chemosensors

MDPI

Article

The Classification of Synthetic- and Petroleum-Based Hydrocarbon Fluids Using Handheld Raman Spectroscopy

Javier E. Hodges, Kailee Marchand, Geraldine Monjardez and Jorn Chi-Chung Yu *

Department of Forensic Science, College of Criminal Justice, Sam Houston State University, Huntsville, TX 77341, USA; jeh064@shsu.edu (J.E.H.); kxm122@shsu.edu (K.M.); gxm073@shsu.edu (G.M.)
* Correspondence: jornyu@shsu.edu; Tel.: +1-(936)-294-4412

Abstract

Hydrocarbon fluids have a widespread presence in modern society due to their role in the global energy and fuel supply. The ability to distinguish between hydrocarbon fluids from different manufacturing processes is essential in industrial and government settings. Currently, performing such analyses is expensive and time-consuming, as standard practice involves sending samples to a laboratory for gas chromatography-mass spectrometry (GC-MS) analysis. The inherent limitations of traditional separation techniques often make them unsuitable for the demands of real-time process monitoring and control. This work proposes the use of handheld Raman spectroscopy for rapid classification of petroleum- and synthetic-based hydrocarbon fluids. A total of 600 Raman spectra were collected from six different hydraulic fluids and analyzed. Preliminary visual observations revealed reproducible spectral differences between various types of hydraulic fluids. Principal component analysis (PCA) and linear discriminant analysis (LDA) were used to investigate the data further. The findings indicate that handheld Raman spectrometers are capable of detecting chemical features of hydrocarbon fluids, supporting the classification of their formulations.

Keywords: hydrocarbons; hydraulic fluids; Raman spectroscopy; principal component analysis (PCA); linear discriminant analysis (LDA)

check for updates

Received: 11 July 2025
Revised: 14 August 2025
Accepted: 25 August 2025
Published: 2 September 2025

Citation: Hodges, J.E.; Marchand, K.; Monjardez, G.; Yu, J.C.-C. The Classification of Synthetic- and Petroleum-Based Hydrocarbon Fluids Using Handheld Raman Spectroscopy. *Chemosensors* **2025**, *13*, 327. https://doi.org/10.3390/chemosensors13090327

Copyright: © 2025 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/).

1. Introduction

Hydrocarbon fluids are very common in modern society. They are used for heating, cooking, drying, and transportation [1]. These materials can be made using petroleum-based or synthetic-based manufacturing processes. While petroleum-based hydrocarbon fluids are refined directly from crude oil distillates, synthetic-based hydrocarbon fluids are manufactured by organic synthesis, enabling these products to achieve their respective purposes more effectively [2]. The major subclasses of synthetic-based hydrocarbon fluids include polyalphaolefin (PAO)-based, polyol ester-based, and phosphate ester-based hydrocarbon fluids, among others. Inter-product variations exist between and within these subclasses due to industrial additives that aim to improve traits including thermal stability, solubility, and fire resistance [2]. Overall, synthetic-based hydrocarbon fluids are chemically engineered for a specific molecular composition with a more tailored and uniform chemical structure [2].

In many scenarios, it is necessary to classify hydrocarbon fluids. For instance, in forensic arson investigations, lab analysts work to classify ignitable liquid residues. Commonly utilized ignitable liquids include a multitude of hydrocarbon fluids [3]. In the fuel industry,

it is essential to classify hydrocarbon fluids to ensure that they comply with governing laws and regulations [4].

To differentiate hydrocarbon fluids produced through different manufacturing processes, analysis methods such as gas chromatography-mass spectrometry (GC-MS) are typically required [5]. GC-MS is a relatively expensive method, requiring the use of tandem instruments, and autosamplers are also used in most cases to ensure data reproducibility [6]. Column carrier gases and large electricity consumption [7] are additional expenses. Cost aside, GC-MS analyses typically require a minimum of 20 min per sample, with some methods requiring even longer runtimes [8–10]. As an alternative, this work analyzed hydrocarbon fluids with different manufacturing processes using a handheld Raman spectrometer.

Handheld Raman spectrometers offer a rapid and field-deployable platform for chemical tests of unknown species. In the case of testing pure chemicals, identification can be achieved by comparing unknown spectra to those collected in a Raman spectral database. In the case of mixture analysis, the chemical signatures of major components may be detected. Thus far, several research groups have demonstrated the use of handheld Raman spectroscopy in forensic science [11–13], medicine [14–16], art [17,18], geology [19,20], and other disciplines. The convenience of handheld Raman spectroscopy has allowed it to be utilized in real-world applications. In this work, we investigated the capability of handheld Raman spectroscopy for classifying hydrocarbon fluid samples.

Previous studies have investigated spectroscopic techniques for analyzing hydrocarbon fluids [4,21–27]. The faster analysis times presented by spectroscopic instruments—less than a minute in some cases—provide an incentive to develop methods to compete with the detailed information output by GC-MS. Raman spectroscopy is especially appealing because it does not require the preparation of liquid samples prior to analysis. Fluorescent signal interference has historically been a key obstacle for the wider implementation of Raman instrumental analysis, especially handheld Raman spectroscopy. However, data preprocessing steps have been demonstrated in the literature to mitigate the issue and allow Raman spectra, both benchtop and handheld, to reveal valuable sample information [4,23].

Chemometrics is a subdiscipline that combines chemistry and statistics and often intersects with machine learning [28–31]. In a 2012 study, benchtop Raman spectroscopy was used to classify three different gasoline products originating from different petroleum refineries [23]. The authors employed chemometric techniques, specifically principal component analysis (PCA) [32] and linear discriminant analysis (LDA). They concluded that LDA was more effective than PCA at sample classification. This is most likely due to LDA being a supervised classification method, which allows the model to analyze examples of the different categories in a training dataset before it is required to make determinations on an unlabeled test dataset [33]. A 2024 study demonstrated the use of a handheld Raman spectroscopy to discriminate between legitimate and illegitimate (designer) diesel fuels [4]. Illegitimate fuels, which contain heavier hydrocarbon fractions, showed spectral differences identifiable by visual observation. The authors supported their analysis with PCA and LDA and reached a similar conclusion regarding LDA's superiority over PCA for hydrocarbon classification tasks.

In the present work, we investigated both chemometric techniques for hydrocarbon classification tasks. While PCA's performance has not matched LDA's in the previous literature studies [4,34], PCA is a non-supervised classification method, and it does not require a training dataset before making determinations on unlabeled data [33]. In the absence of training data, LDA is unusable, making PCA appealing in such situations. Therefore, investigative value can be found in further examinations of PCA's ability to classify hydrocarbons. The primary LDA model in the present work uses the Moore–Penrose

pseudoinverse to handle covariance [35]. A solution for covariance is often necessary in cases such as our own, when the number of predictors (datapoints per spectra) is larger than the number of spectra in each classification group ($p > n$). Shrinkage-regularized linear discriminants [36] are another solution for covariance and were also explored in the present work. Alternative supervised chemometric classification models include Random Forests and Support Vector Machines (SVMs), which are high-capacity models. LDA was favored over Random Forests because high-capacity models tend to overfit data when $p > n$ [37]. SVMs, however, are relatively resistant to overfitting and are worth exploring in tasks that LDA performs poorly [38].

The hydrocarbon fluid products in this study were formulated and manufactured as hydraulic fluids. Hydraulic fluids are used in engine-powered vehicles to transfer energy through transmission systems, provide lubrication, regulate temperatures, and seal internal systems [2]. In addition to ensuring regulatory compliance, it is necessary to regularly inspect fuel systems for cross-contamination, as improperly introduced hydrocarbon fluids can have severe consequences for engine-powered vehicles, such as military jets [39–41]. It is increasingly important to employ efficient and rapid methods of classifying hydrocarbon fluids as fleet sizes continue to expand [42].

2. Materials and Methods

2.1. Hydraulic Fluid Samples

Four synthetic-based hydraulic fluids were tested. MIL-PRF-87257C (polyol ester-based) was obtained from Radco Industries (Batavia, IL, USA). MIL-PRF-83282D (PAO-based) was obtained from Lanxess Royco (Fords, NJ, USA). MIL-PRF-87252C (PAO-based) was obtained from Castrol Branco (Wayne, NJ, USA). HyJet IV-A (phosphate ester-based) was obtained from Exxon Mobil (Port Allen, LA, USA).

Two petroleum-based hydraulic fluids were tested. MIL-PRF-5606J was obtained from Shell Aeroshell (Houston, TX, USA). MIL-PRF-5606H was obtained from Lanxess Royco (Baytown, TX, USA).

2.2. Use of Handheld Raman Spectroscopy

A handheld Raman spectrometer (HandyRam™, Field Forensic Inc., St. Petersburg, FL, USA) with a 785 nm, 70 mW laser was employed for spectral collection. A closed vial compartment was used to eliminate ambient light variability. The spectrometer has an 8 mm working distance and a sampling area that is 2.5 mm in diameter. The integration time of 5.0 s for each spectrum was determined by the autointegration feature. The spectral range was recorded from 400 to 2300 cm^{-1} at 1 cm^{-1} intervals, with a spectral resolution of 12–14 cm^{-1}. The measurements were conducted over five days. Although factors such as instrument drift or ambient condition variability were not directly monitored, samples from each hydraulic fluid were measured each day to avoid systemic bias. Spectral comparisons were performed using the Peak software (V1.01.0068, Snowy Range Instruments, Laramie, WY, USA). Samples were each placed into 2 mL type 1 borosilicate amber vials (Glass Vials, Hanover, MD, USA) for Raman spectral measurements.

Singular stock containers of each hydraulic fluid were used for testing. In total, 42 samples were aliquoted from each of the 7 hydraulic fluids. Each sample was measured ten times, resulting in a dataset of 420 spectra used to perform principal component analysis (PCA) as well as to train a linear discriminant analysis (LDA) model. Additionally, a second dataset was generated by measuring an additional 18 samples, aliquoted from 3 different classes of hydraulic fluids. The test set samples were also measured ten times each, resulting in a total of 180 spectra in the test dataset. This test dataset served as unseen data for the LDA model.

2.3. Pre-Processing and Chemometric Analysis

Spectral data were imported to MATLAB R2025a software (version 25.1.0.2943329, MathWorks, Natick, MA, USA) [43] for data treatment and PCA, followed by LDA. MATLAB R2025a was used to generate the figures in this paper.

The data were smoothed and baselined using the asymmetric least squares model [44] with a smoothing parameter (λ) of 1×10^6 and an asymmetry parameter (p) of 0.05. Afterwards, the data were vector-normalized using the L2 norm as demonstrated by other groups [45,46]. Per convention [47], the data were mean-centered at the onset of each chemometric analysis, enabling the algorithms to assign scores based on interclass variation. The LDA model used a Moore–Penrose pseudoinverse solution for covariance [35]. Uniform LDA class priors were used to avoid bias from unequal class sizes.

Because of the limited number of samples tested in this work, ten repeated measurements from each sample were treated as independent observations in both PCA and LDA. This approach could address potential instrument drift and ambient condition variability during data collection; however, this strategy could potentially lead to model overfitting or an over-optimistic accuracy assessment.

3. Results and Discussion

3.1. Pre-Processed Raman

3.1.1. Spectra Peak Characteristics

Visual observation reveals several notable differences in the raw spectra of the synthetic- and petroleum-based hydraulic fluids, as shown in Figure 1. The upper end of the spectral range has been shortened for visual clarity.

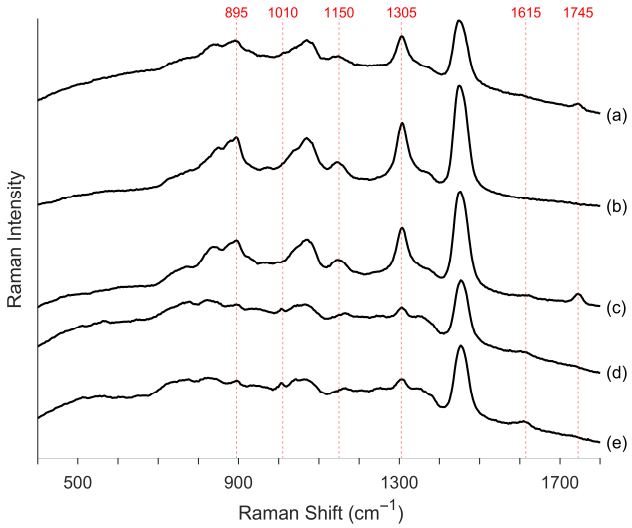


Figure 1. Raw spectra of the synthetic-based hydraulic fluids: (a) 83282D; (b) 87252C; (c) 87257C; and the petroleum-based hydraulic fluids: (d) 5606H; (e) 5606J.

At 895 cm^{-1}, 1070 cm^{-1}, 1305 cm^{-1}, and 1455 cm^{-1}, the synthetic-based samples have more intense Raman scattering signals compared to the petroleum-based samples. All spectra exhibit a relative minimum (intensity dip) at 860 cm^{-1}, which is sharper in the synthetic-based samples. A small, sharp peak in the petroleum-based samples can be observed at 1010 cm^{-1}, as well as a small shoulder peak at 1350 cm^{-1}, that are not present in the synthetic-based samples. Additionally, the spectra of the petroleum-based samples exhibit a peak at 1615 cm^{-1}, which appears to have a small, broad counterpart in the synthetic-based samples at 1620 cm^{-1}. Conversely, a peak in the synthetic-based samples (excluding 87252C) at 1745 cm^{-1} appears to have a small, broad counterpart in the

petroleum-based samples at 1730 cm^{-1}. Similarly, a peak in the synthetic-based samples at 1150 cm^{-1} has a smaller counterpart in that of the petroleum-based samples at 1170 cm^{-1}. These observations, while not exhaustive, are summarized in Table 1.

Table 1. Summary of the spectra differences used to distinguish synthetic- and petroleum-based hydraulic fluids.

Spectra Difference	Peak Wavenumber (cm^{-1})
Larger peak in synthetic-based spectra	895 1070 1150 [1] 1305 1455 1745 [2]
Larger peak in petroleum-based spectra	1615 [3]
Sharper minimum in the synthetic-based spectra	860
Unique peak in petroleum-based spectra	1010
	1350

[1] The corresponding peak in the petroleum-based spectra is at 1170 cm^{-1}. [2] The corresponding peak in the petroleum-based spectra is at 1730 cm^{-1}. Neither peak is present in MIL-PRF-87252C spectra. [3] The corresponding peak in the synthetic-based spectra is at 1620 cm^{-1}.

The vibrational assignments for these Raman spectral peaks are provided in Table 2.

Table 2. Molecular vibrations associated with selected Raman spectral peaks that occur in hydraulic fluid compounds.

Molecular Vibration	Peak Wavenumber (cm^{-1})
Alkane C-C stretching [48,49]	860
	1070
In-plane H-C-H scissoring [50]	895
Monocyclic aromatic H breathing [48]	1010
Iso-alkane C-C skeletal stretching [48]	1150
Alkane bending [48]	1305
	1455
PAH [1] H [48]	1350
Alkene C=C stretching [51]	1615
Carbonyl C=O stretching [49]	1745

[1] Polycyclic aromatic hydrocarbon.

3.1.2. Baseline Intensities

The baseline intensities of the spectra vary widely between hydraulic fluids, and even between different samples originating from the same hydraulic fluid. Most importantly, the petroleum-based spectra consistently have higher baseline intensities than spectra from two out of the three synthetic-based hydraulic fluids: 87257C and 87252C. The median intensity at 400 cm^{-1} is 19,247 for the petroleum-based spectra and 6456 (a 66.8% decrease) for the synthetic-based spectra. Using a one-tailed Welch *t*-test [52] to compare the mean 400 cm^{-1} intensities between the two groups yields a *p*-value of 2.7×10^{-63}. However, the relative standard deviation of 400 cm^{-1} intensities is 11.5% for the petroleum-based spectra and 64.2% for the synthetic-based spectra. Note that smaller-scale variance was observed when

comparing different samples of the same hydraulic fluid. When different synthetic-based fluids were compared to each other, 83282D consistently had a higher baseline (approximately twice as intense). This could be caused by the fluorescent variability of each product. Factors such as sample heterogeneity are not likely to explain the baseline variations, as synthetic-based hydrocarbon fluids are typically uniform in structure. Therefore, the high variance was mainly a result of intra-class, inter-product variation. Despite the statistically significant difference in baseline intensities between these two groups, peak shapes and positions were more consistent and predictable within each category and would serve as better discriminating factors if the experiment was replicated across different instruments or laboratories. Therefore, baseline intensity differences were not used for discriminating between synthetic- and petroleum-based hydraulic fluid spectra. All data were normalized and baseline-corrected before PCA and LDA analysis.

3.1.3. Pre-Processing Approach

Standard Normal Variate (SNV) baseline-correction is a commonly employed Raman pre-processing technique [53]. It corrects baseline intensities by subtracting the intensity mean from each spectral data point. As illustrated in Figure 2 below, the raw baseline intensities in the present work adhere to a curved shape.

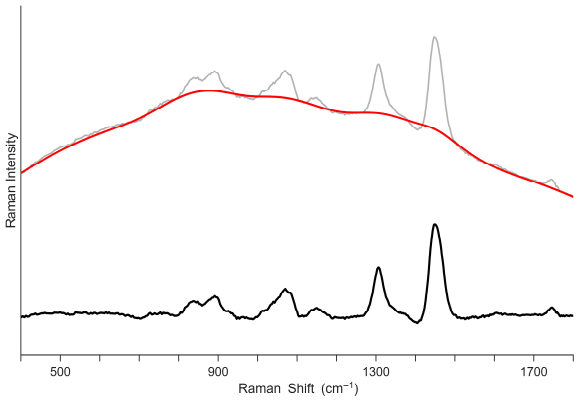


Figure 2. A visual demonstration of the asymmetric least squares (ALS) baselining model applied to 83282D. The spectral baseline (red) is computed and subtracted from the pre-processed spectrum (grey), resulting in the baselined spectrum (black).

When SNV is applied to a curved baseline, it lowers the baseline, yet the baseline retains its curved form because all spectral intensities are reduced by a fixed amount. To avoid this issue, we instead applied asymmetric least squares (ALS) baseline-correction, which uses a penalized least-squares equation to variably subtract baseline intensities from each spectral point. ALS is a more appropriate technique for curved, asymmetric baselines such as those discussed in this section. Unlike SNV, ALS does not include a normalization equation. Therefore, ALS was complemented with vector normalization.

Savitzky–Golay smoothing is a common pre-processing technique that addresses high-frequency noise in spectra [54]. It was explored in the present work but removed from the final workflow due to its negligible effect on the spectra, which do not contain high-frequency noise.

3.1.4. Phosphate Ester-Based Hydraulic Fluid

HyJet IV-A belongs to a subclass of synthetic-based hydraulic fluids, specifically phosphate ester-based hydraulic fluids. Its unique spectral features can be seen to deviate from the other synthetic hydraulic fluids, as shown in Figure 3. PCA analyses revealed

that this sample did not cluster with the other synthetic hydraulic fluids; therefore, it was excluded from the main PCA analysis but included in the LDA analysis as its own category.

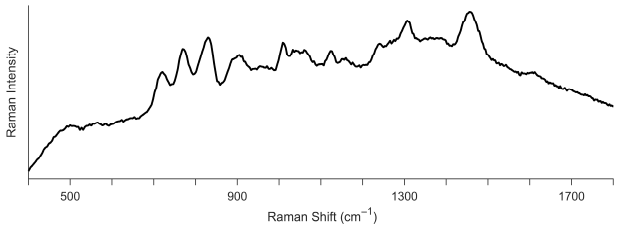


Figure 3. Raw spectra of HyJet IV-A, a phosphate ester-based hydraulic fluid.

3.2. Principal Component Analysis of Processed Spectra

3.2.1. Principal Component Analysis with a Phosphate Ester-Based Hydraulic Fluid

The distinctions discussed in Section 3.1.1. and summarized in Table 1 were still observed after the data were baseline-corrected and normalized. After pre-processing, principal component analysis was performed on the training dataset. In Figure 4, individual spectra are plotted using two-dimensional scores derived from PC1 and PC2.

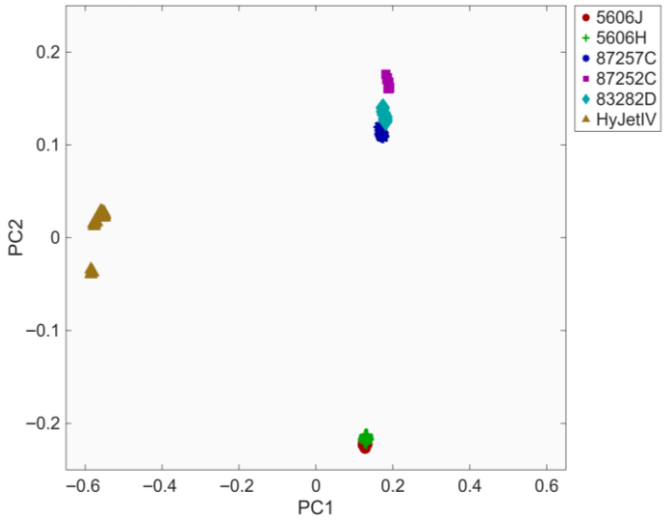


Figure 4. Principal Component 2 versus Principal Component 1 for all training set spectra.

As discussed previously, HyJet IV-A is a phosphate ester-based hydraulic fluid and did not cluster with the other synthetic hydraulic fluids. The following section covers PCA performed on a version of the training dataset with HyJet IV-A samples removed.

3.2.2. Principal Component Analysis Without a Phosphate Ester-Based Hydraulic Fluid

The plots in Figure 5 display the relative contributions of each wavenumber towards Principal Component 1 (PC1) (a) and Principal Component 2 (PC2) (b). Wavenumbers with higher absolute values have a larger effect on the model's classifications.

As expected, the wavenumbers featured in PC1 correlate to the visual observations discussed in Section 3.1.1. The peaks at 895 cm^{-1}, 1305 cm^{-1}, and 1455 cm^{-1} have the largest contributions to PC1. The peak at 1745 cm^{-1} has a small contribution to PC1 and, notably, the largest contribution to PC2.

It is also notable that each of the three most weighted peaks in PC1 are present in both classes and more intense in the synthetic-based samples. As seen in Table 2 (Section 3.1.1), those peaks are associated with aliphatic chains. Compared to petroleum-based hydrocarbon fluids, synthetic-based samples are expected to contain a higher aliphatic proportion due to the backbones and various sidechains of PAO and polyol ester-based compounds.

By contrast, the peaks associated with aromatic content—1010 cm^{-1}, 1350 cm^{-1}, and 1615 cm^{-1}—are each more pronounced in, or exclusive to, the petroleum-based samples. Aromatic compounds lower the thermal and oxidative stability of hydrocarbon fluid products [2]. In addition, PAHs are an environmental hazard and industry regulations limit their proportions in manufactured fluids [55]. As a result, aromatic compounds are scarce in synthetic-based products, which have selectively tailored traits. Aromatics are generated during the petroleum refining process and are difficult to fully remove during downstream treatment processes, leading to their presence in petroleum-based products [55]. Nonetheless, the peaks indicative of aromatic content have relatively minor intensities compared to the major peaks that signal the tailored aliphatic content of synthetic-based samples.

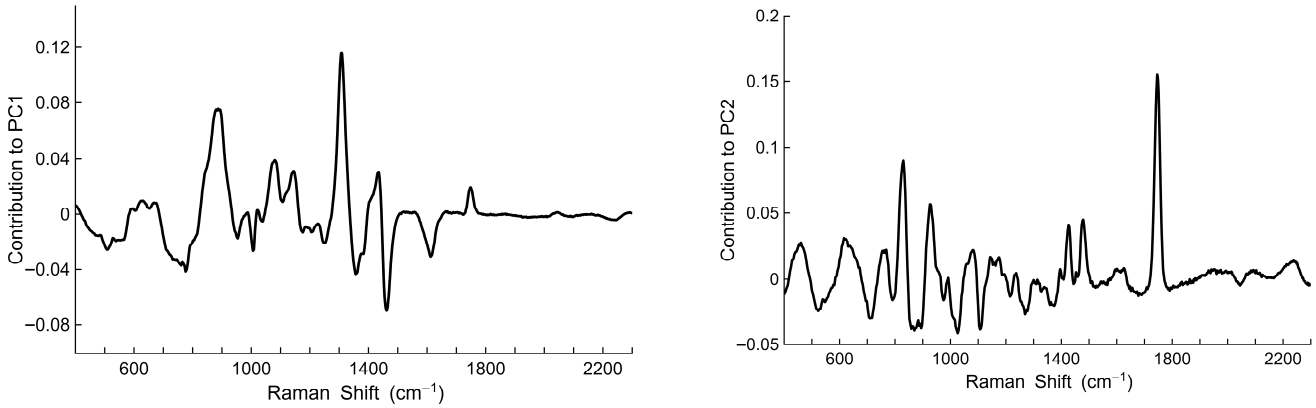


Figure 5. A graph of spectral wavenumbers and their relative contributions to (**a**) PC1 and (**b**) PC2.

As visualized by a scree plot (Figure 6), PC3 and higher did not exhibit significant variance among the samples; therefore, the PCA in this study is limited to the first two principal components. The scree plot also reveals that PC2 accounts for less than 5% of the variance between samples. This explains the large contribution made to PC2 by the 1745 cm^{-1} peak—a minor feature in the context of the entire spectrum. A principal component that explains a low percentage of the overall variance can be expected to highlight features that are slightly visible or not visible at all through direct examination.

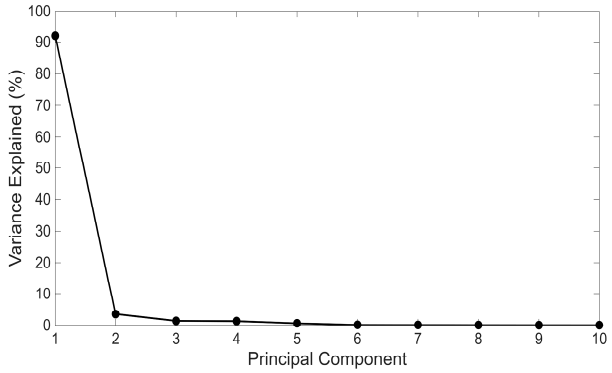


Figure 6. A scree plot displaying the percentage of the dataset's total variance captured by each principal component.

In Figure 7, individual spectra are plotted using two-dimensional scores derived from PC1 and PC2. While PC2 provided intraclass separation using small spectral features, PC1 effectively addressed the research question by providing a clear separation between synthetic- and petroleum-based hydraulic fluids.

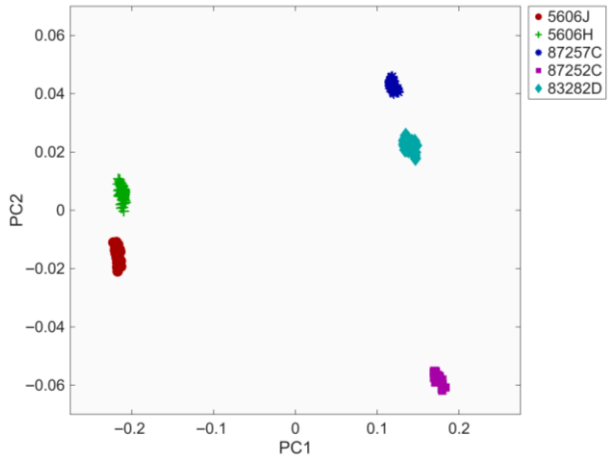


Figure 7. Principal Component 2 versus Principal Component 1 for all training set spectra excluding phosphate ester-based hydraulic fluid HyJet IV-A.

3.3. Linear Discriminant Analysis of Processed Spectra

Linear discriminant analysis (LDA) is a supervised classification method [33]. The user inputs labeled data, which is used to create a best-fit model. The model is then capable of predicting classifications of future data points based on the labeled observations used during the training process. In the present work, the training set was used to generate a model that classifies hydraulic fluids as synthetic- or petroleum-based. Figure 8 is a visual plot of the training data using Linear Discriminant 1 (LD1) and Linear Discriminant 2 (LD2).

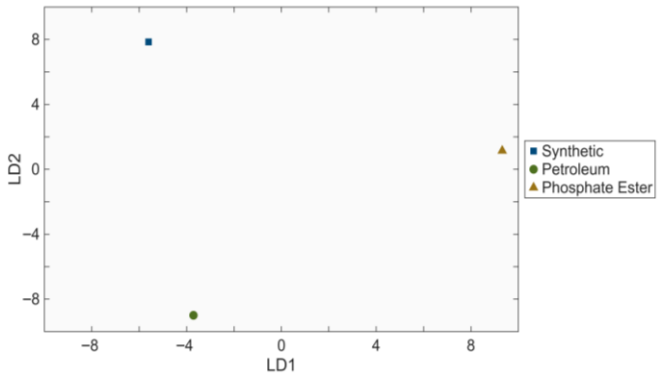


Figure 8. A plot of Linear Discriminant 2 (LD2) versus Linear Discriminant 1 (LD1) for the LDA model generated by the training dataset.

The LDA model was asked to classify the test set (180 spectra) data into three established categories. Figure 9 is a visual plot of the LD scores assigned to the test set. The model classified the dataset with 100% accuracy.

The results were validated using *K*-fold cross-validation (K = 10), which produced 100% model accuracy. The 95% confidence intervals (CIs) for the model's precision (selectivity) were 96.0–100.0% for the synthetic class, 94.0–100.0% for the petroleum class, and 88.4–100.0% for the phosphate ester class. The same intervals were produced upon calculating the model's recall (sensitivity) 95% CIs. Therefore, 95% CIs for the F1-scores (harmonic mean of precision and recall) were 0.96–1.0 for the synthetic class, 0.94–1.0 for the petroleum class, and 0.884–1.0 for the phosphate ester class. Figure 10 shows the confusion matrix depicting the model's performance by class.

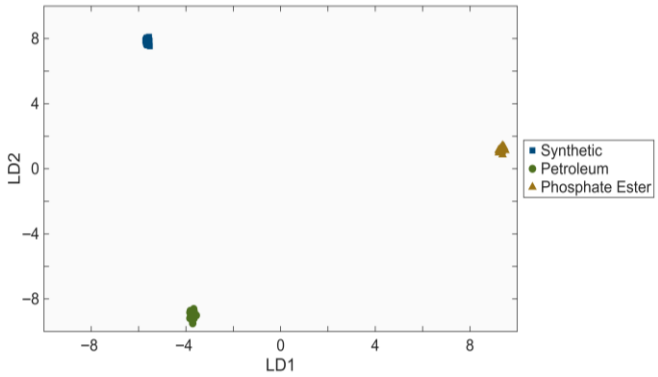


Figure 9. A plot of the test dataset analyzed by the LDA model.

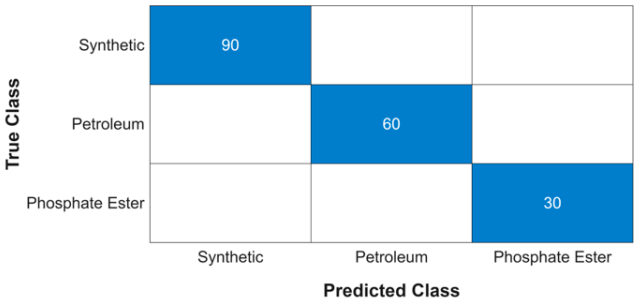


Figure 10. A confusion matrix for the LDA model showing the number of spectra for each possible classification outcome.

A training/test split of 70:30 is often recommended for classification models [56,57]. With the limited number of hydraulic fluid products in this study, the LDA model was refit after removing 83282D, 87252C, and 5606H from the training dataset, resulting in one representative product per classification group. The products that were removed or retained in the training set were chosen randomly. The size of the training set was therefore reduced from $n = 420$ to $n = 210$, while the test set remained at $n = 180$ with no additions or removals. Figure 11 shows the model's performance after the training set removals. Ultimately, the model classified all samples, including the holdout samples in the test set, with 100% accuracy.

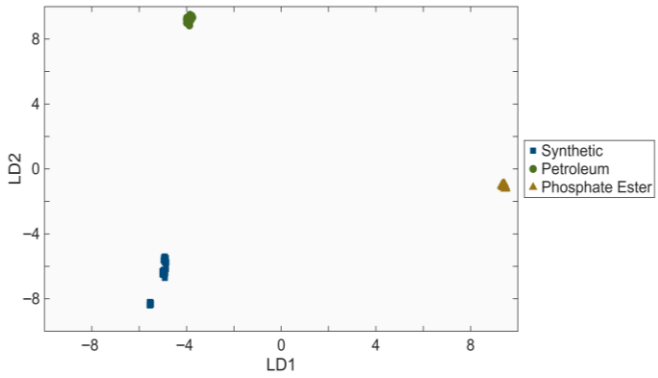


Figure 11. The prediction outcome for testing data by the LDA model generated with a reduced training dataset.

The LDA model shown thus far used a Moore–Penrose solution for covariance [35]. Another viable approach is to use shrinkage regularized linear discriminants [36]. In the present work, a *K*-fold cross-validated ($K = 10$) grid search was performed, after removing constant predictors, to determine the optimal shrinkage parameter, γ. A total of 21 γ values,

from 0 to 1 in intervals of 0.05, were covered in the grid search. The optimal γ value was determined to be 0, which produced a *K*-fold accuracy of 100%. When the shrinkage parameter is 0, no shrinkage occurs. This result indicates that shrinkage was not necessary for the LDA model to effectively separate the classes, despite the condition of $p > n$. We hypothesize that this is due to the combination of using low-noise data and removing constant predictors before applying shrinkage.

4. Conclusions

In summary, hydraulic fluids were used to demonstrate the analysis of hydrocarbon fluids via a handheld Raman spectroscopy. Synthetic-, petroleum-, and phosphate ester-based hydrocarbon fluids were all classified by visual observation of the spectra. LDA is generally a better technique than PCA for categorizing samples that belong to a few distinct classes. PCA was nonetheless able to separate hydrocarbon fluids by class using the first principal component, and the second principal component provided intraclass separation, although PC2's separation criteria were unclear. The variability generated by PC3 and above was negligible due to the overall similarities between spectra. Regardless, the results of this work support the use of a handheld Raman spectroscopy for the classification of hydrocarbon fluids with different manufacturing processes.

Sample diversity is a limitation of this study. The hydraulic fluid products were each obtained in singular quantities, preventing the analysis of batch variation as well as inter-manufacturer variability. The generalizability of the PCA and LDA models in this work cannot be fully assessed with the six hydrocarbon fluid sources that were used. In future work, a more robust model could be developed with a diversified set of training samples.

Furthermore, future work can investigate the clear differentiation of hydrocarbon fluid compounds within the same classification group. In fuel supply, forensic science, and other field applications, the implementation of this technique could improve resource optimization by serving as a reliable, quick screening method. On-site detection of ignitable liquids in forensic casework could prove to be a strong tool since the target samples are highly volatile and cannot always be preserved. In the case of jet fuel analysis, samples containing unexpected fluids could be addressed sooner, thereby preventing the adverse outcomes associated with using contaminated jet fuel.

In addition, this study also draws attention to the expanding applications of a handheld Raman. The instrument's prevalence may continue to grow as technical leaders seek to implement newer technologies in fieldwork, replacing simpler screening methods such as color tests. Some handheld Raman spectrometers allow users to build customizable spectral databases, which is an invaluable feature for rapid chemical identification.

Author Contributions: Conceptualization, J.C.-C.Y.; methodology, J.C.-C.Y. and J.E.H.; software, J.E.H.; validation, J.E.H., K.M. and J.C.-C.Y.; formal analysis, J.E.H.; investigation, K.M.; resources, J.C.-C.Y., J.E.H. and K.M.; data curation, J.E.H.; writing—original draft preparation, J.E.H.; writing—review and editing, J.C.-C.Y., G.M., J.E.H. and K.M.; visualization, J.E.H.; supervision, J.C.-C.Y.; project administration, J.C.-C.Y.; funding acquisition, J.C.-C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Department of Forensic Science and the Graduate School of Sam Houston State University. No funding number was assigned for this project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in this study are openly available in Figshare at https://doi.org/10.6084/m9.figshare.c.7902725.v4 (accessed on 17 July 2025).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Forsberg, C. What Is the Long-Term Demand for Liquid Hydrocarbon Fuels and Feedstocks? *Appl. Energy* **2023**, *341*, 121104. [CrossRef]
2. Bart, J.; Gucciardi, E.; Cavallaro, S. Bioloubricant Product Groups and Technological Applications. In *Biolubricants*; Woodhead Publishing: Sawdust, UK, 2013; pp. 565–711, ISBN 978-0-85709-263-2.
3. Stauffer, E.; Dolan, J.A.; Newman, R. *Fire Debris Analysis*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2007; ISBN 978-0-12-663971-1.
4. Picardi, G.; Cattaruzza, F.; Mangione, D.; Manzo, F.; Terracciano, A.; Proposito, A. Rapid Screening of Designer Fuel Frauds by Raman Spectroscopy. *Talanta Open* **2024**, *9*, 100333. [CrossRef]
5. Doble, P.; Sandercock, M.; Du Pasquier, E.; Petocz, P.; Roux, C.; Dawson, M. Classification of Premium and Regular Gasoline by Gas Chromatography/Mass Spectrometry, Principal Component Analysis and Artificial Neural Networks. *Forensic Sci. Int.* **2003**, *132*, 26–39. [CrossRef]
6. Ryan, P.; Denne, D.; Wakefield, C.; Warburton, G.; Hazelby, D. An Investigation of the Reproducibility of Results of an Automatic GC/MS/DS Method for the Detection of Organic Contaminants. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *48*, 283–286. [CrossRef]
7. Nowak, P.; Bis, A.; Rusin, M.; Woźniakiewicz, M. Carbon Footprint of the Analytical Laboratory and the Three-Dimensional Approach to Its Reduction. *Green Anal. Chem.* **2023**, *4*, 100051. [CrossRef]
8. Woodrow, J. *The Laboratory Characterization of Arco Jet Fuel Vapor and Liquid*; University of Nevada: Reno, NV, USA, 2000.
9. Mana Kialengila, D.; Wolfs, K.; Bugalama, J.; Van Schepdael, A.; Adams, E. Full Evaporation Headspace Gas Chromatography for Sensitive Determination of High Boiling Point Volatile Organic Compounds in Low Boiling Matrices. *J. Chromatogr. A* **2013**, *1315*, 167–175. [CrossRef]
10. Pires, A.; Han, Y.; Kramlich, J.; Garcia-Perez, M. Chemical Composition and Fuel Properties of Alternative Jet Fuels. *BioResources* **2018**, *13*, 2632–2657. [CrossRef]
11. Fujihara, K.; Fujita, Y.; Yamamoto, T.; Nishimoto, N.; Kimura-Kataoka, K.; Kurata, S.; Takinami, Y.; Yasuda, T.; Takeshita, H. Blood Identification and Discrimination between Human and Nonhuman Blood Using Portable Raman Spectroscopy. *Int. J. Leg. Med.* **2017**, *131*, 319–322. [CrossRef]
12. Huang, T.-Y.; Yu, J.C.-C. Development of Crime Scene Intelligence Using a Hand-Held Raman Spectrometer and Transfer Learning. *Anal. Chem.* **2021**, *93*, 8889–8896. [CrossRef]
13. Hargreaves, M.; Page, K.; Munshi, T.; Tomsett, R.; Lynch, G.; Edwards, H. Analysis of Seized Drugs Using Portable Raman Spectroscopy in an Airport Environment—A Proof of Principle Study. *J. Raman Spectrosc.* **2008**, *39*, 873–880. [CrossRef]
14. Navin, C.; Tondepu, C.; Toth, R.; Lawson, L.; Rodriguez, J. Quantitative Determinations Using Portable Raman Spectroscopy. *J. Pharm. Biomed. Anal.* **2017**, *136*, 156–161. [CrossRef]
15. Wang, J.; Koo, K.; Trau, M. Tetraplex Immunophenotyping of Cell Surface Proteomes via Synthesized Plasmonic Nanotags and Portable Raman Spectroscopy. *Anal. Chem.* **2022**, *94*, 14906–14916. [CrossRef]
16. Daoust, F.; Nguyen, T.; Orsini, P.; Bismuth, J.; de Denus-Baillargeon, M.; Veilleux, I.; Wetter, A.; Mckoy, P.; Dicaire, I.; Massabki, M.; et al. Handheld Macroscopic Raman Spectroscopy Imaging Instrument for Machine-Learning-Based Molecular Tissue Margins Characterization. *J. Biomed. Opt.* **2021**, *26*, 022911. [CrossRef]
17. Lauwers, D.; Hutado, A.; Tanevska, V.; Moens, L.; Bersani, D.; Vandenabeele, P. Characterisation of a Portable Raman Spectrometer for In Situ Analysis of Art Objects. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2014**, *118*, 294–301. [CrossRef]
18. Gueli, A.; Galvagno, R.; Incardona, A.; Pappalardo, E.; Politi, G.; Paladini, G.; Stella, G. Correlation of Visible Reflectance Spectrometry and Portable Raman Data for Red Pigment Identification. *Heritage* **2024**, *7*, 2161–2175. [CrossRef]
19. Jehlička, J.; Vítek, P.; Edwards, H.; Heagreaves, M.; Čapoun, T. Application of Portable Raman Instruments for Fast and Non-Destructive Detection of Minerals on Outcrops. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2009**, *73*, 410–419. [CrossRef] [PubMed]
20. Culka, A.; Jehlička, J.; Opluštil, S. Evaluation of Carbonification of Coals Using a Portable Raman Spectrometer. *J. Raman Spectrosc.* **2023**, *54*, 1220–1232. [CrossRef]
21. Pereira, R.; Skrobot, V.; Castro, E.; Fortes, I.; Pasa, V. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminant Analysis Applied to FTIR Spectra. *Energy Fuels* **2006**, *20*, 1097–1102. [CrossRef]
22. Barbeira, P.; Pereira, R.; Corgozinho, C. Identification of Gasoline Origin by Physical and Chemical Properties and Multivariate Analysis. *Energy Fuels* **2007**, *21*, 2212–2215. [CrossRef]

23. Li, S.; Dai, L. Classification of Gasoline Brand and Origin by Raman Spectroscopy and a Novel R-Weighted LSSVM Algorithm. *Fuel* **2012**, *96*, 146–152. [CrossRef]
24. Yousefinejad, S.; Aalizadeh, L.; Honarasa, F. Application of ATR-FTIR Spectroscopy and Chemometrics for the Discrimination of Furnace Oil, Gas Oil and Mazut Oil. *Anal. Methods* **2016**, *8*, 4640–4647. [CrossRef]
25. Khanmohammadi Khorrami, M.; Sadrara, M.; Mohammadi, M. Quality Classification of Gasoline Samples Based on Their Aliphatic to Aromatic Ratio and Analysis of PONA Content Using Genetic Algorithm Based Multivariate Techniques and ATR-FTIR Spectroscopy. *Infrared Phys. Technol.* **2022**, *126*, 104354. [CrossRef]
26. Xu, J.; Liu, S.; Gao, M.; Zuo, Y. Classification of Lubricating Oil Types Using Mid-Infrared Spectroscopy Combined with Linear Discriminant Analysis–Support Vector Machine Algorithm. *Lubricants* **2023**, *11*, 268. [CrossRef]
27. Biaktluanga, L.; Lalhruaitluanga, J.; Lalramnghaka, J.; Thanga, H. Analysis of Gasoline Quality by ATR-FTIR Spectroscopy with Multivariate Techniques. *Results Chem.* **2024**, *8*, 101575. [CrossRef]
28. Shang, L.; Bao, Y.; Tang, J.; Ma, D.; Fu, J.; Zhao, Y.; Wang, X.; Yin, J. A Novel Polynomial Reconstruction Algorithm-based 1D Convolutional Neural Network Used for Transfer Learning in Raman Spectroscopy Application. *J. Raman Spectrosc.* **2022**, *53*, 237–246. [CrossRef]
29. Zhou, W.; Qian, Z.; Ni, X.; Tang, Y.; Guo, H.; Zhuang, S. Dense Convolutional Neural Network for Identification of Raman Spectra. *Sensors* **2023**, *23*, 7433. [CrossRef] [PubMed]
30. Zhou, Y.; Tang, X.; Zhang, D.; Lee, H. Machine Learning Empowered Coherent Raman Imaging and Analysis for Biomedical Applications. *Comms. Eng.* **2025**, *4*, 8. [CrossRef]
31. Wang, Z.; Ranasinghe, J.; Wu, W.; Chan, D.; Gomm, A.; Tanzi, R.; Zhang, C.; Zhang, N.; Allen, G.; Huang, S. Machine Learning Interpretation of Optical Spectroscopy Using Peak-Sensitive Logistic Regression. *ACS Nano* **2025**, *19*, 15457–15473. [CrossRef] [PubMed]
32. Srivastava, S.; Wang, W.; Zhou, W.; Jin, M.; Vikesland, P. Machine Learning-Assisted Surface-Enhanced Raman Spectroscopy Detection for Environmental Applications: A Review. *Environ. Sci. Technol.* **2024**, *58*, 20830–20848. [CrossRef] [PubMed]
33. Mishra, C.; Gupta, D. Deep Machine Learning and Neural Networks: An Overview. *IAES Int. J. Artif. Intell. IJ-AI* **2017**, *6*, 66–73. [CrossRef]
34. Amouzgar, M.; Glass, D.; Baskar, R.; Averbukh, I.; Kimmey, S.; Tsai, A.; Hartmann, F.; Bendall, S. Supervised Dimensionality Reduction for Exploration of Single-Cell Data by HSS-LDA. *Patterns* **2022**, *3*, 100536. [CrossRef] [PubMed]
35. Gyamfi, K.; Brusey, J.; Hunt, A.; Gaura, E. Linear Classifier Design under Heteroscedasticity in Linear Discriminant Analysis. *Expert Syst. Appl.* **2017**, *79*, 44–52. [CrossRef]
36. Krzanowski, W.; Jonathan, P.; McCarthy, W.; Thomas, M. Discriminant Analysis with Singular Covariance Matrices: Methods and Applications to Spectroscopic Data. *J. R. Stat.* **1995**, *44*, 101. [CrossRef]
37. Barreñada, L.; Dhiman, P.; Timmerman, D.; Boulesteix, A.; Van Calster, B. Understanding Overfitting in Random Forest for Probability Estimation: A Visualization and Simulation Study. *Diagn. Progn. Res.* **2024**, *8*, 14. [CrossRef]
38. Mirjalili, S.; Powell, P.; Strunk, J.; James, T.; Duarte, A. Evaluation of Classification Approaches for Distinguishing Brain States Predictive of Episodic Memory Performance from Electroencephalography. *Neuroimage* **2022**, *247*, 118851. [CrossRef]
39. Federal Aviation Administration. *Jet Fuel Contamination with Diesel Exhaust Fluid (DEF)*; Federal Aviation Administration: Washington, DC, USA, 2023.
40. Stamker, D.G.; Tartakovsky, K.; Rabaev, M. Identification and Quantification of Phosphate Ester-Based Hydraulic Fluid in Jet Fuel. *SAE Int. J. Fuels Lubr.* **2019**, *12*, 43–50. [CrossRef]
41. U.S. Air Force. Department of the Air Force (DAF) 23.3 Small Business Innovation Research (SBIR) Direct to Phase II (D2P2) Proposal Submission Instructions Amendment 2. Available online: https://media.defense.gov/2023/Aug/22/2003285647/-1/-1/0/AF_SBIR_233_DP2.PDF (accessed on 17 January 2023).
42. Secretary of the Air Force. *The Department of the Air Force in 2050*; Department of Defense: Arlington, VA, USA, 2024.
43. *MATLAB R2025a*, version 25.1.0.2943329; MathWorks: Natick, MA, USA, 2025.
44. Eilers, P.; Boelens, H. *Baseline Correction with Asymmetric Least Squares Smoothing*; Leiden University Medical Centre: Leiden, The Netherlands, 2005.
45. Kanno, N.; Kato, S.; Ohkuma, M.; Matsui, M.; Iwasaki, W.; Shigeto, S. Machine Learning-Assisted Single-Cell Raman Fingerprinting for In Situ and Nondestructive Classification of Prokaryotes. *iScience* **2021**, *24*, 102975. [CrossRef]
46. Yang, Y.; Ling, X.; Qiu, W.; Bian, J.; Zhang, X.; Chen, Q. Surface-Enhanced Raman Scattering Spectroscopy Reveals the Phonon Softening of Yttrium-Doped Barium Zirconate Thin Films. *J. Phys. Chem. C* **2022**, *126*, 10722–10728. [CrossRef]
47. Wise, B.; Gallagher, N. The Process Chemometrics Approach to Process Monitoring and Fault Detection. *J. Process Control* **1996**, *6*, 329–348. [CrossRef]
48. Kuptsov, A.; Arbuzova, T. A Study of Heavy Oil Fractions by Fourier-Transform near-Infrared Raman Spectroscopy. *Pet. Chem.* **2011**, *51*, 203–211. [CrossRef]

49. Böke, J.; Popp, J.; Krafft, C. Optical Photothermal Infrared Spectroscopy with Simultaneously Acquired Raman Spectroscopy for Two-Dimensional Microplastic Identification. *Sci. Rep.* **2022**, *12*, 18785. [CrossRef]
50. Moosavinejad, S.; Madhoushi, M.; Vakili, M.; Rasouli, D. Evaluation of Degradation in Chemical Compounds of Wood in Historical Buildings Using FT-IR and FT-Raman Vibrational Spectroscopy. *Maderas Cienc. Tecnol.* **2019**, *21*, 381–392. [CrossRef]
51. Gieleciak, R.; Hall, A.; Michaelian, K.; Chen, J. Exploring the Potential of Raman Spectroscopy for Characterizing Olefins in Olefin-Containing Streams. *Energy Fuels* **2023**, *37*, 13698–13709. [CrossRef]
52. Ruxton, G. The Unequal Variance *t*-Test Is an Underused Alternative to Student's *t*-Test and the Mann-Whitney U Test. *Behav. Ecol.* **2006**, *17*, 688–690. [CrossRef]
53. Genkawa, T.; Shinzawa, H.; Kato, H.; Ishikawa, D.; Murayama, K.; Komiyama, M.; Ozaki, Y. Baseline Correction of Diffuse Reflection Near-Infrared Spectra Using Searching Region Standard Normal Variate (SRSNV). *Appl. Spectrosc.* **2015**, *69*, 1432–1441. [CrossRef] [PubMed]
54. Bai, Y.; Liu, Q. Denoising Raman Spectra by Wiener Estimation with a Numerical Calibration Dataset. *Biomed. Opt. Express* **2020**, *11*, 200. [CrossRef]
55. Ravindra, K.; Sokhi, R.; Vangrieken, R. Atmospheric Polycyclic Aromatic Hydrocarbons: Source Attribution, Emission Factors and Regulation. *Atmos. Environ.* **2008**, *42*, 2895–2921. [CrossRef]
56. Dobbin, K.; Simon, R. Optimally Splitting Cases for Training and Testing High Dimensional Classifiers. *BMC Med. Genom.* **2011**, *4*, 31. [CrossRef]
57. Gholamy, A.; Kreinovich, V.; Kosheleva, O. *Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation*; University of Texas at El Paso: El Paso, TX, USA, 2018.

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.