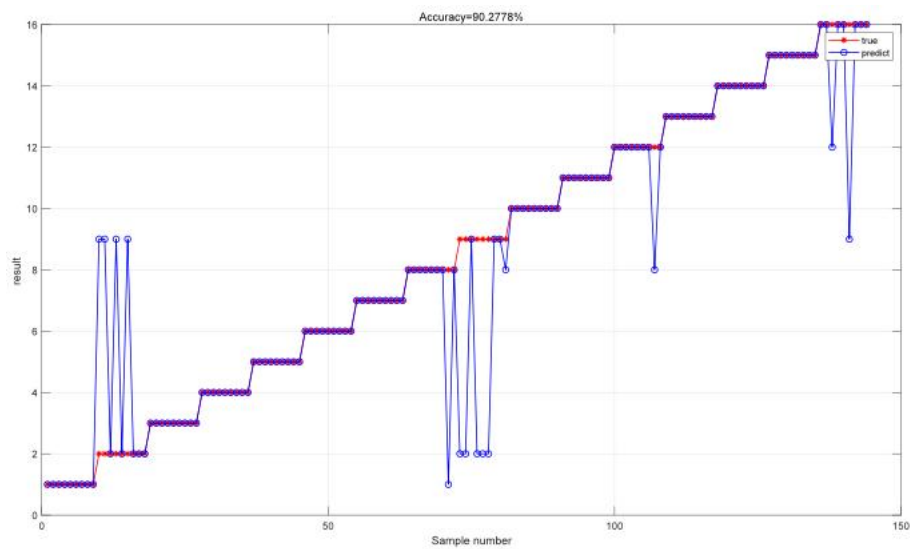


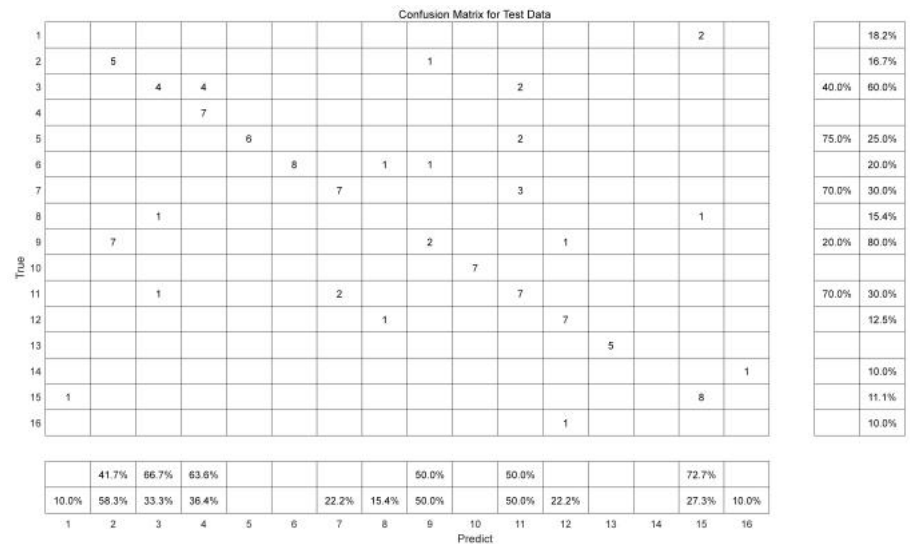
Figure 6. SVM algorithm corresponding to the test set of Raman raw data recognition rate.



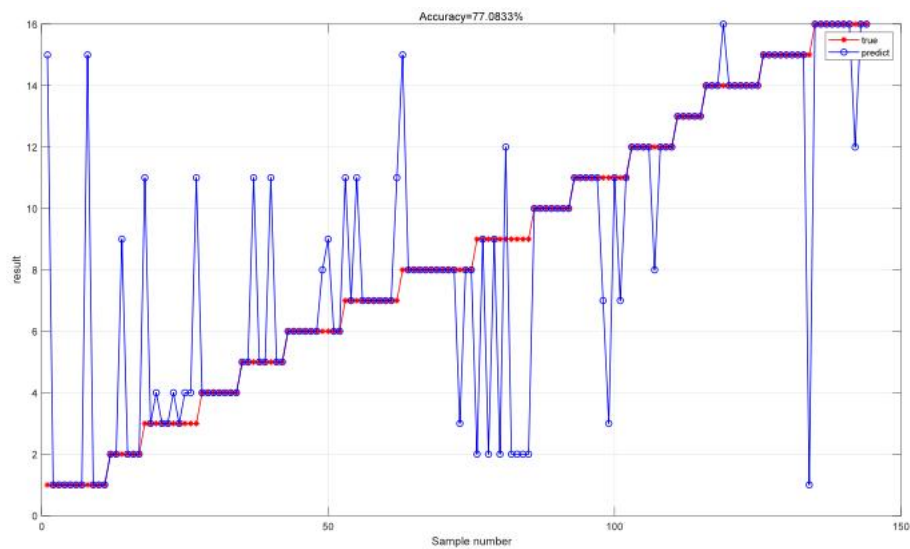
**Figure 7.** Test recognition rate corresponding to each brand for Raman raw data.



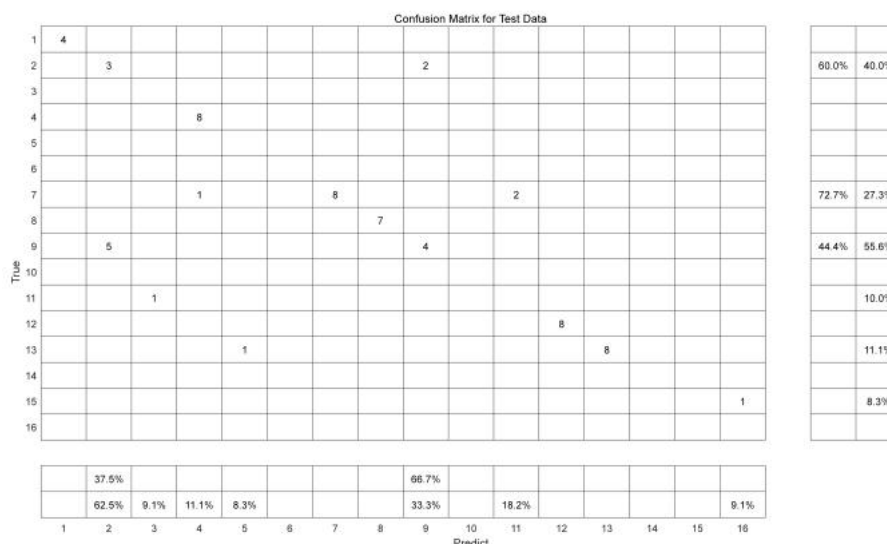
**Figure 8.** SVM algorithm corresponding to Raman preprocessed data test set recognition rate.



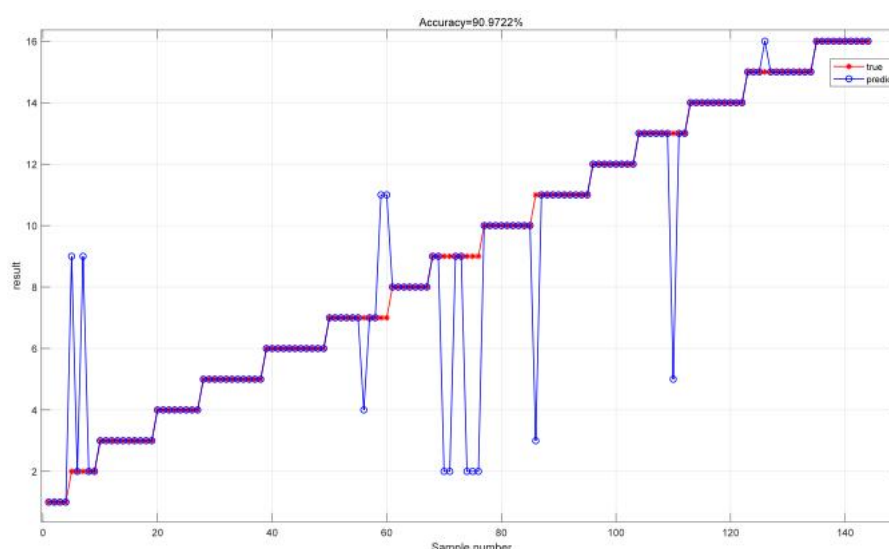
**Figure 9.** RF algorithm corresponding to the original data each brand recognition rate.



**Figure 10.** RF algorithm corresponding to Raman raw data test set recognition rate.



**Figure 11.** RF algorithm corresponding to Raman preprocessing data for each brand recognition rate.



**Figure 12.** Test set recognition rate of RF algorithm corresponding to Raman preprocessed data.

In this study, the samples are labeled according to the brand serial number and divided into test set and training set in the ratio of 7:3. After erasing the corresponding labels of the test set, we use Support Vector Machine (SVM) and Random Forest (RF) algorithms to detect the recognition effect on the test set, respectively. If the recognition result matches with the real data categories, it is recognized correctly, otherwise it is recognized incorrectly. After parameter tuning, the best recognition effect of the test set under the SVM algorithm corresponds to a penalty parameter C of 1 and a kernel function coefficient g of 10; while under the RF algorithm, the best recognition effect is achieved when the number of decision trees is 15 and the minimum number of samples of leaf nodes min\_samples\_leaf is 1. The results show that after the preprocessing step, the recognition rate of SVM on the spectral data is significantly improved from the original 57.63% to 90.27%, while the RF algorithm also improves the recognition effect on the original data from 77.08% to 90.97%.

## 5. Conclusions and Outlook

By combining Raman spectroscopy with machine learning algorithms, this study realizes the accurate recognition of the brand of the printing color material, which provides a potential and feasible scientific identification means for the future document seal inspection and authenticity identification of documents. However, we also realize that although the current recognition accuracy is already relatively high, there is still some room for improvement. Constrained by the information limitations of single spectra, new technological approaches need to be further explored and researched in the future to enhance the recognition accuracy. The future research will therefore focus on the exploration and application of spectral fusion technology, aiming to provide more accurate and reliable technical support for the identification of printing color brands by fusing information from different spectra.

## Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

## Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## References

- [1] Li, B. Dating of Seals Produced with Stamp-Pad Ink Using Gas Chromatography Method. *Journal of Forensic Sciences*, 2014, 59(5), 1403-9.
- [2] Styazhkina, S.A. Illegal Turnover of Official Documents: Criminal Legal Aspect. *Bulletin of Udmurt University Series Economics and Law*, 2019, 29(6), 892-8.
- [3] Song, J.H. Exploration of Judicial Practice Issues on the Crime of Counterfeiting Company Seal. *China Prosecutor*, 2013, 14, 3.
- [4] Gál, L.; Oravec, M.; Gemeiner, P., et al. Principal component analysis for the forensic discrimination of black inkjet inks based on the Vis-NIR fibre optics reflection spectra. *Forensic Sci Int*, 2015, 257, 285-92.
- [5] Saini, K.; Rathore, R. Differentiation of Red and Black Ballpoint Pen Inks using High Performance Thin Layer Chromatography and Gas Chromatography-Mass Spectrometry. *Arab Journal of Forensic Sciences and Forensic Medicine*, 2018, 1(7), 829-41.
- [6] Oravec, M.; Sasinková, V.; Tomanová, K., et al. In-situ surface-enhanced Raman scattering and FT-Raman spectroscopy of black prints. *Vibrational Spectroscopy*, 2017, 94, 16-21.
- [7] Jan, A. Changes in composition of ballpoint pen inks on aging in darkness. *Journal of Forensic Sciences*, 2002, 47(2), 324-7.
- [8] Samanidou, V.F.; Nikolaidou, K.I.; Papadoyannis, I.N. Development and Validation of a Gradient-HPLC-PDAD Method for the Identification of Ballpoint Pen Ink Components: Study of Their Decomposition on Aging for Forensic Science Applications. *Journal of Liquid Chromatography & Related Technologies*, 2004, 27(2), 215-35.

- [9] Fikiet, M.A.; Khandasammy, S.R.; Mistek, E, et al. Surface enhanced Raman spectroscopy: A review of recent applications in forensic science. *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy*, 2018, S138614251830163X.
- [10] Tian, K.; Wang, W.; Yao, Y, et al. Rapid identification of gutter oil by detecting the capsaicin using surface enhanced Raman spectroscopy. *Journal of Raman Spectroscopy*, 2018.
- [11] Colomban, P. The on-site/remote Raman analysis with mobile instruments: a review of drawbacks and success in cultural heritage studies and other associated fields. *Journal of Raman Spectroscopy*, 2012, 43(11), 1529-35.
- [12] Colomban, P.; Treppoz, F.O. Identification and differentiation of ancient and modern European porcelains by Raman macro - and micro - spectroscopy. *Journal of Raman Spectroscopy*, 2001, 32(2), 93-102.
- [13] Han, W.; Huang, J.T.; Wang, H. Identification of printing clay and printing oil types using Raman spectroscopy. *Journal of Chinese People's Public Security University: natural science edition*, 2016, 2, 5.
- [14] Identification of printing clay and printing oil types using Raman spectroscopy. *Journal of Chinese People's Public Security University: natural science edition*, 2016, 2, 5.
- [15] Li, W.; Li, X.; Yang, T, et al. Detection of saliva morphine using surface-enhanced Raman spectroscopy combined with immunochromatographic assay. *Journal of Raman Spectroscopy*, 2020, 51.
- [16] Oravec, M.; Sasinková, V.; Tomanová, K, et al. In-situ surface-enhanced Raman scattering and FT-Raman spectroscopy of black prints. *Vibrational Spectroscopy*, 2017, 94, 16-21.
- [17] Jan, A. Changes in composition of ballpoint pen inks on aging in darkness. *Journal of Forensic Sciences*, 2002, 47(2), 324-7.
- [18] Braz, A.; López-López, M.; García-Ruiz, C. Raman spectroscopy for forensic analysis of inks in questioned documents. *Forensic Science International*, 2013, 232(1-3), 206-12.
- [19] Carlos, L.J.O.; De, L.C.J.; Moreira, L.M, et al. Analysis of the alteration in the optical configuration of Raman spectrometer: Optimization of signal-to-noise ratio (SNR) in a specific wavelength range of clinical interest. *Spectroscopy*, 2008, 22(6), 467-74.
- [20] Ryabchykov, O.; Guo, S.; Bocklitz, T. Analyzing Raman spectroscopic data. *Physical Sciences Reviews*, 2018.
- [21] Awal, M.A.; Mostafa, S.S.; Ahmad, M. Performance Analysis of Savitzky-Golay Smoothing Filter Using ECG Signal. 2011.
- [22] Zhang, H.; Cheng, C.; Gao, R, et al. Rapid identification of cervical adenocarcinoma and cervical squamous cell carcinoma tissue based on Raman spectroscopy combined with multiple machine learning algorithms. *Photodiagnosis and Photodynamic Therapy*, 2021, 33(21), 102104.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)