

## Article

# Efficient Identification of Crude Oil via Combined Terahertz Time-Domain Spectroscopy and Machine Learning

Fan Yang, Huifang Ma \*, Haiqing Huang and Dehua Li \*

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 202182130038@sdust.edu.cn (F.Y.)

\* Correspondence: mahuifang@sdust.edu.cn (H.M.); jcbwl@sdust.edu.cn (D.L.)

**Abstract:** The quality of crude oil varies significantly according to its geographical origin. The efficient identification of the source region of crude oil is pivotal for petroleum trade and processing. However, current methods, such as mass spectrometry and fluorescence spectroscopy, suffer problems such as complex sample preparation and a long characterization time, which restrict their efficiency. In this work, by combining terahertz time-domain spectroscopy (THz-TDS) and a machine learning analysis of the spectra, an efficient workflow for the accurate and fast identification of crude oil was established. Based on THz-TDS of 83 crude oil samples obtained from six countries, a machine learning protocol involving the dimension reduction of spectra and classification was developed to identify the geological origins of crude oil, with an overall accuracy of 96.33%. This work demonstrates that THz spectra combined with a modern numerical scheme analysis can be readily employed to categorize crude oil products efficiently.

**Keywords:** crude oil; terahertz time-domain spectroscopy; principal component analysis; convolution neural network

## 1. Introduction

Crude oil is a complex mixture consisting of a variety of compounds [1]. Alongside liquid hydrocarbons like alkanes, cycloalkanes, and aromatic hydrocarbons, it also contains compounds containing nitrogen, sulfur, oxygen, and other elements, including hydrogen sulfide, quinoline, and phenols. Additionally, crude oil contains trace metals such as nickel, iron, and vanadium [2–4]. The composition and properties of crude oil are greatly influenced by the formation environment, resulting in variations in the quality and commercial value of crude oil from across the globe [5,6]. Rapid and accurate source identification of crude oil holds significant importance for crude oil trading and subsequent processing.

The existing methods for crude oil identification mainly include mass spectrometry methods such as gas chromatography–mass spectrometry (GC-MS), inductively coupled plasma–mass spectrometry (ICP-MS), Fourier transform–ion cyclotron resonance–mass spectrometry (FT-ICR MS), atmospheric pressure photoionization Fourier transform–ion cyclotron resonance mass spectrometry (APPI FT-ICR MS) [7–10], fluorescence spectroscopy [11–13], Raman spectroscopy [14,15], X-ray analysis [16,17], and infrared spectroscopy [18–20], etc. However, these methodologies exhibit distinct limitations. Mass spectrometry requires complex sample pretreatment, has a long experimental process, and yields complex fingerprint spectra [21], making it difficult to achieve classification analysis of multitudinous samples. Fluorescence spectroscopy fluctuates due to interference from scattered light, mutual interference between elements, and overlapping peaks. Raman spectroscopy suffers from weak Raman signal intensity and a low signal-to-noise ratio. X-ray spectroscopy is hazardous to human health and not extensively applicable. Currently, the most widely used methods in the field of crude oil detection and identification are attenuated total reflection infrared spectroscopy (ATR-IR) and Fourier transform attenuated



**Citation:** Yang, F.; Ma, H.; Huang, H.; Li, D. Efficient Identification of Crude Oil via Combined Terahertz Time-Domain Spectroscopy and Machine Learning. *Photonics* **2024**, *11*, 155. <https://doi.org/10.3390/photonics11020155>

Received: 21 December 2023

Revised: 24 January 2024

Accepted: 3 February 2024

Published: 6 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

total reflection infrared spectroscopy (ATR-FTIR), which are affected by uncertainties such as interfering molecules like resins in the obtained spectral range [22].

As a subfield of crude oil identification, crude oil classification most widely adopts the American Petroleum Institute (API) gravity method. Based on API gravity, crude oil can be classified into three categories: “light”, “medium”, and “heavy”. However, the use of standardized experiments to obtain data for crude oil analysis cannot be applied to geographical origin recognition. In order to achieve the faster and more efficient classification of crude oil, the combination of spectroscopic techniques with chemometrics or neural networks has attracted more and more scientific interest. As is shown in Table 1, the most commonly used methods for crude oil classification are a combination of mass spectrometry, ATR-FTIR, and chemometrics. However, inherent defects are inevitable in these methods, such as complex sample pretreatment and low sensitivity. In order to overcome these shortages and expand the application of spectroscopic techniques in the field of crude oil classification, this study employed terahertz time-domain spectroscopy (THz-TDS) to measure crude oil samples.

**Table 1.** Applications of spectroscopic techniques: spectroscopic techniques combined with chemometrics or neural network methods in crude oil classification.

Type of Classification	Instrumental Methods	Chemometric Methods	Neural Networks	Ref.
Characteristics of trace elements	ICP-MS, GS-MS	CA <sup>1</sup>	-	[9]
API gravity	FT-ICR MS	-	-	[10]
API gravity	ATR-FTIR	PCA, PLS-DA <sup>2</sup>	-	[19]
Fuel and crude oil types	GS-MS	COW-PCA-LDA <sup>3</sup>	-	[21]
Origin of crude oils	ATR-FTIR	PCA, SIMCA <sup>4</sup>	-	[23]
Geographical origin	GS-MS	-	Kohonen self-organizing maps	[24]
Well and geographical origin	APPI FT-ICR MS	PCA, HCA <sup>5</sup>	-	[25]
Certified reference materials	GC-MS	MPCA <sup>6</sup> , PARAFAC <sup>7</sup>	-	[26]

<sup>1</sup> CA: cluster analysis. <sup>2</sup> PLS-DA: partial least squares-discriminant analysis. <sup>3</sup> COW-PCA-LDA: correlation optimized warping plus principal component analysis plus linear discriminant analysis. <sup>4</sup> SIMCA: soft independent modeling of class analogies. <sup>5</sup> HCA: hierarchical clustering. <sup>6</sup> MPCA: multi-way principal components analysis. <sup>7</sup> PARAFAC: principal factors analysis.

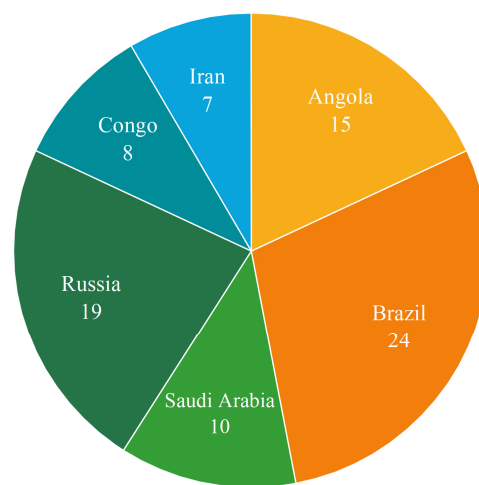
THz-TDS is a spectral measurement technique based on ultrafast femtosecond laser pulses [27]. Due to the high transmittance of THz waves in crude oil, it can measure the amplitude attenuation and time delay of the incident terahertz pulse without sample pretreatment, making it operationally simple. Additionally, the THz-TDS can be applied to analyze the compound structures and intermolecular interaction among organic molecules (such as hydrogen bonding, Van der Waals forces, dipole rotation, and vibrational transitions) [28–31]. It has been demonstrated that THz-TDS is applicable in qualitative and quantitative analysis of the crude oil properties [32,33], but its application in crude oil classification has not yet been explored.

THz-TDS with a frequency range from 0.2 to 2.5 THz was employed in this study to measure the terahertz time-domain spectra of 83 types of crude oil from six countries: Angola, Brazil, Saudi Arabia, Russia, Congo, and Iran. The refractive index spectra and absorption coefficient spectra of these crude oils were extracted and subjected to principal component analysis (PCA). Subsequently, the principal component of the refractive index spectra and absorption coefficient spectra were input into a convolutional neural network (CNN) to classify crude oils of different origins. The method of THz-TDS combined with PCA and CNN shows the new application of THz-TDS combined with machine learning in crude oil origin classification, which provides a new solution for the on-site classification of crude oil origin.

## 2. Experiment

### 2.1. Spectral Acquisition

The TDS-1008 system (BATOP GmbH, Jena, Germany, Figure S1), driven by a Mai Tai femtosecond laser (Spectra-Physics, Milpitas, CA, USA, with a wavelength of 780–980 nm, a pulse width of 80 fs, and an output power of 1.5–2.5 W) was used to conduct the THz-TDS measurements. The time domain signals were Fourier transformed to the frequency range of 0.2–2.5 THz, with a signal-to-noise ratio higher than 65 dB. The poly-4-methyl-1-pentene (TPX) cell adopted in the experiment had dimensions of 46 mm × 46 mm × 3 mm (length, width, and thickness, respectively). Every sample in the cell was placed into the THz-TDS chamber for measurement, and then when switching to a new sample, the previous sample was taken out, and the cell was cleaned and dried, and then injected with a new sample. The number of crude oil samples from different countries measured in the experiment is shown in Figure 1.



**Figure 1.** Pie chart of crude oil sample quantity in different countries.

In addition, the refractive index and absorption coefficient of the sample are given by [34,35]

$$n(\omega) = \left[ \frac{\varphi(\omega)c}{\omega d} \right] + 1 \quad (1)$$

$$\alpha(\omega) = \frac{2}{d} \ln \left[ \frac{4n(\omega)}{\rho(\omega)[n(\omega) + 1]^2} \right] \quad (2)$$

where  $\omega$  is the angular frequency;  $c$  is the velocity of light in vacuum;  $d$  is the sample thickness; and  $\varphi(\omega)$  and  $\rho(\omega)$  are the phase difference and amplitude ratio between the reference signal and the sample signal, respectively.

### 2.2. Principal Component Analysis (PCA)

PCA is a multivariate statistical analysis method that has been widely applied to deal with multivariable data, as it is capable of reducing the dimensionality of the data [33]. PCA transforms the original data into a set of linearly independent representations of each dimension via linear transformation, while retaining as much information present in the original dataset as possible [21]. By applying linear transformation, the original data variables are converted into a new set of variables known as principal components (PCs). Each principal component (PC) contains two types of information: scores and loadings. Scores represent the projection values of the original data on each PC, while loadings represent the weight information between the original variables and each PC [36,37]. The variance contribution rate is used to measure the importance of PCs, indicating the

proportion of variance accounted for by a particular PC. The calculation formula for the variance contribution rate  $R_i$  is as follows:

$$R_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}, i = 1, 2, \dots, p \tag{3}$$

where  $R_i$  represents the eigenvalues (i.e., variances) obtained through eigenvalue decomposition of the covariance matrix of the original data and  $i$  indicates the  $i$ -th PC. PCs with higher variance contribution rates have a greater degree of explanation for the original data. The cumulative variance contribution rate  $CR_i$  refers to the proportion of the total variance of the first  $i$  PCs, given by

$$CR_i = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k}, i = 1, 2, 3, \dots, p. \tag{4}$$

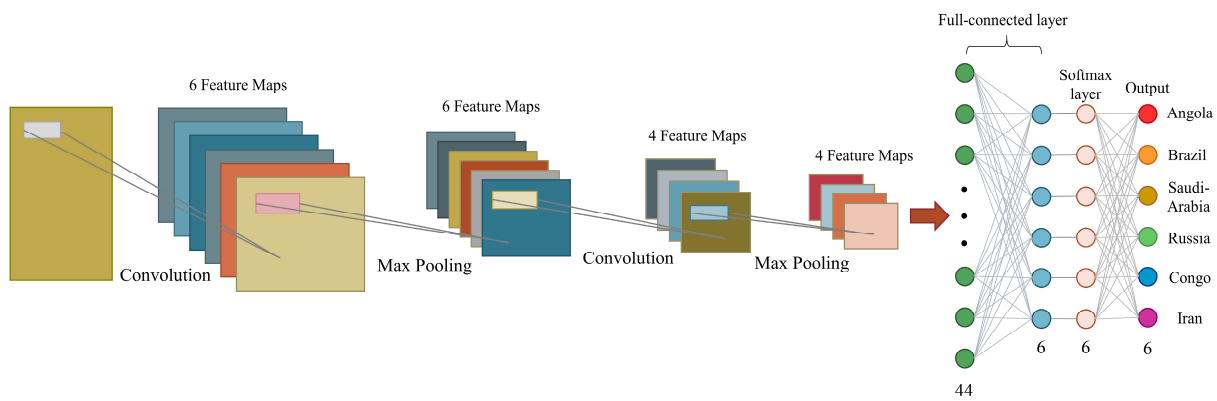
### 2.3. Convolutional Neural Networks (CNNs)

CNNs, as one of the representative algorithms in deep learning, have been involved in remarkable achievements in various fields such as image recognition [38] and spectral recognition [39,40]. CNNs typically consist of convolutional layers, pooling layers, and fully connected layers. Generally, non-linear activation functions are added between the convolutional and pooling layers in CNNs to process the linear output data from the convolutional layers non-linearly, thereby enhancing the network’s representational capacity.

The two-dimensional CNN structure used in this study is illustrated in Figure 2. It comprises an input layer, two convolutional layers, two pooling layers, two fully connected layers, a SoftMax layer, and an output layer. The activation function applied in each convolutional layer is the exponential linear unit (ELU). This definition is shown in Equation (5):

$$ELU(x) = \begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases} \tag{5}$$

where  $\alpha$  is an adjustable parameter, usually with a value of 1.



**Figure 2.** CNN structural framework.

The first convolutional layer consists of 6 convolutional kernels (or filters), while the second convolutional layer consists of 4 convolutional kernels. Both layers have a kernel size of  $2 \times 1$  and a stride of 1. The input layer has a matrix size of  $11 \times 1 \times 1$ , and the pooling layer has a matrix size of  $2 \times 1$  and a stride of 1. The cross-entropy loss function (also known as the SoftMax function) is introduced to convert the model output into a

classification probability, which measures the discrepancy between the output  $q$  and the true labels  $p$  for each class, and is given by

$$Loss = - \sum_{j=1}^m p_j \log(q_j) \tag{6}$$

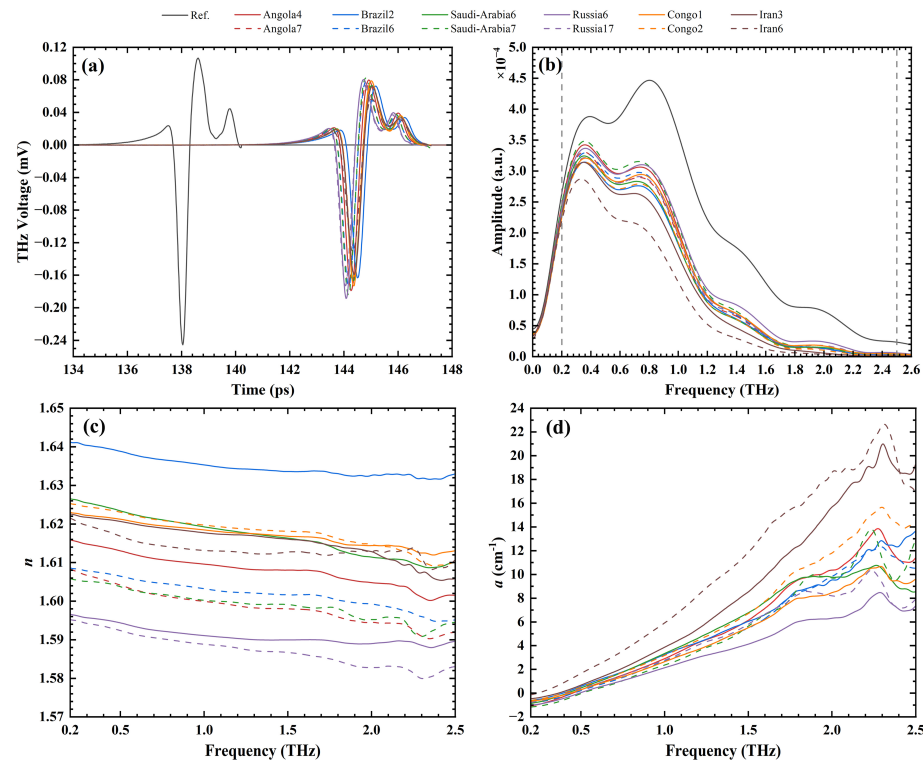
where  $m$  is the number of categories.

The Adam optimizer with an initial learning rate of 0.001 and a learning rate decay factor of 0.1 was used to train the model, and the trainable parameters were updated by minimizing the cross-entropy loss function in the training phase. The maximum number of iterations was set to 450, and at iteration 400, the learning rate was reduced to 0.0001. Furthermore, an L2 regularization parameter with a value of 0.0001 was introduced.

### 3. Results and Discussion

#### 3.1. Terahertz Spectra of Different Origin Crude Oils

In Table S1, the origins and names of all crude oil samples are presented, along with their assigned numbers. Figure 3a,b displays the spectra of a subset of samples (two samples per origin) in the time and frequency domain, respectively. The complete THz spectroscopy for the crude oil samples is exhibited in Figures S2 and S3. In the figures, the solid black line represents the reference signal, while each color corresponds to a specific origin. The reference signal is the terahertz time-domain waveform of a terahertz wave passing through an empty sample cell. Solid and dashed lines of the same color distinguish different samples from the same origin.



**Figure 3.** Terahertz spectra of crude oil samples (two samples of each origin were taken randomly). (a) Time-domain spectra of crude oil samples; (b) frequency spectra of crude oil samples; (c) refractive index spectra of crude oil samples; (d) absorption coefficient spectra crude oil samples.

Figure 3a shows the time delays and amplitude attenuations of the signals from different crude oil samples. The time delay and amplitude magnitude of the sample signals are influenced by the asphaltene content in the crude oil [22]. A lower asphaltene content results in a reduced time delay and amplitude attenuation transmitted terahertz pulse. The

water content in the crude oil is another factor influencing the time delay and amplitude magnitude of terahertz waves [32]. Increased water content in the crude oil leads to greater time delay and amplitude attenuation in transmitted terahertz waves. Furthermore, the time delay and amplitude attenuation of the transmitted terahertz wave also fluctuate due to the complexity of crude oil composition and the presence of trace metals, sulfur, and nitrogen [40]. As shown in Figure 3b–d, the spectral features of different crude oil samples differ greatly in the range of 0.2–2.5 THz. Therefore, spectral data in the range of 0.2–2.5 THz were subjected to principal component analysis.

### 3.2. PCA-CNN Classification Model

In this study, PCA was adopted to investigate the refractive index spectrum and absorption coefficient spectrum of crude oil. The matrix was of the scale  $83 \times 1150$  for both spectra, in which each row represented a sample and each column represented data points at various frequencies. PCs with a variance contribution rate exceeding 0.03% and the cumulative variance contribution rate of the total PCs more than 99.92% were chosen as the variables for the analysis. The variance contribution rates and cumulative variance contribution rates of the principal components of the refractive index spectra and absorption coefficient spectra are shown in Tables 2 and 3, respectively.

**Table 2.** Variance contribution rate and cumulative variance contribution rate of the first three PCs of the refractive index spectra.

Principal Component	Variance Contribution Rate/%	Cumulative Variance Contribution Rate/%
PC 1	99.688	99.688
PC 2	0.258	99.946
PC 3	0.033	99.979

**Table 3.** Variance contribution rate and cumulative variance contribution rate of the first eight PCs of the absorption coefficient spectra.

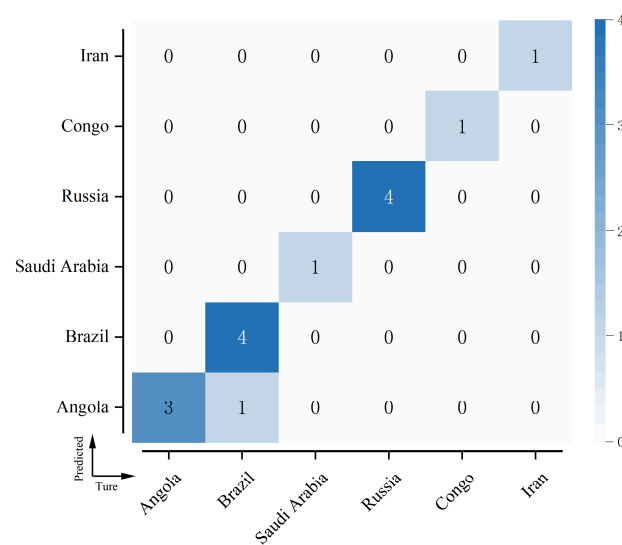
Principal Component	Variance Contribution Rate/%	Cumulative Variance Contribution Rate/%
PC 1	74.485	74.485
PC 2	22.301	96.786
PC 3	2.069	98.855
PC 4	0.647	99.503
PC 5	0.239	99.742
PC 6	0.099	99.841
PC 7	0.052	99.893
PC 8	0.035	99.928

By using PCA, the 2300 original spectral data including 1150 variables of the refractive index and absorption coefficient spectra separately were reduced to 11 variables including the first three PCs of the refractive index spectra and the first eight PCs of the absorption coefficient spectra. To eliminate the impact of differences in range of values between variables, the first three PCs of the extracted refractive index spectra and the first eight PCs of the absorption coefficient spectra were normalized. Subsequently, these normalized components were fed into a CNN for training and testing and the normalization formula is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

where  $X$  is the absolute value,  $X_{min}$  is the minimum value, and  $X_{max}$  is the maximum value of the dataset.

For countries for which the number of crude oil samples was less than or equal to 10, one sample was randomly chosen for testing, while the remaining samples were the training set. For countries with more than 10 crude oil samples, the training and test sets were divided at a ratio of 4:1. The k-fold cross-validation with  $k = 5$  was implemented to validate the model. In this study, the training set data were divided into five mutually exclusive subsets utilizing fivefold cross-validation, among which one was selected for model verification and the remaining four were used for training in each iteration. This process was repeated five times, with a different subset acting as the validation set each time. The average classification accuracy of the five iterations was 97.14%. The confusion matrix of the test set, depicted in Figure 4, serves as a visual comparison between the actual category and the model prediction category. The figure reveals that only one sample was misclassified. The experiment was carried out 20 times and the average classification accuracy was 96.33%.



**Figure 4.** Confusion matrix results for the test set.

#### 4. Conclusions

A methodology is proposed for categorizing the source of crude oil by employing THz spectroscopy in conjunction with PCA and CNN. To establish a THz spectroscopy dataset, THz-TDS technology was adopted to measure the spectra of 83 distinct crude oil samples from six countries. The PCA technique was then employed to reduce dimensionality and extract features from the refractive index spectra and absorption coefficient spectra of the crude oil samples, resulting in a combined PC dataset. Subsequently, a CNN was utilized for feature learning and classification recognition of the combined PC data. The experimental results demonstrate that the proposed approach, which combined THz-TDS, PCA, and a CNN, achieved a classification accuracy of 96.33% for distinguishing crude oil from different origins. This accomplishment confirms the feasibility and promising prospects of this method in the field of crude oil classification, extending the existing classification methods of crude oil.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/photonics11020155/s1>. Figure S1: Schematic diagram of transmission THz-TDS setup. M1, M2, and M3 are reflectors; L1 and L2 are lenses; and OPM1, OPM2, OPM3, and OPM4 are off-axis parabolic mirrors. Table S1: Country, name, and number of crude oil samples. Figure S2: (a1)–(a12) THz time-domain spectra of crude oils; (b1)–(b12) THz frequency-domain spectra of crude oils. Figure S3: (a1)–(a6) The refractive index spectra of crude oils; (b1)–(b6) the absorption coefficient spectra of crude oils.

**Author Contributions:** Conceptualization, F.Y. and D.L.; methodology, F.Y. and H.H.; software, F.Y.; validation, D.L.; investigation, F.Y. and H.H.; writing—original draft preparation, F.Y.; writing—review and editing, D.L., H.M. and F.Y.; supervision, D.L. and H.M.; project administration, D.L.; funding acquisition, D.L. and H.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (No. 2017YFA0701000), the National Natural Science Foundation of China (Nos, 42072163 and 62305196) and the China Postdoctoral Science Foundation (No, GZC20231498).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within this article and the Supplementary Materials.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Vieira, L.V.; Rainha, K.P.; de Castro, E.V.R.; Filgueiras, P.R.; Carneiro, M.T.W.D.; Brandão, G.P. Exploratory Data Analysis Using API Gravity and V and Ni Contents to Determine the Origins of Crude Oil Samples from Petroleum Fields in the Espírito Santo Basin (Brazil). *Microchem. J.* **2016**, *124*, 26–30. [[CrossRef](#)]
2. López, L.; Lo Mónaco, S. Geochemical Implications of Trace Elements and Sulfur in the Saturate, Aromatic and Resin Fractions of Crude Oil from the Mara and Mara Oeste Fields, Venezuela. *Fuel* **2004**, *83*, 365–374. [[CrossRef](#)]
3. Duyck, C.; Miekeley, N.; Porto da Silveira, C.L.; Szatmari, P. Trace Element Determination in Crude Oil and Its Fractions by Inductively Coupled Plasma Mass Spectrometry Using Ultrasonic Nebulization of Toluene Solutions. *Spectrochim. Acta Part B At. Spectrosc.* **2002**, *57*, 1979–1990. [[CrossRef](#)]
4. Akinlua, A.; Sigidle, A.; Buthelezi, T.; Fadipe, O.A. Trace Element Geochemistry of Crude Oils and Condensates from South African Basins. *Mar. Pet. Geol.* **2015**, *59*, 286–293. [[CrossRef](#)]
5. Ellrich, J.; Hirner, A.; Stärk, H. Distribution of Trace Elements in Crude Oils from Southern Germany. *Chem. Geol.* **1985**, *48*, 313–323. [[CrossRef](#)]
6. Rodrigues, É.V.A.; Silva, S.R.C.; Romão, W.; Castro, E.V.R.; Filgueiras, P.R. Determination of Crude Oil Physicochemical Properties by High-Temperature Gas Chromatography Associated with Multivariate Calibration. *Fuel* **2018**, *220*, 389–395. [[CrossRef](#)]
7. El Nady, M.M.; El-Naggar, A.Y. Occurrences and Distributions of Normal Alkanes and Biological Markers to Detections of Origin, Environments, and Maturation of Crude Oils in El Hamed Oilfield, Gulf of Suez, Egypt. *Energy Sources Part A Recover. Util. Environ. Eff.* **2016**, *38*, 3338–3347. [[CrossRef](#)]
8. Gajdosechova, Z.; Dutta, M.; Lopez-Linares, F.; de Azevedo Mello, P.; Dineck Iop, G.; Moraes Flores, E.M.; Mester, Z.; Pagliano, E. Determination of Chloride in Crude Oil Using Isotope Dilution GC–MS: A Comparative Study. *Fuel* **2021**, *285*, 119167. [[CrossRef](#)]
9. Niu, Z.; Meng, W.; Wang, Y.; Wang, X.; Li, Z.; Wang, J.; Liu, H.; Wang, X. Characteristics of Trace Elements in Crude Oil in the East Section of the South Slope of Dongying Sag and Their Application in Crude Oil Classification. *J. Pet. Sci. Eng.* **2022**, *209*, 109833. [[CrossRef](#)]
10. Santos, J.M.; Wisniewski, A., Jr.; Eberlin, M.N.; Schrader, W. Comparing Crude Oils with Different API Gravities on a Molecular Level Using Mass Spectrometric Analysis. Part 1: Whole Crude Oil. *Energies* **2018**, *11*, 2766. [[CrossRef](#)]
11. Wang, C.; Shi, X.; Li, W.; Wang, L.; Zhang, J.; Yang, C.; Wang, Z. Oil Species Identification Technique Developed by Gabor Wavelet Analysis and Support Vector Machine Based on Concentration-Synchronous-Matrix-Fluorescence Spectroscopy. *Mar. Pollut. Bull.* **2016**, *104*, 322–328. [[CrossRef](#)] [[PubMed](#)]
12. Huang, X.-D.; Wang, C.-Y.; Fan, X.-M.; Zhang, J.-L.; Yang, C.; Wang, Z.-D. Oil Source Recognition Technology Using Concentration-Synchronous-Matrix-Fluorescence Spectroscopy Combined with 2D Wavelet Packet and Probabilistic Neural Network. *Sci. Total Environ.* **2018**, *616*, 632–638. [[CrossRef](#)] [[PubMed](#)]
13. Steffens, J.; Landulfo, E.; Courrol, L.C.; Guardani, R. Application of Fluorescence to the Study of Crude Petroleum. *J. Fluoresc.* **2011**, *21*, 859–864. [[CrossRef](#)] [[PubMed](#)]
14. Flórez, M.A.; Guerrero, J.E.; Cabanzo, R.; Mejía-Ospino, E. SARA Analysis and Conradson Carbon Residue Prediction of Colombian Crude Oils Using PLSR and Raman Spectroscopy. *J. Pet. Sci. Eng.* **2017**, *156*, 966–970. [[CrossRef](#)]
15. Sadeghtabaghi, Z.; Rabbani, A.R.; Hemmati-Sarapardeh, A. Experimental Evaluation of Thermal Maturity of Crude Oil Samples by Asphaltene Fraction: Raman Spectroscopy and X-Ray Diffraction. *J. Pet. Sci. Eng.* **2021**, *199*, 108269. [[CrossRef](#)]
16. Álvarez, E.; Trejo, F.; Marroquín, G.; Ancheyta, J. The Effect of Solvent Washing on Asphaltenes and Their Characterization. *Pet. Sci. Technol.* **2015**, *33*, 265–271. [[CrossRef](#)]
17. Lam-Maldonado, M.; Melo-Banda, J.A.; Macias-Ferrer, D.; Portales-Martínez, B.; Dominguez, J.M.; Silva-Rodrigo, R.; Páramo-García, U.; Mata-Padilla, J.M. Transition Metal Nanocatalysts by Modified Inverse Microemulsion for the Heavy Crude Oil Upgrading at Reservoir. *Catal. Today* **2020**, *349*, 81–87. [[CrossRef](#)]
18. Li, J.; Chu, X.; Tian, S.; Lu, W. The Identification of Highly Similar Crude Oils by Infrared Spectroscopy Combined with Pattern Recognition Method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2013**, *112*, 457–462. [[CrossRef](#)]



19. Mohammadi, M.; Khanmohammadi Khorrami, M.; Vatani, A.; Ghasemzadeh, H.; Vatanparast, H.; Bahramian, A.; Fallah, A. Rapid Determination and Classification of Crude Oils by ATR-FTIR Spectroscopy and Chemometric Methods. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *232*, 118157. [[CrossRef](#)]
20. Huang, H.; Yuan, E.; Zhang, D.; Sun, D.; Yang, M.; Zheng, Z.; Zhang, Z.; Gao, L.; Panezai, S.; Qiu, K. Free Field of View Infrared Digital Holography for Mineral Crystallization. *Cryst. Growth Des.* **2023**, *23*, 7992–8008. [[CrossRef](#)]
21. Sun, P.; Bao, K.; Li, H.; Li, F.; Wang, X.; Cao, L.; Li, G.; Zhou, Q.; Tang, H.; Bao, M. An Efficient Classification Method for Fuel and Crude Oil Types Based on m/z 256 Mass Chromatography by COW-PCA-LDA. *Fuel* **2018**, *222*, 416–423. [[CrossRef](#)]
22. Matoug, M.M.; Gordon, R. Crude Oil Asphaltenes Studied by Terahertz Spectroscopy. *ACS Omega* **2018**, *3*, 3406–3412. [[CrossRef](#)]
23. Garmarudi, A.B.; Khanmohammadi, M.; Fard, H.G.; de la Guardia, M. Origin Based Classification of Crude Oils by Infrared Spectrometry and Chemometrics. *Fuel* **2019**, *236*, 1093–1099. [[CrossRef](#)]
24. Fonseca, A.M.; Biscaya, J.L.; Aires-de-Sousa, J.; Lobo, A.M. Geographical Classification of Crude Oils by Kohonen Self-Organizing Maps. *Anal. Chim. Acta* **2006**, *556*, 374–382. [[CrossRef](#)]
25. Chiaberge, S.; Fiorani, T.; Savoini, A.; Bionda, A.; Ramello, S.; Pastori, M.; Cesti, P. Classification of Crude Oil Samples through Statistical Analysis of APPI FTICR Mass Spectra. *Fuel Process. Technol.* **2013**, *106*, 181–185. [[CrossRef](#)]
26. Fortunato de Carvalho Rocha, W.; Schantz, M.M.; Sheen, D.A.; Chu, P.M.; Lippa, K.A. Unsupervised Classification of Petroleum Certified Reference Materials and Other Fuels by Chemometric Analysis of Gas Chromatography-Mass Spectrometry Data. *Fuel* **2017**, *197*, 248–258. [[CrossRef](#)] [[PubMed](#)]
27. van Exter, M.; Grischkowsky, D.R. Characterization of an Optoelectronic Terahertz Beam System. *IEEE Trans. Microw. Theory Tech.* **1990**, *38*, 1684–1691. [[CrossRef](#)]
28. Zheng, Z.-P.; Fan, W.-H.; Liang, Y.-Q.; Yan, H. Application of Terahertz Spectroscopy and Molecular Modeling in Isomers Investigation: Glucose and Fructose. *Opt. Commun.* **2012**, *285*, 1868–1871. [[CrossRef](#)]
29. Zhan, H.; Chen, M.; Zhao, K.; Li, Y.; Miao, X.; Ye, H.; Ma, Y.; Hao, S.; Li, H.; Yue, W. The Mechanism of the Terahertz Spectroscopy for Oil Shale Detection. *Energy* **2018**, *161*, 46–51. [[CrossRef](#)]
30. Wang, Y.; Ma, H.; Yang, Y.; Qi, J.; Zhang, G.; Ren, H.; Guo, W. First Principles Terahertz Spectroscopy of Molecular Crystals: The Crucial Role of Periodic Boundary Conditions Benchmarked with Experimental L-Ascorbic Acid Spectra. *Phys. Chem. Chem. Phys.* **2023**, *25*, 12252–12258. [[CrossRef](#)]
31. Ma, H.; Yang, Y.; Jing, H.; Jiang, W.; Guo, W.; Ren, H. Semi-Empirical Model to Retrieve Finite Temperature Terahertz Absorption Spectra Using Morse Potential. *Chin. J. Chem. Phys.* **2023**, *36*, 15–24. [[CrossRef](#)]
32. Jin, W.-J.; Zhao, K.; Yang, C.; Xu, C.-H.; Ni, H.; Chen, S.-H. Experimental Measurements of Water Content in Crude Oil Emulsions by Terahertz Time-Domain Spectroscopy. *Appl. Geophys.* **2013**, *10*, 506–509. [[CrossRef](#)]
33. Zhan, H.; Wu, S.; Bao, R.; Ge, L.; Zhao, K. Qualitative Identification of Crude Oils from Different Oil Fields Using Terahertz Time-Domain Spectroscopy. *Fuel* **2015**, *143*, 189–193. [[CrossRef](#)]
34. Duvillaret, L.; Garet, F.; Coutaz, J.-L. Highly Precise Determination of Optical Constants and Sample Thickness in Terahertz Time-Domain Spectroscopy. *Appl. Opt.* **1999**, *38*, 409–415. [[CrossRef](#)] [[PubMed](#)]
35. Dorney, T.D.; Baraniuk, R.G.; Mittleman, D.M. Material Parameter Estimation with Terahertz Time-Domain Spectroscopy. *J. Opt. Soc. Am. A* **2001**, *18*, 1562–1571. [[CrossRef](#)] [[PubMed](#)]
36. Brereton, R.G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2003.
37. Sad, C.M.S.; da Silva, M.; dos Santos, F.D.; Pereira, L.B.; Corona, R.R.B.; Silva, S.R.C.; Portela, N.A.; Castro, E.V.R.; Filgueiras, P.R.; Lacerda, V. Multivariate Data Analysis Applied in the Evaluation of Crude Oil Blends. *Fuel* **2019**, *239*, 421–428. [[CrossRef](#)]
38. Yang, X.; Ye, Y.; Li, X.; Lau, R.Y.K.; Zhang, X.; Huang, X. Hyperspectral Image Classification with Deep Learning Models. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5408–5423. [[CrossRef](#)]
39. Ren, H.; Li, H.; Zhang, Q.; Liang, L.; Guo, W.; Huang, F.; Luo, Y.; Jiang, J. A Machine Learning Vibrational Spectroscopy Protocol for Spectrum Prediction and Spectrum-Based Structure Recognition. *Fundam. Res.* **2021**, *1*, 488–494. [[CrossRef](#)]
40. Qin, F.; Li, Q.; Zhan, H.; Jin, W.; Liu, H.; Zhao, K. Probing the Sulfur Content in Gasoline Quantitatively with Terahertz Time-Domain Spectroscopy. *Sci. China Phys. Mech. Astron.* **2014**, *57*, 1404–1406. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.