Development of An Intelligent Handheld Raman Analyzer for Scene Investigation

Ting-Yu Huang[1], M.S.*, and Jorn Chi-Chung Yu, PhD
[1]Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340

Forensic Technology CENTER OF EXCELLENCE

NIJ | National Institute of Justice — STRENGTHEN SCIENCE. ADVANCE JUSTICE.

RTI INTERNATIONAL

1. Introduction

Gas chromatography-mass spectrometry (GC/MS) is the commonly accepted analytical instrumentation to identify gasoline, which is one of the most widely used ignitable liquids in arson cases. However, the use of standard GC/MS testing is subject to the accredited laboratories and is time-consuming in sample preparation and chromatographic separation. Besides, the interpretation of GC/MS data is labor-intensive and relies on an analyst's skill and experience.

Raman spectroscopy provides spectral fingerprints of the analyte by observing photon scattering and the change in photon energy. Considering its portability and short spectra acquisition time, the technique offers potential for serving as a screening tool for chemicals in the field.

In order to increase the efficiency of Raman spectrum comparison for identification, we proposed to develop an artificial intelligence (AI) algorithm based on transfer learning of a pre-trained convolutional neural network (CNN). A CNN has multiple layers in the architecture acting as filters to extract features from input data without manual engineering. Since developing a CNN requires a large reference data set and is computationally expensive, reusing the knowledge of an existing CNN and revising its network, viewed as transfer learning, is a desirable alternative. In this work, an intelligent Raman analyzer was developed by combining a handheld Raman spectrometer and transfer learning for on-site gasoline grade discrimination to assist fire investigation.

2. Objectives

Create a smart AI classifier to discriminate gasoline samples with various octane numbers.

- ▶Rapid, automated, and accurate workflow
- ▶On-site screening test

Explore the feasibility of continuous wavelet transform (CWT) in image analysis of Raman scattering signals.

Explore how transfer learning can be applied to develop an AI classifier and evaluate the performance of the AI classifier.

3. Methods

Sample Description

Gasoline samples with three octane numbers: 87 regular, 89 mid-grade, and 93 premium, were collected from one fuel station in Houston, Texas. The Raman database comprised of training, verification, and evaluation data sets. The training and verification data sets were created from neat gasoline samples with triplicates of 1.0 mL aliquots sampled from each grade of gasoline. The evaluation data set was created from two levels (50% and 25%) of weathering process with duplicates of 1.0 mL aliquots sampled from each grade of gasoline. The weathering process was to reduce the volume of gasoline sample by evaporation.

Raman Spectra Collection

A total of 450 Raman spectra were collected from a handheld Raman spectrometer (HandyRam^{TM}, Field Forensic Inc., USA) with a 785 nm laser wavelength for training (345 spectra), verification (45 spectra), and evaluation (60 spectra). The spectral range was recorded from 400 to 2300 cm^{-1} with autointegration time. Baseline subtraction was conducted to correct the background signal.

Raman Spectra Processing

CWT with the analytic Morse wavelet filter bank was utilized to create the scalograms, the absolute value of CWT coefficients of a signal in the spectra (Figure 1). Table 1 lists the parameters for CWT computations. Each scalogram was generated in an RGB image that was an array of size 224-by-224-by-3.

Transfer Learning

GoogLeNet, a pre-trained CNN, was revised and fine-tuned to learn the task of gasoline grade classification. Revision of the network included replacing the final three layers (dropout, loss3-classifier, output) with new layers in order to adapt to the new training data. Fine-tuning was to adjust training options as shown in Table 1.

4. Results

Model Training

Figure 2 shows the training progress of the proposed model (GoogLeNet). The final validation accuracy achieved at 95.65%, which indicates that the pre-trained CNN has successfully learned to differentiate the pattern of scalograms generated from three grades of gasoline Raman spectra.

Model Verification and Evaluation

Compared with the six classical machine learning models, the GoogLeNet model was the only classifier that could correctly classify all the scalograms in the verification data set into correct gasoline grades. Further investigation implies that the most easily confused classes for classical machine learning algorithms were 89 mid-grade and 93 premium gasoline. The reasons were due to the highly similar spectra profiles between the two grades (Figure 3a) and the variations in peak height and ratio between different sampling aliquots (Figure 3b).

Though the spectra altered when gasoline samples were evaporated (Figure 3c), the GoogLeNet model could still offer 73% and 53% accuracy for 50% and 25% weathered gasoline samples respectively, which outperformed other classical machine learning algorithms.

The proposed approach was more capable of discriminating spectral data with complicated features, such as inter similarities or degrading peaks. The performance indicators of the GoogLeNet model all reached to the highest values among all models (Figure 4).

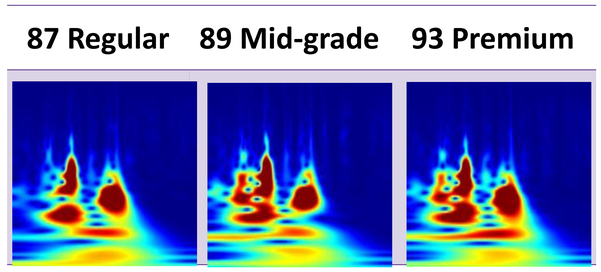


Figure 1: Scalograms generated from three grades of raw gasoline Raman spectra by CWT.

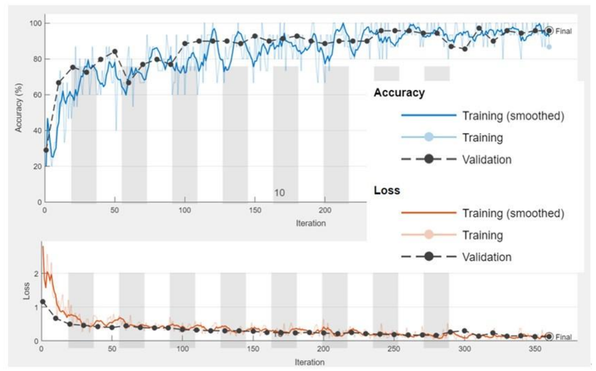
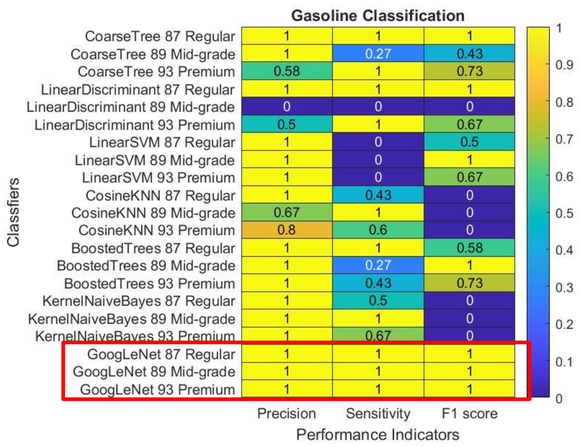
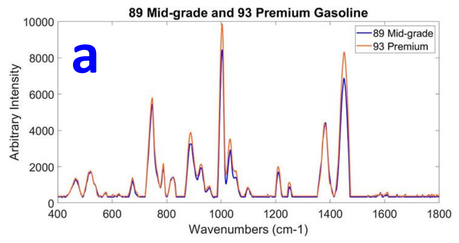
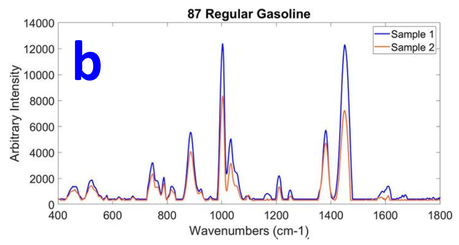


Figure 2: Training progress of the GoogLeNet model.







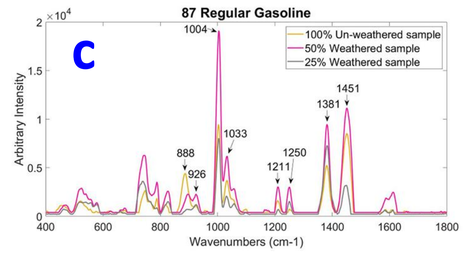


Figure 3: Examples of Raman spectra from the verification and evaluation data sets (a) similar spectral profiles between 89 mid-grade and 93 premium gasoline (b-c) variations of peak intensity, peak ratio, and Raman shift in different aliquots of 87 regular samples and weathered 87 regular samples.

Figure 4: Comparison of the precision, sensitivity, and F1-score of the verification data set.

5. Conclusions

- An AI-powered Raman analyzer was developed for on-site gasoline grade classification.
- Spectral data processing did not depend on hand-crafted engineering.
- The AI analyzer offered high data prediction performance.
- Probability-based reports allow better evidence interpretation in legal proceedings.

3. Methods (Continued)

Classification Methodology

During model training, the scalograms in the training data set were randomly divided into 80% (276 scalograms) and 20% (69 scalograms) for training and validating the model, respectively. The experiments were conducted on Matlab 2021a (MathWorks, Natick, Massachusetts, USA).

Performance Comparison

Six classical machine learning models (decision tree, linear discriminant analysis, SVM, KNN, ensemble method, and Naive Bayes) were constructed using the same training data. The models were verified and evaluated using the same corresponding data sets. The results of verification and evaluation of the proposed GoogLeNet model were compared with the results obtained from the six classical machine learning models. Evaluation measures included accuracy, precision, sensitivity, and F1 score.

Types of Analysis	Parameter	Input
CWT	Signal Length	1000
	Sampling Frequency	Fs
	Voices Per Octave	12
Transfer Learning	Mini Batch Size	15
	Max Epochs	20
	Initial Learn Rate	0.0001
	Validation Frequency	10

6. Potential for Impact

The study provides the forensic science community a new approach for gasoline grade classification by adopting the AI technique. A real-time AI assistant based on Raman spectrometry can be developed for scene investigation.

The study also demonstrates that wavelet analysis is effective in presenting features when transforming Raman scattering signals into 2-dimensional (2D) representations.

7. Next Steps

The proposed method was not yet validated by using gasoline samples from different suppliers. Future work will focus on investigating the impact of various gasoline sources on the performance of the proposed classifier. The proposed workflow can also be applied to fire debris analysis of GC/MS data.

References and Suggested Citation

1. Huang, TY.; Yu, JCC. Development of Crime Scene Intelligence Using a Hand-Held Raman Spectrometer and Transfer Learning. Anal. Chem. 2021, 93(25), 8889–8896.
2. Huang, TY.; Yu, JCC. 2021 Online Forensic Symposium - Current Trends in Forensic Trace Analysis. (2021). Application of Deep Learning and Raman Spectroscopy for On-site Gasoline Grade Classification.
3. Thenmozhi, K.; Srinivasulu Reddy, U. Crop pest classification based on deep convolutional neural network and transfer learning. Computers and Electronics in Agriculture. 2019, 164, 104906.
4. Lilly, JM.; Olhede, SC. Higher-order Properties of Analytic Wavelets. IEEE Trans. on Signal Processing. 2009, 57 (1), 146–160.
5. Lilly, JM.; Olhede, SC. Generalized Morse Wavelets as a Superfamily of Analytic Wavelets. IEEE Trans. Signal Process. 2012, 60 (11), 6036–6041.

Acknowledgments and Disclaimer

The authors thank Sam Houston State University staff for their insight and cooperation in collecting preliminary data.

The study has been accepted for publication in Analytical Chemistry. The project has also partly presented at the 2021 Online Forensic Symposium - Current Trends in Forensic Trace Analysis.

Disclaimer: This presentation was supported by Award No. 2016-MUBX-K110, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

More Information

***Presenting author**: Ting-Yu Huang (Email: txh038@shsu.edu)